

Bangor University

DOCTOR OF PHILOSOPHY

Style in the vernacular and on the radio: code-switching and mutation as stylistic and social markers in Welsh

Prys, Myfyr

Award date:
2016

Awarding institution:
Bangor University

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**Style in the vernacular and on the radio: code-switching and mutation as
stylistic and social markers in Welsh**

Myfyr Prys

School of Linguistics and English language

Bangor University

PhD 2016

Abstract

This thesis seeks to analyse two types of linguistic features of Welsh, code-switching and mutation, as sociolinguistic variables: features which encode social information about the speaker and/or stylistic meaning. Developing a study design that incorporates an analysis of code-switching and mutation in naturalistic speech has demanded a relatively novel methodological approach. The study combined a variationist analysis of the vernacular use of both variables in the 40-hour Siarad corpus (Deuchar 2014) with a technique that ranks radio programmes in order of formality through the use of channel cues and other criteria (Ball et al 1988). This allows for a comparison of the use of code-switching and mutation in multiple stylistic contexts, each of which show varying degrees of emotional engagement and self-monitoring by speakers.

The analysis found that code-switching was strongly correlated with the level of formality of each radio programme, and that at least one aspirate mutation trigger, (a), also patterned in a similar way. Some other mutation triggers, most notably including the nasal possessive trigger (fy), seemed to be primarily affected by the speakers' backgrounds and their relative ages in particular. A qualitative analysis of the type of discourse found in each radio programme made further links between the institutional style of each programme and their use of the stylistically controlled 'marker' variables, with non-standard variants appearing to be indexical of solidarity, subversion and irony, while standard variants indexed prestige, authority and earnestness.

Acknowledgment

Gyntaf oll, hoffwn i ddiolch i fy ngoruchwyliwr Peredur Davies am ei amynedd hirfaith wrth i'r doethuriaeth yma fynd drwy sawl iteriad gwahanol, ac am ei holl gyngor craff a threiddgar. Hoffwn i hefyd ddiolch i Margaret Deuchar, y person ddaru fy nechrau i ym myd ieithyddiaeth, ac a roddodd yr hyder i mi gario ymlaen. Diolch hefyd i fy rhieni a 'nheulu am fy nghefnogi mewn mwy nac un ffordd drwy gydol y broses.

Diolch i Ros Temple am y geiriau o gefnogaeth yn gynnar ymlaen, ac yn enwedig am fy nghyflwyno i waith yr anwogawn Eckert. Diolch i holl gyfranwyr BBC Radio Cymru, nid yn unig am fod yn ffynhonnell ddata gwerth chweil, ond am greu cynnwys diddorol, modern, gwych - rydych chi'n haeddu pob canmoliaeth. Diolch i Fraibet Avaledo, Alberto Rosignoli, Sarah Cooper, Hanna Binks a phawb arall o adran Ieithyddiaeth Bangor ddoe a heddiw, a roddodd eu hamser i drafod materion academaidd diflas gyda fi.

Finally, thanks to Rebecca for all the moral and emotional support over these last years.

Where would I be without you? Go raibh maith agat.

Declaration and Consent

Details of the Work

I hereby agree to deposit the following item in the digital repository maintained by Bangor University and/or in any other repository authorized for use by Bangor University.

Author Name:

Title:

Supervisor/Department:

Funding body (if any):

Qualification/Degree obtained:

This item is a product of my own research endeavours and is covered by the agreement below in which the item is referred to as “the Work”. It is identical in content to that deposited in the Library, subject to point 4 below.

Non-exclusive Rights

Rights granted to the digital repository through this agreement are entirely non-exclusive. I am free to publish the Work in its present version or future versions elsewhere.

I agree that Bangor University may electronically store, copy or translate the Work to any approved medium or format for the purpose of future preservation and accessibility. Bangor University is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Bangor University Digital Repository

I understand that work deposited in the digital repository will be accessible to a wide variety of people and institutions, including automated agents and search engines via the World Wide Web.

I understand that once the Work is deposited, the item and its metadata may be incorporated into public access catalogues or services, national databases of electronic theses and dissertations such as the British Library’s EThOS or any service provided by the National Library of Wales.

I understand that the Work may be made available via the National Library of Wales Online Electronic Theses Service under the declared terms and conditions of use (<http://www.llgc.org.uk/index.php?id=4676>). I agree that as part of this service the National Library of Wales may electronically store, copy or convert the Work to any approved medium or format for the purpose of future preservation and accessibility. The National Library of Wales is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Statement 1:

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless as agreed by the University for approved dual awards.

Signed (candidate)

Date

Statement 2:

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

All other sources are acknowledged by footnotes and/or a bibliography.

Signed (candidate)

Date

Statement 3:

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loan and for electronic storage (subject to any constraints as defined in statement 4), and for the title and summary to be made available to outside organisations.

Signed (candidate)

Date

NB: Candidates on whose behalf a bar on access has been approved by the Academic Registry should use the following version of **Statement 3:**

Statement 3 (bar):

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loans and for electronic storage (subject to any constraints as defined in statement 4), after expiry of a bar on access.

Signed (candidate)

Date

Statement 4:

Choose **one** of the following options

a) I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University and where necessary have gained the required permissions for the use of third party material.	
b) I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University when the approved bar on access has been lifted.	
c) I agree to submit my thesis (the Work) electronically via Bangor University's e-submission system, however I opt-out of the electronic deposit to the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University, due to lack of permissions for use of third party material.	

Options B should only be used if a bar on access has been approved by the University.

In addition to the above I also agree to the following:

1. That I am the author or have the authority of the author(s) to make this agreement and do hereby give Bangor University the right to make available the Work in the way described above.
2. That the electronic copy of the Work deposited in the digital repository and covered by this agreement, is identical in content to the paper copy of the Work deposited in the Bangor University Library, subject to point 4 below.
3. That I have exercised reasonable care to ensure that the Work is original and, to the best of my knowledge, does not breach any laws – including those relating to defamation, libel and copyright.
4. That I have, in instances where the intellectual property of other authors or copyright holders is included in the Work, and where appropriate, gained explicit permission for the inclusion of that material in the Work, and in the electronic form of the Work as accessed through the open access digital repository, *or* that I have identified and removed that material for which adequate and appropriate permission has not been obtained and which will be inaccessible via the digital repository.
5. That Bangor University does not hold any obligation to take legal action on behalf of the Depositor, or other rights holders, in the event of a breach of intellectual property rights, or any other right, in the material deposited.
6. That I will indemnify and keep indemnified Bangor University and the National Library of Wales from and against any loss, liability, claim or damage, including without limitation any related legal fees and court costs (on a full indemnity bases), related to any breach by myself of any term of this agreement.

Signature: Date :

“The curse of mongrel Welsh had been a source of considerable vexation to commentators from the beginnings of Welsh broadcasting”

- R. Smith, *Broadcasting and the Welsh Language* (2000)

“It is necessary that heteroglossia wash over a culture’s awareness of itself and its language, penetrate to its core, relativize the primary language system underlying its ideology and literature and deprive it of its naïve absence of conflict”

- Mikhail Bakhtin, *Discourse in the Novel* (1934-35)

“Ffwcio eich gwaith a cyfalafiaeth”

- Pep Le Pew, *Y Mwyafrif* (2001)

Table of Contents

1. Introduction	27
1.1 Style in the Welsh context.....	28
1.1.1 The social distribution of Welsh.....	31
1.1.2 Diglossia in Wales.....	34
1.1.2.1 Bilingual diglossia.....	36
1.1.2.2 Dialectal diglossia and standard Welsh.....	37
1.1.3 Differences between Standard and Vernacular Welsh.....	38
1.1.3.1 Standard Welsh and English transfer.....	39
1.1.3.2 Standard Welsh and mutation.....	42
1.1.4 The current situation and Welsh speakers' exposure to Welsh.....	42
1.1.4.1 Literature and reading in Welsh.....	43
1.1.4.2 Religious services.....	44
1.1.4.3 Welsh in the public sector.....	45
1.1.4.4 The private sector and the private sphere.....	46
1.1.4.5 Internet-based social networks.....	47
1.1.4.6 The education system.....	48
1.1.4.7 Welsh in the media.....	50
1.1.4.7.1 De-standardization of <i>Radio Cymru</i> ?	51
1.1.4.7.2 Reactions to de-standardization.....	52
1.2 Study design.....	59
1.3 Thesis structure.....	65
2. Style	66
2.1 Introduction.....	66

2.2 The variationist model.....	67
2.2.1 Labov’s attention to speech methodology.....	68
2.2.2 Sociolinguistic variables.....	70
2.2.3 Markers, indicators and stereotypes.....	72
2.2.4 Inter-speaker variation.....	73
2.2.4.1 Change from above and below	74
2.2.4.2 Patterns of inter-speaker variation.....	76
2.2.4.2.1 Apparent time and real time.....	76
2.2.4.2.2 Stability.....	77
2.2.4.2.3 Generational change and lifespan change.....	77
2.2.4.2.4 Age grading.....	78
2.2.4.3 Gender and variation.....	79
2.2.4.4 Class and variation.....	79
2.2.4.5 The social network approach.....	80
2.2.5 Criticism of the variationist approach to style.....	81
2.2.6 Variationism – Summary.....	83
2.3 Communication Accommodation theory.....	84
2.3.1 Accommodation theory – Summary.....	86
2.4 Audience design theory.....	87
2.4.1 Audience design.....	87
2.4.2 Referee design.....	88
2.4.2.1 Referee design and the media.....	90
2.4.3 Stylistic variation reflects social variation.....	90
2.4.4 Empirical evidence for audience design.....	92
2.4.5 Summary of audience design.....	94

2.5 Speaker design approaches.....	95
2.5.1 Indexicality and stances.....	96
2.5.2 Style as practice.....	97
2.5.3 The community of practice.....	99
2.5.4 Authenticity and stylization.....	101
2.5.5 Speaker design approaches – Summary.....	103
2.6 Register variation theory.....	105
2.6.1 Criticism of Register Design.....	106
2.7 Style in minority language contexts.....	108
2.8 Summary.....	109
3. Mutation	112
3.1 Introduction.....	112
3.2 The Welsh mutation system.....	112
3.2.1 Mutation triggers.....	114
3.2.2 Triggering environments.....	114
3.2.3 Internal factors affecting mutation.....	117
3.2.4 Functional load.....	119
3.2.5 Variable zero triggers.....	120
3.2.6 Mutation in Gwynedd.....	123
3.3 Variationist studies of mutation.....	124
3.3.1 P.W. Thomas’ mutation study.....	125
3.3.2 Ball’s Cwmtawe Study.....	128
3.3.2.1 Elicitation Methodology of Bell’s Study.....	130
3.3.2.2 Results of Ball’s study.....	133
3.3.2.3 Index scores in Bell’s study.....	134

3.3.2.4	Inter-speaker variation in Ball's study.....	135
3.3.2.5	Style Shifting in the Cwmtawe study.....	135
3.3.2.6	Markers, indicators and stereotypes in Ball's study.....	136
3.3.2.7	The SM variables in Ball's study.....	137
3.3.2.8	Criticism of Ball's study.....	138
3.3.3	Hatton's school-based mutation study.....	139
3.3.4	Ball, Griffith and Jones' <i>Radio Cymru</i> study.....	142
3.3.5	Jones' language obsolescence study.....	146
3.4	Psycholinguistics and mutation.....	149
3.5	Mutation and attitudes.....	151
3.6	Summary.....	151
4.	Code-switching	153
4.1	Structural factors in code-switching.....	153
4.1.1	Defining borrowings and switches.....	156
4.1.2	The Matrix Language Frame and Welsh.....	158
4.2	Extra-linguistic factors and code-switching.....	161
4.3	Style and code-switching.....	164
4.3.1	Direct speech.....	166
4.3.2	Flagging.....	168
4.3.3	Repairing speech.....	170
4.3.4	Code-switching and humour.....	172
4.4	Variationist approaches to code-switching and style.....	175
4.5	Attitudes toward code-switching.....	177
4.5.1	Empirical evidence on attitudes towards CS in Welsh.....	178
4.6	Code-switching and the Welsh media.....	180

4.7 Summary.....	182
5. <i>Siarad</i> Corpus Analysis Methodology.....	183
5.1 The <i>Siarad</i> corpus.....	183
5.1.1 The collection of the <i>Siarad</i> corpus.....	184
5.1.2 The transcription of the <i>Siarad</i> corpus.....	185
5.1.2.1 Transcription of non-standard forms.....	185
5.1.2.2 Language classification.....	187
5.1.2.3 Other tiers.....	188
5.1.2.4 The quality of transcription.....	188
5.1.3 The make-up of the <i>Siarad</i> corpus.....	188
5.1.3.1 Region.....	189
5.1.3.2 Age and gender.....	190
5.1.3.3 Level of education.....	191
5.2 Formality and the <i>Siarad</i> corpus.....	192
5.3 The mutation variables.....	194
5.4 Finding mutation environments.....	195
5.4.1 Dealing with dropped triggers.....	196
5.5 Classifying mutations.....	199
5.5.1 Classifying binary variables.....	199
5.5.2 Classifying intermediate variants for index-scored variables.....	200
5.5.3 Classifying tokens for (ei.m) and (ei.f).....	202
5.5.4 Classifying undetermined variables.....	205
5.6 Calculating scores for mutation variables	205
5.6.1 Calculating percentage scores for binary variables.....	206
5.6.2 Calculating index scores for non-binary variables.....	207

5.7 Identifying undetermined tokens.....	210
5.7.1 Ambiguous language classification.....	210
5.7.2 Words that are never mutated	211
5.7.3 ‘Fossilized’ formulaic constructions.....	212
5.7.4 ‘Relexified’ dialectal variants.....	216
5.7.5 Dropped items which blocked mutation.....	218
5.7.6 Unknown antecedent possessors for (ei.m) and (ei.f).....	218
5.7.7 Unclear audio.....	221
5.7.8 Affricate mutations.....	221
5.8 RBRUL analysis.....	222
5.8.1 Age.....	222
5.8.2 Gender.....	223
5.8.3 Level of education.....	223
5.8.4 Initial consonant.....	223
5.8.5 Place names.....	224
5.9 Variables not considered	225
5.10 Code-switching analysis.....	226
5.10.1 Coding and tagging code-switches.....	226
5.10.2 Calculating group scores for code-switching.....	227
5.10.3 Statistical analysis of code-switching.....	227
5.11 Summary.....	228
6. <i>Siarad Corpus Analysis Results</i>	229
6.1 Mutation token frequencies.....	229
6.2 Internal factors.....	230
6.2.1 Place names.....	231

6.2.1.1	Place names - (o).....	231
6.2.1.2	Place names - (am).....	233
6.2.1.3	Place names - (yn).....	234
6.2.1.4	Place names – summary.....	235
6.2.1.5	Frequency of place names.....	235
6.2.2	Initial consonant.....	238
6.2.2.1	Initial consonant – (fy).....	238
6.2.2.2	Initial consonant – (am).....	239
6.2.2.3	Initial consonant – (o).....	240
6.2.2.4	Initial consonant – (yn).....	241
6.2.2.5	Initial consonant – (a).....	242
6.2.2.6	Initial consonant – (ei.m).....	242
6.2.2.7	Initial consonant – (ei.f).....	243
6.2.2.8	Initial consonant – summary.....	243
6.3	Language external variables.....	244
6.3.1	Age.....	244
6.3.1.1	Age - (fy).....	244
6.3.1.2	Age - (am).....	244
6.3.1.3	Age - (o).....	245
6.3.1.4	Age - (yn).....	245
6.3.1.5	Age - (a).....	246
6.3.1.6	Age - (ei.m).....	246
6.3.1.7	Age - (ei.f).....	246
6.3.1.8	Age - (â/gyda).....	247
6.3.1.9	Age – summary.....	247

6.3.2	Gender.....	247
6.3.2.1	Gender (fy).....	248
6.3.2.2	Gender (am).....	248
6.3.2.3	Gender (o).....	249
6.3.2.4	Gender (a).....	249
6.3.2.5	Gender (ei.m).....	250
6.3.2.6	Gender (ei.f).....	250
6.3.2.7	Gender – Summary.....	251
6.3.3	Level of Education.....	251
6.3.3.1	Level of education (fy).....	251
6.3.3.2	Level of education (am).....	252
6.3.3.3	Level of education (o).....	253
6.3.3.4	Level of education (yn).....	253
6.3.3.5	Level of education (a).....	254
6.3.3.6	Level of education (ei.m).....	254
6.3.3.7	Level of education (ei.f).....	255
6.3.3.8	Level of education – summary.....	256
6.3.4	Attitudes.....	256
6.3.4.1	Attitudes (fy).....	256
6.3.4.2	Attitudes (am).....	257
6.3.4.3	Attitudes (o).....	257
6.3.4.4	Attitudes (yn).....	258
6.3.4.5	Attitudes (a).....	258
6.3.4.6	Attitudes (ei.m).....	259
6.3.4.7	Attitudes (ei.f).....	259

6.3.4.8 Attitudes – summary.....	260
6.4 Code-switching results	260
6.4.1 Code-switching – age.....	260
6.4.2 Code-switching – gender.....	261
6.4.3 Code-switching – level of education.....	261
6.4.4 Code-switching – attitudes.....	262
6.5 Summary of RBRUL results.....	262
6.6 Apparent Time analysis.....	263
6.6.1 Apparent Time – (fy).....	264
6.6.2 Apparent Time - (am)	265
6.6.3 Apparent Time – (o).....	266
6.6.4 Apparent Time – (yn).....	267
6.6.4.1 Distribution of mutation type following (yn).....	270
6.6.5 Apparent Time – (a).....	270
6.6.6 Apparent Time – (ei.m).....	271
6.6.7 Apparent Time – (ei.f).....	273
6.6.7.1 Distribution of mutation type following (ei.f).....	274
6.6.8 Apparent Time – (â/gyda).....	275
6.6.9 Apparent time – Code-Switching.....	276
6.7 Summary of Results.....	277
6.8 Discussion of Inter-speaker variation	277
6.8.1 The aspirate mutations.....	278
6.8.2 (ei.m).....	279
6.8.3 (ei.f).....	280
6.8.4 (o) and (am).....	281

6.8.5	(fy).....	282
6.8.6	(yn).....	283
6.8.7	Level of education and gender.....	285
6.8.8	Inter-speaker variation – summary.....	286
7.	Radio Corpus Methodology	288
7.1	Selecting programmes.....	288
7.1.1	<i>Post Prynhawn</i>	289
7.1.2	<i>Dewi Llwyd Ar Fore Sul</i>	291
7.1.3	<i>Geth a Ger</i>	293
7.1.4	<i>Tudur Owen</i>	295
7.1.5	Summary.....	297
7.2	Data collection.....	299
7.3	Transcription.....	301
7.3.1	Tier structure.....	301
7.3.2	Mutation	301
7.3.3	Transcribing dropped mutation triggers.....	302
7.3.4	Language tagging.....	303
7.4	Calculating scores.....	304
7.4.1	Calculating scores – code-switching.....	304
7.4.2	Calculating scores – mutation.....	305
7.5	Categorizing programmes.....	305
7.5.1	The style axis.....	306
7.5.2	Analysing channel cues.....	308
7.5.2.1	Laughter.....	308
7.5.2.2	Overlapping speech.....	309

7.5.3	The social axis.....	310
7.5.3.1	The social axis – age.....	310
7.5.3.2	The social axis – education.....	311
7.6	Summary.....	311
8.	Radio Corpus Results	312
8.1	Categorizing programmes by level of formality.....	312
8.1.1	Laughter.....	312
8.1.2	Overlapping speech.....	313
8.1.3	Formality – summary.....	314
8.2	The style axis.....	315
8.2.1	<i>Post Prynhawn</i> – formality.....	316
8.2.2	<i>Dewi Llwyd Ar Fore Sul</i> – formality.....	317
8.2.3	<i>Geth a Ger</i> – formality.....	319
8.2.4	<i>Tudur Owen</i> – formality.....	320
8.2.5	<i>Summary</i> – formality.....	320
8.3	The social axis.....	321
8.3.1	<i>Dewi Llwyd Ar Fore Sul</i> – The social axis.....	321
8.3.2	<i>Post Prynhawn</i> – The social axis.....	323
8.3.3	<i>Tudur Owen</i> – The social axis.....	325
8.3.4	<i>Geth a Ger</i> – The social axis.....	325
8.3.5	<i>Summary</i> – social axis.....	326
8.4	Results for linguistic variables.....	326
8.4.1	Results – (am).....	326
8.4.2	Results – (o).....	327
8.4.3	Results – (yn).....	330

8.4.4	Results – (fy).....	331
8.4.5	Results – (a).....	333
8.4.6	Results – (â/gyda).....	334
8.4.7	Results – (ei.m).....	335
8.4.8	Results – (ei.f).....	336
8.4.9	Results – code-switching.....	337
8.5	Results – summary.....	341
9.	Qualitative Analysis	343
9.1	Retracing.....	343
9.1.1	Retracing - <i>Geth a Ger</i>	344
9.1.2	Retracing - <i>Tudur Owen</i>	348
9.1.3	Retracing - <i>Dewi Llwyd AFS</i>	350
9.1.4	Retracing – summary.....	351
9.2	Flagging.....	351
9.2.1	Flagging - <i>Post Prynawn</i>	352
9.2.2	Flagging – <i>Dewi Llwyd AFS</i>	353
9.2.3	Flagging – <i>Geth a Ger</i>	355
9.2.4	Flagging – <i>Tudur Owen</i>	356
9.2.5	Flagging – summary.....	358
9.3	Reinforcing comprehension.....	360
9.3.1	Reinforcing comprehension – <i>Post Prynawn</i>	361
9.3.2	Reinforcing comprehension – <i>Dewi Llwyd AFS</i>	362
9.3.3	Reinforcing comprehension – <i>Geth a Ger</i>	363
9.3.4	Reinforcing comprehension – summary.....	364
9.4	Direct speech.....	365

9.5 Humour.....	368
9.5.1 ‘Seriousness’.....	374
9.6 Summary.....	376
10. Discussion.....	378
10.1 Classifying variables.....	378
10.1.1 Code-switching as a marker.....	378
10.1.2 (a) and (â/gyda).....	379
10.1.3 (fy) as an indicator.....	380
10.1.4 (ei.m) (am) and (o).....	382
10.1.5 (yn) – an ambiguous case.....	383
10.1.6 (ei.f) Language internal factors.....	385
10.1.7 Summary.....	385
10.2 Stylistic analysis.....	386
10.2.1 Applying variationist theory to the data.....	387
10.2.2 Applying speaker-centred theory to the data.....	386
10.2.3 Audience design and referee design on <i>Radio Cymru</i>	390
11. Conclusion.....	392
11.1 Drawbacks and advantages of the method.....	392
11.2 Future research.....	394
11.3 Implications of the project.....	395

List of Figures

Figure 1 The distribution of gender among participants in the *Siarad* corpus

Figure 2 The age demographics of *Siarad* corpus participants

Figure 3 The number of *Siarad* participants by level of education

Figure 4 Frequency of tokens of mutation for each variable in the *Siarad* corpus

Figure 5 The proportion of tokens for each variable that were followed by a place name

Figure 6 Standard mutation of (o) across three conditions, with, without or only place names

Figure 7 Standard mutation of (am) across three conditions, with, without or only place names

Figure 8 Standard mutation of (yn) across three conditions, with, without or only place names

Figure 9 Standard mutation and frequency of place names in the *Siarad* corpus

Figure 10 Standard mutation in the *Siarad* corpus following (fy) by age group

Figure 11 Standard mutation following (am) by age group

Figure 12 Standard mutation following (o) both with and without place names

Figure 13 Standard mutation following (yn) for the ‘including place names’ condition by age

Figure 14 Standard mutation following (yn) for the ‘without place names’ condition by age

Figure 15 Distribution of mutation types following (yn) in the *Siarad* corpus

Figure 16 Standard mutation in the *Siarad* corpus following (a) by age group

Figure 17 Standard mutation in the *Siarad* corpus following (ei.m) by age group

Figure 18 Standard mutation in the *Siarad* corpus following (ei.f) by age group

Figure 19 Distribution of mutation type following (ei.f) in the *Siarad* corpus

Figure 20 Standard mutation in the *Siarad* corpus following (â/gyda) by age group

Figure 21 Average proportions of code-switching by age group

Figure 22 The average frequency of laughter in each programme

Figure 23 The mean frequency of lazy overlaps for each programme

Figure 24 The style axis distribution of each programme by level of formality

Figure 25 The social axis distribution of programmes according to prestige

Figure 26 Distribution of standard tokens of (am) by radio programme

Figure 27 Standard mutation of (o) without place names by radio programme

Figure 28 Standard mutation of (o) with only place names by radio programme

Figure 29 Standard mutation of (yn) by radio programme

Figure 30 Standard mutation of (fy) by radio programme

Figure 31 Standard mutation of (a) by radio programme

Figure 32 Standard mutation of (â/gyda) by radio programme

Figure 33 Standard mutation of (ei.m) by radio programme

Figure 34 Standard mutation of (ei.f) by radio programme

Figure 35 Proportion of code-switching by radio programme

Figure 36 Average of code-switching means for all individual speakers averaged across all episodes

Figure 37 Code-switching means for each individual presenter

Figure 38 Each presenter's use of code-switching compared with their age group

Figure 39 Each presenter's use of (fy) compared with their age group

List of Tables

Table 1 Distribution of features between standard and vernacular north western Welsh

Table 2 Hierarchical ranking of the difficulty of acquisition of linguistic features

Table 3 Consonant changes in soft mutation

Table 4 Consonant changes in nasal mutation

Table 5 Consonant changes in aspirate mutation

Table 6 Selected mutation triggers for the SM system

Table 7 Selected mutation triggers for the NM system

Table 8 Selected mutation triggers for the AM system

Table 9 Variables analysed in P.W Thomas' (1984) study of mutation

Table 10 Variables analysed in Ball's (1984) study of mutation

Table 11 Standard mutation of percentage-scored variables in Ball (1984)

Table 12 Standard mutation of index-scored variables in Ball (1984)

Table 13 Mean style shifting for three acculturation groups in Ball (1984)

Table 14 Mean percentage standard use of SM triggers in Ball (1984: 284)

Table 15 Percentage standard use of (fy) by age in free conversation in Hatton (1988: 248)

Table 16 Percentage standard use of (yn) by age in Hatton (1988: 253)

Table 17 Percentage standard use of (yn) by background in Hatton (1988: 253)

Table 18 Standard use of AM mutations by programme in (Ball et al 1988: 188)

Table 19 Mean standard use of all AM variables in a community based study and a radio study (Ball et al 1988: 189)

Table 20 Variables used in Jones' (1998) study

Table 21 Code-switching patterns in different styles (Zentella 1990: 83)

Table 22 General information about the mutation variables selected for the study

Table 23 Constructions with mutation triggers which are variably dropped

- Table 24** The number and types of mutation that each variable undergoes
- Table 25** A list of formulaic constructions with ‘fossilized’ mutation triggers
- Table 26** A list of words which have dialectal variants
- Table 27** The range of consonants mutated by each mutation trigger
- Table 28** Results of analysis of factor place names for (o)
- Table 29** Results of analysis of factor place names for (am)
- Table 30** Results of analysis of factor place names for (yn)
- Table 31** Results of analysis of place name frequency factor for (yn)
- Table 32** Results of RBRUL analysis of mean frequency of place names by age group
- Table 33** Results of analysis of variable (fy) and initial consonant
- Table 34** Results of analysis of variable (am) and initial consonant
- Table 35** Results of analysis of variable (o) and initial consonant
- Table 36** Results of analysis of variable (yn) and initial consonant
- Table 37** Results of analysis of initial consonant (a) and initial consonant
- Table 38** Results of analysis of initial consonant (ei.m) and initial consonant
- Table 39** Results of analysis of initial consonant (a) and initial consonant
- Table 40** Results of RBRUL analysis of standard mutation after (fy) and age
- Table 41** Results of RBRUL analysis of age and standard use of (am)
- Table 42** Results for (o) without place-names analysis with age
- Table 43** Result of one-way analysis of age and (yn) without place-names
- Table 44** Results of one-way analysis of age and (a)
- Table 45** Results of RBRUL analysis of age and (ei.m)
- Table 46** Results of RBRUL analysis of age and (ei.f)
- Table 47** Results of the RBRUL analysis of (fy) and gender
- Table 48** Results of the RBRUL analysis of (am) and gender

- Table 49** Results of the RBRUL analysis of (yn) and gender
- Table 50** Results of the RBRUL analysis of (a) and gender
- Table 51** Results of the RBRUL analysis of (ei.m) and gender
- Table 52** Results of the RBRUL analysis of (ei.f) and gender
- Table 53** Standard use of (fy) and level of education
- Table 54** Standard use of (am) and level of education
- Table 55** Standard use of (o) and level of education
- Table 56** Standard use of (yn) and level of education
- Table 57** Standard use of (a) and level of education
- Table 59** Standard use of (ei.m) and level of education
- Table 60** Standard use of (ei.f) and level of education
- Table 61** Standard use of (fy) and attitudes
- Table 62** Standard use of (am) and attitudes
- Table 63** Standard use of (o) and attitudes
- Table 64** Standard use of (yn) and attitudes
- Table 65** Standard use of (a) and attitudes
- Table 66** Standard use of (ei.m) and attitudes
- Table 67** Standard use of (ei.f) and attitudes
- Table 68** Results of RBRUL analysis of age and code-switching
- Table 69** Results of RBRUL analysis of gender and code-switching
- Table 70** Results of RBRUL analysis of level of education and code-switching
- Table 71** Results of RBRUL analysis of attitudes and code-switching
- Table 72** Summary of the statistical analysis of all variables
- Table 73** Number of SM tokens and radical tokens in *Siarad* corpus following (fy) by age

Table 74 Number of SM tokens and radical tokens in *Siarad* corpus following (am) by age group (including place names)

Table 75 Number of SM tokens and radical tokens in *Siarad* corpus following (o) by age group (including place names)

Table 76 Number of SM, NM and radical tokens in *Siarad* corpus following (yn) by age group

Table 77 Number of SM, NM and radical tokens in *Siarad* corpus following (yn) by age group

Table 78 Number of SM tokens and radical tokens in *Siarad* corpus following (a) by age group

Table 79 Number of SM tokens and radical tokens in *Siarad* corpus following (ei.m) by age group

Table 80 Number of SM tokens and radical tokens in *Siarad* corpus following (ei.f) by age group

Table 81 Number of SM tokens and radical tokens in *Siarad* corpus following (a) by age group

Table 82 Summary of the four *Radio Cymru* programmes

Table 83 The number of words and length of recordings (h/m/s) of each programme post-editing

Table 84 Variables used to measure formality in the radio programmes

Table 85 Number of standard and non-standard tokens of (am) in each programme

Table 86 Number of standard and non-standard tokens of (o) ‘without place names’ in each programme

Table 87 Number of standard and non-standard tokens of (o) ‘only place names’ in each programme

Table 88 Number of standard, intermediate SM and non-standard tokens of (yn) in each programme

Table 89 Number of standard and non-standard tokens of (fy) in each programme

Table 90 Number of standard and non-standard tokens of (a) in each programme

Table 91 Number of standard and non-standard tokens of (â/gyda) in each programme

Table 92 Number of standard and non-standard tokens of (ei.m) in each programme

Table 93 The number of standard, intermediate SM and non-standard tokens of (ei.f) in each programme

Table 94 The total number of Welsh and English words in each programme

Table 95 A comparison of the frequency and direction of the retracing strategy across all four programmes

Table 96 Frequency and language of flagging across all programme

Table 97 Flagged words and relative frequencies in the CEG corpus

Table 98 Frequency and language of words reinforced by each radio programme

Table 99 A comparison of the frequency of direct speech translated into Welsh or maintained in English

1. Introduction

The study of style concerns the exploration of socially meaningful variation within, and between, the speech of individuals. However, the concept crucially spans the individual experience, social groups, and broader society. As Irvine (2001) explains “‘style’ crucially involves distinctiveness; though it may concern the individual, it does so only within a social framework [...]; it thus depends on social evaluation and, perhaps aesthetics; and it interacts with ideologized representation” (Irvine 2001: 21). It is a widely accepted sociolinguistic verity that “all speakers and social groups have sociolinguistic repertoires” (Rickford and Eckert 2001: 1) and that inferring the stylistic meanings indexed by linguistic features forms part of speakers’ *communicative competence* (Dell Hymes 1972). In a sense, style empowers speakers by providing them with a set of communicative options and creative possibilities. However, style also entraps and constrains speakers, by imbuing certain acquired linguistic varieties with social stigma, which can be difficult for individuals who speak these codes to avoid (Gumperz 1982). One of the central tensions within the field of sociolinguistics is between researchers who have preferred to explore style as a potentially liberating conduit of self-expression (Coupland 2007; Eckert 2001) and others who emphasize that speakers are constrained to particular styles of speech by macro-level social processes (Labov 1972, Finegan and Biber 1994).

The seminal model of style in sociolinguistics has been the variationist paradigm (Labov 1966), which has shown that stylistic variation is tightly interwoven with social variation (Romaine 1980, Bell 1984), and can be legitimately considered a reflection of social processes, including both power disparities (Labov 1972) and linguistic vitality (Dorian 1994). In terms of power, Bourdieu (1991) described linguistic features as constituting *cultural capital*, with certain variants being imbued with prestige due to their association with powerful socioeconomic elites. Bourdieu argued that access to social mobility for individuals in capitalist societies can

often be dependent upon successful harnessing of these prestigious linguistic forms. Prestigious linguistic forms are typically connected with the notion of *standard language* varieties. Standard languages are prestigious codes that normally share certain characteristics: they are codified and static varieties, they are associated with historically entrenched elites, and they are often associated with prestigious written registers (Jones 1998). Standard languages are legitimized by *standard language ideologies* (SLIs), which influence speakers to evaluate deviations from the standard norm negatively (Lippi-Green 1997). Aside from these overt power relations, Dorian's (1994) work on the dying linguistic variety known as East Sutherland Gaelic (ESG) explored the relationship between sociolinguistic vitality and style. Her study described the way in which processes associated with language death, such as domain attrition and the narrowing range of available interlocutors, can atrophy the set of stylistic options available to minority language speakers, limiting speakers' possibilities of linguistic self-expression, thus rendering the language yet more redundant and catalysing the socially driven process of linguistic decline.

In this thesis, I intend to explore style within a sociolinguistic context which received relatively little scholarly attention: that of Welsh-English bilinguals in Wales. To do this, I will be drawing upon a wealth of recent sociolinguistic studies in order to analyse the connection between individual styles, social frameworks and ideological representations in the Welsh language setting (Irvine 2001). However, before I go on to discuss these issues, I will first summarize the sociolinguistic background of Welsh-speaking Wales.

1.1 Style in the Welsh context

There are several factors which emerge as particularly important for understanding the Welsh language context. Here I will list some of the main details which need to be taken into account:

1. The first is that Welsh is a minority language in Wales, and is only spoken by around 19% (562,000) of a population of 3.1 million (ONS 2011). Proportions of Welsh speakers are declining in heartland areas, but have increased recently in some urban and anglicized areas due to the influence of the education system, which promotes Welsh (Jones 1998, Robert 2009).
2. Wales is officially a bilingual country, with Welsh “a language with official status” (Welsh Government 2011: 1) in Wales, and English having *de facto* official status (J. Davies 2014). All Welsh speakers are bilingual, also being able to speak English (ONS 2011).
3. As is implied above, the government promotes Welsh in accordance with the Welsh Language Measure of 2011, with the aim of realizing “the principle that persons in Wales should be able to live their lives through the medium of the Welsh language if they choose to do so” (Welsh Government 2011: 3). The creation of the Measure involved the appointment of a language commissioner with powers to enforce language standards in the public sector, the voluntary sector and among private sector companies which provide utilities to the public.
4. For those who do speak Welsh, it is often the case that they do not use the language in all functions of their daily lives. In fact, in making stylistic choices, Welsh speakers may draw from three broad sources of linguistic forms – vernacular (or dialect) Welsh, standard Welsh or English (A.R Thomas 1987). This situation has been described as being *diglossic* (Ferguson 2007), involving both *bilingual diglossia* between Welsh and English (E. Williams 2009) and *dialectal diglossia* between vernacular and standard/literary Welsh (Fife 1986).
5. Welsh society has often been described as inherently heterogeneous (Coupland and Ball 1989). Welsh speaking communities are no exception to this rule, with Welsh speaking

bilinguals living in strongly contrastive sociological contexts. These include rural and urban areas, as well as mainly Welsh speaking and more anglicized areas.

6. Welsh speakers have also been described as tending to fit a particular socioeconomic profile, which varies by area – often working class in North Wales and typically middle class in South Wales, and there are very few urban communities in which Welsh speakers are numerous enough to constitute the whole social strata (Giggs and Pattie 1991). These kinds of factors have led most researchers to argue that analyses based on socioeconomic class are not feasible in the Welsh-speaking context (P.W. Thomas 1984, Ball 1984, R.O. Jones 1976), whereas class has been central in Anglophone variationist studies on style (e.g. Labov 1966).
7. Welsh is undergoing a process of linguistic convergence triggered by language contact with English. This is claimed to involve a general convergence towards English across a range of features including explicit transfer of surface features such as lexical items and morphology (A.R. Thomas 1982), as well as deeper changes to syntax (D. Philipps 2008) and the loss of traditional Welsh features (Jones 1998). However, the extent and scale of this change has been debated (Davies and Deuchar 2010).
8. Associated with the above factor is a broadly negative cultural attitude towards elements of English transfer in Welsh (Robert 2011). This has been described as a reaction to the English language's influence upon Welsh, and the fears that this has elicited for the future of the language (P. Davies 2010). This tendency has led some Welsh speakers to evaluate individuals who use elements of English transfer negatively (Robert 2009), as well as to activist campaigns which have targeted the quality of the Welsh used in the media (see below).
9. As a result of this, Robert (2011) argues that Standard Language Ideology in Wales largely centres on condemnation of, and political pressure against, English transfer -

particularly in the public sphere. This can be seen most clearly in the overt prescriptions against lexical borrowing in standard Welsh Grammars, such as P.W. Thomas' (1996) widely used *Gramadeg y Gymraeg*.

The factors above provide a broad picture of the Welsh sociolinguistic and political context. Understanding this fully is important because it provides the background against which Welsh speakers make stylistic decisions, and from which they develop their notions of *communicative competence*. In the next few sub-sections, I will describe these factors in more detail, before moving on to outline my study design.

1.1.1 The Social distribution of Welsh

Giggs and Pattie (1991) have argued that Wales cannot be easily described as a single homogenous entity: “contemporary Welsh society is plural rather than homogenous, divided rather than cohesive” (Giggs and Pattie 1991: 28). In the south east, Welsh speakers tend to be thinly scattered amongst monolingual English speaking majorities, while in more rural areas in the west and the north there are communities such as Caernarfon and Pwllheli in which the majority of inhabitants are bilingual (Aitchison and Carter 2000, E. Williams 2009). Coupland et al (2005) thus describe Gwynedd, the North Western focus of this study, as:

“[...] distinctive for its high overall percentage [...] of Welsh speakers and its wider ‘Welsh heartland’ character. It is rural, mountainous and generally sparsely populated. It is one of the areas of Wales (also including Ceredigion in south-west Wales) foregrounded in current political antagonisms over ‘English incomers’, as they are identified by the language and cultural pressure group Cymuned” (Coupland et al 2005: 7).

With its heartland profile, Gwynedd has the highest proportion of Welsh speakers over the age of three who are able to speak, read and write in Welsh, 56%, in comparison with 7.8% for

Blaenau Gwent (the ‘least Welsh speaking’ area of Wales) in the South East (ONS 2011). This filters through to the makeup of the education system in Gwynedd, where Coupland et al (2005) report that 93% of secondary schools in 2005 were Welsh medium, which they contrast with the existence of only a single Welsh medium school in the much more densely populated Vale of Glamorgan at that time.

Another level of heterogeneity can be seen in terms of the class distribution of Welsh speakers in different areas of Wales. Aitchison and Carter (1987) noted a reversal of Hechter’s (1978) cultural division of labour (whereby minority social groups tend to appear in lower social classes in most societies) in the capital city of Cardiff, where Welsh speakers from other areas of Wales were moving in large numbers in pursuit of new Welsh-essential jobs, created by the historic reversal of domain attrition which took place during the 20th century (see section 1.1.2.1 below). As a consequence of this transformation, many researchers argue that Welsh speakers in the south of Wales are more likely to be from higher social classes than in North Wales, and that anglicized communities such as Cardiff by contrast normally have Welsh speakers distributed unequally among different social classes, but being mainly middle-class (Aitchison and Carter 2000). As Giggs and Pattie found in their survey study:

“Welsh-speakers collectively were an important component of the elite service class chiefly in metropolitan South Wales and its industrial hinterland and secondarily in north-east Wales. In contrast, in the Welsh-speaking rural heartland of Wales (i.e. Gwynedd, Dyfed and Powys, Welsh speakers were actually under-represented in the service class” (Giggs and Pattie 1991: 27).

While the majority of mainstream variationist analyses have focused on class as the primary independent variable (Meyerhoff 2011), most sociolinguistic studies looking at Welsh have made the case that class stratification is not a factor shaping the distribution of linguistic

features, and that it can be disregarded in sociolinguistic analyses (Ball 1984; P.W. Thomas 1984, R.O. Jones 1976). The rationale behind this argument is that, as mentioned above, communities where Welsh speakers live are generally linguistically heterogeneous, and there are very few communities where it can be said that the entire social strata of a Welsh town participate together as a cohesive speech community (Ball 1984). It has also been argued that rural communities differ fundamentally from urban towns and cities in such a way that class does not manifest in the same manner (P.W. Thomas 1984). Analyses of mainly Welsh speaking and rural communities have thus arranged speakers into alternative social groups based around broad notions of prestige. For example, Jenkins' (1962) analysis of a rural community during the Fifties uses the term *buchedd* ("way of life") to categorize speakers into two categories, A and B. In the first category were speakers who were positively oriented towards the value of education for themselves and their children, who attended the chapel and observed Sabbath, and who held high status jobs. The second category included speakers who were less positively oriented towards education, who frequented the tavern rather than the chapel, and who held lower status jobs. Although these categories seem to overlap with some aspects of class, Jenkins argued that a member of such a community could have a low status job and still belong to the high prestige group A due to other factors, such as attitudes towards education, religious lifestyle, and so on.

More recently, some researchers have argued against the idea that class is irrelevant to the Welsh sociolinguistic context. For example, G. Williams (1987) predicted that social changes in Wales were leading to stronger class-based distinctions amongst Welsh speakers. The role of transnational institutions, such as Welsh medium schools and the media, was expected to be crucial in this process, which would disseminate a standard non-localized variety of Welsh nationally and place a minority of Welsh speakers in positions of relative power and prestige. Williams suggested that these changes could lead to a Welsh defined by class differences, as a

result of the subsequent differing levels of access to prestigious varieties among the population: “I would claim that we are beginning to see the emergence of a class variety of Welsh which derives not from the conscious process of language standardization but from the process of the extension of the prestige of the Welsh language” (G. Williams 1987: 96).

A compelling recent study by Madoc-Jones et al (2013) has also argued for the consideration of class among Welsh speakers. The authors focus on perceptions of Welsh identity among two highly socially divergent groups of Welsh speakers; service users in the criminal justice system (i.e. offenders who had served in prison within the last five years) and service providers, who were successful, highly placed executives. The service users made frequent allusions towards their own marginality outside of mainstream conceptions of Welshness, and often presented themselves as being unable to participate in more prestigious domains of Welsh. Madoc-Jones et al’s study provides powerful evidence that class may indeed be an important ideological frame for Welsh speakers. However, thus far there has been little real empirical or linguistic evidence to prove whether the kind of Welsh that speakers use is stratified socioeconomically. In terms of attitudinal data, a recent matched-guise test study by Robert (2009) suggested that one speaker from an L2 south-eastern background was well evaluated because of her use of what judges described as *media Welsh*, a variety that reflects a relatively non-regional professional standard Welsh increasingly used in the media and other transnational institutions. However, in the absence of more research specifically addressing this question, the relationship between class and Welsh is unclear. Because of this, one of the main conceptual explanatory devices used in mainstream analyses of style remains unexplored in the Welsh context.

1.1.2 Diglossia in Wales

Ball et al (1988) sum up the Welsh linguistic context in the following way: “(1) It is diglossic and (2) it is bilingual” (Ball et al 1988: 192). Many other researchers have also described the

Welsh linguistic situation as diglossic (Coupland and Ball 1989; Fife 1986; Madoc-Jones et al 2013). Diglossia refers to a situation in which a speech community uses two highly divergent codes in a complimentary fashion on a societal level:

“Diglossia is a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation.” (Ferguson 2007: 42)

The code retained for high functions is often referred to as the H code, while the code used in low functions is referred to as the L code. Ferguson’s distinction between H varieties which come from “an earlier period” or “another speech community” was developed by later researchers into the concepts of dialectal and bilingual diglossia (Fishman 1980). Examples of dialectal diglossia include areas in the Arabic-speaking world where classical Arabic is often used for high functions, and where various regional varieties are used for L functions. Examples of contexts of bilingual diglossia include the linguistic situation of Paraguay described in Fishman (1980), where the native language Guarani is the L code, and the colonial language Spanish is the H code. Diglossia in Wales has been described both in terms of bilingual diglossia, with English as the H code and Welsh as the L code (E. Williams 2009), and in terms of dialectal diglossia, with Literary Welsh taking the role of the H code and the various regional vernacular dialects of Welsh taking the role of the L code (Fife 1986; Coupland and Ball 1989).

1.1.2.1 Bilingual Diglossia

The roots of bilingual diglossia in Wales lie in the fact that for a lengthy historical period following the annexation of Welsh-speaking areas to the English state, Welsh was not used for high functions, such as in the judicial and education systems, but was generally restricted to vernacular and religious contexts (E. Williams 2009). This situation persisted until relatively recently, when alarm at the decline in the numbers of Welsh speakers triggered a wave of nationalist social activism (Aitchison and Carter 2000). One of the most significant events for this process of revitalization was nationalist leader Saunders Lewis' influential radio broadcast *Tynged yr Iaith* ("The Fate of the Language") in 1962, where Lewis called for the complete expansion of Welsh into the public domain, encouraging listeners to "make it impossible to conduct local or central government business in Wales without the Welsh language" (Davies 2014: 120). This led directly to the creation of the nationalist campaign group *Cymdeithas yr Iaith* ("The Language Society") which carried out a strategy of non-violent social action over the subsequent decades, including damaging monolingual road signs and invading courtrooms. This strategy was hugely successful and led to a significant reversal of domain attrition in the public sphere (Aitchison and Carter 2000).

Welsh entered the domain of the media with the first monolingual Welsh language radio station *Radio Cymru* in 1979 (Smith 2000) and also gained a Welsh language television station in 1982, following a lengthy period of campaigning (Jones 1998). As a direct result of the road sign damaging protest, the Bowen Report of 1972 mandated that all public signs in Wales be bilingual (J. Davies 2014). The Welsh Language Act of 1993 created the Welsh Language Board, a commission whose mandate was to oversee public bodies' adherence to the legislation, as well as to develop strategies of promoting the use of Welsh institutionally and in public in general. Although the Welsh Language Board was given powers to investigate situations in which speakers' rights to use Welsh were violated, some criticized these powers

as being largely symbolic, as they could not fine bodies or compensate individuals (C. Williams 2011: 49). In 2011 the National Assembly passed a new Language Measure which abolished the Welsh Language Board, replacing it with a language commissioner. The language commissioner's role is to oversee the adherence of public bodies to the standards set forth in the measure, using various means to enforce compliance, such as: publicising the failures of a particular body, drawing up action plans to prevent repetitions, or imposing civil penalties (Welsh Government 2011: 40).

1.1.2.2 Dialectal diglossia and standard Welsh

Since the process of reversing domain attrition, and the expansion of Welsh into new domains, was begun in the mid-twentieth century, one of the primary questions has been: what variety of Welsh is most appropriate for use in public and formal domains? One answer has been a variety known as *standard* or *literary Welsh*. Standard Welsh has its roots in the Welsh Bible translation of 1588, commissioned by Elizabeth I to shore up Protestantism in Wales (Roberts 2011). This in turn is claimed to be based on the translator William Morgan's fusion of dialectal varieties used in earlier poetry (D.G. Jones 1988). From this period up until the mid-twentieth century the primary domain of standard Welsh was religion (Jenkins 2000). The majority of Welsh speakers by the eighteenth century are thought to have been able to understand the standard Welsh of the Bible, due to the influence of travelling schools and religious services (Roberts 2011). Coupland and Ball (1989) claim that despite this exposure to standard Welsh through religious services and schools, the majority of speakers would only have had passive competence in the variety: "[...] many speakers only had productive access to the vernacular spoken forms, not speaking, reading or writing this standard Welsh (though probably able to understand it when read aloud)" (Coupland and Ball 1989: 11).

The practice of using standard Welsh for writing and high functions has persisted to the current day, raising questions about the appropriateness of the form for modern speakers (Robert 2011). It should be noted that the gap between standard Welsh and the dialects is greater than that between Standard English and most British or American English dialects, with standard Welsh possessing, for example, a full system of inflected verbs which is only infrequently used in the vernacular (Fife 1986). However, the importance of having a standard variety has been persistently defended, mainly on the basis that a national variety is needed due to relatively strong differences between the four main area dialects of Wales (Coupland and Ball 1989).

1.1.3 Differences between Standard and Vernacular Welsh

Standard Welsh has many features which differentiate it from colloquial Welsh. Perhaps one of the most well-attested and important differences is that the colloquial varieties of all dialects show a preference for periphrastic verb constructions, while standard Welsh has both periphrastic forms and a rich inflected verb system (Fife 1986)¹. Other features in complementary distribution between the two codes include, amongst others, the forms of verb paradigms, the use of the negative particle *ni(d)* in formal Welsh versus *ddim* in colloquial Welsh (Coupland and Ball 1989), reflexive verbs, and more. As an example, table 1 provides a selection of Welsh linguistic features that are in complimentary distribution between the vernacular of north-west Wales and standard Welsh:

Linguistic Feature	Standard Welsh	North Western Vernacular
Negation - nid (H) ddim (L)	nid wyf yn mynd NEG be.1S.PRES PRT go.NONFIN ("I'm not going")	dw i (ddi)m yn mynd be.1S.PRES PRON.1S PRT go.NONFIN

¹ Although there may well be dialectal differences.

Relative particle “a” (H)	beth a glywaist? what REL hear.2S.PAST (“ <i>what did you hear?</i> ”)	be glywaist di? what hear.2S.PAST PRON.2S
Synthetic forms of the pluperfect (H)	buasai ef yno o’r blaen be.3S.PLU PRON.3S from DET front (“ <i>he had been there before</i> ”)	oedd o wedi bod yna o’r blaen be.3S.IMP PRON.3S PRT.PAST be.NONFIN there from DET front
Vingesimal (H) /decimal (L) counting system	deg, un ar ddeg, deuddeg, tair ar ddeg, pedair ar ddeg, pymtheg (“ <i>10, 11, 12, 13, 14, 15</i> ”)	deg, un deg un, un deg dau, un deg tri, un deg pedwar, un deg pump
Synthetic (H) / periphrastic (L) clause construction	cerdd y dyn walk.3S.PRES DET man (“ <i>the man is walking</i> ”)	mae ’r dyn yn cerdded be.3S.PRES man PRT walk.NONFIN

Table 1 Distribution of features between standard and vernacular North Western Welsh

1.1.3.1 Standard Welsh and English transfer

Alan R. Thomas (1982) claimed that language contact between English and Welsh was causing rapid linguistic change in Welsh, with convergence towards English occurring across multiple linguistic features. One of the most salient aspects of this process was “lexical borrowing at an exaggerated rate, regardless of the availability of an indigenous term or not” (A. R. Thomas 1982: 210), but it also involved the importation of English morphology and the calquing of English constructions such as phrasal verbs (A. R. Thomas 1987). Thomas also described some changes which are less explicitly rooted in the English language as also ultimately being caused by language contact. These include the movement towards periphrastic constructions, facilitated according to Thomas by the development of four verbs *gallu*, *dylu*, *medru* and *cael*

into modal auxiliaries with semantics and syntactic use which “closely parallels that of English” (A.R Thomas 1982:213).

D. Philipps (2008) has recently taken the same argument further, proposing that convergence towards English systems is occurring across a broad range of marked Welsh features. In this category Philipps places the vingesimal numbering system, inflected prepositions, comparative forms of adjectives and aspects of the mutation system (D. Philipps 2008: 96). The perception that the erosion of long-standing characteristic features of Welsh is related to the influence of English has led to considerable wariness concerning the use of elements of English origin in Welsh, and some Welsh linguists have taken a particularly prescriptive line towards such phenomena in the past (Robert 2011: 140). Perhaps because of this, one of the clearest distinguishing features of modern standard Welsh has been described as its rejection of English-origin material (D.G. Jones 1988). For example P.W. Thomas, a sociolinguist and author of one of the most influential standard grammars, argues that: “[...] no matter how persuasive the arguments in favour of bilingualism, the harsh truth is that the influence of English is seeping deeper into the essence of the Welsh language” (P.W. Thomas 1996: 11, author’s translation in Robert 2011: 140). For Robert (2011: 140), this is proof that “Welsh standard language ideology is fervently monolingual and protectionist” and is primarily concerned with shielding Welsh from English transfer, which has been viewed as an exclusively negative phenomenon.

Arguably, this tendency can be traced back at least to the genesis of the modern nationalist movement, which explicitly made the elevation of the standard language one of its central goals (Aitchison and Carter 2000). Saunders Lewis, who is still admired by many modern Welsh speakers as a pioneer of language rights protest in Wales, as well as for his influential *Tynged Yr Iaith* radio broadcast, was also a strong critic of bilingualism, stating: “Drwg, a drwg yn unig, yw fod y Saesneg yn iaith lafar yng Nghymru. Rhaid ei dileu hi o’r tir a elwir Cymru:

delenda est Carthago” (“It is an evil, and solely an evil, that English is a spoken language in modern Wales. It must be annihilated from the land called Wales: *delenda est Carthago*”) (Lewis, 1938: 59). Lewis also founded the Welsh nationalist party *Plaid Cymru* and was an editor of the official party publication *Y Ddraig Goch*. Many of the anonymous editorials made no secret of their disdain for the admixture of Welsh and English, describing it in terms which show a deep aversion towards bilingualism:

“[...] ac i’r sawl a fydd byw mewn dinasoedd fel Caerdydd neu mewn rhai o gymoedd poblog y deheudir, daw’r Gymraeg nad yw Gymraeg a’r Saesneg nad yw Saesneg, y bobl nad ydynt na Chymru na Saeson, y diwylliant nad yw’n ddim ond hanner peth, y bywyd nad yw’n ddim ond basedd salw afiach, daw’r pethau hyn ar ei draws o hyd ac o hyd nid i’w ddiflasu na’i droi’n ddigalon yn gymaint ag i’w ysbardynnu i fynnu’r peth cyfan neu ddim [...]”.

“[...] *and to those who live in cities like Cardiff or in some of the populous valleys of the South, comes the Welsh that is not Welsh and the English that is not English, people who are neither Welsh nor English, a culture that is only a half-thing, a life that is nothing but vile, ugly superficiality, these things will come up again and again not to dismay us or dishearten us so much as to compel us to demand the whole thing or nothing [...]*” (*Y Ddraig Goch*, May 1927: 3)

The influence of this cultural ideology is arguably still felt in the prioritization of neologisms that avoid the influence of English in the development of Welsh language terminology. Robert (2011) describes this work as feeding through to a small industry of language professionals whose aim is to ensure the availability of Welsh equivalents (or ‘doublets’) for English terms in any possible field. This has led to a situation in which standard Welsh has a preference for obscure or novel terminology, while the majority of colloquial speakers may find it easier to

use an equivalent established English loan word or code-switch (Ball et al 1988). This distribution of elements of transfer between standard and vernacular varieties can thus be described as one of the most salient aspects of style for modern Welsh speakers.

1.1.3.2 Standard Welsh and mutation

Another distinction between standard and vernacular Welsh lies in the use of mutation. The actual colloquial use of this system is widely reported to be substantially narrower than in the standard code. According to D. G. Jones (1988), there are around 40 triggers in the standard language, many of which are unlikely to be seen or heard outside of formal written registers. There is some controversy over the extent to which nasal and aspirate mutations are now being used in colloquial speech. Some studies (Ball 1984; Jones 1998) show a decline in general use, and others (e.g. Awbery 1986) argue that speakers are moving towards a two tier system of mutation, with only radical forms (where mutation is omitted and the initial consonant is maintained) and soft mutations continuing in widespread vernacular use. It is clear then, that the use of mutation in standard and vernacular Welsh is likely to differ significantly and that mastery of some of the more obscure mutations, which are considered essential in written Welsh, may not be accessible to all speakers.

1.1.4 The current situation and Welsh speakers' exposure to Welsh

As I have claimed that style in Welsh may be affected by the diglossic interplay between the standard and vernacular codes in Welsh, this section will aim to establish to what extent the standard code is present in the lives of modern Welsh speakers. Despite the expansion of Welsh into some high domains of Welsh life, bilingual diglossia with English remains prevalent in many areas, with English still being used, for example, as the predominant code in commercial domains and in the private sector (C. Williams 2010). Given that English is the dominant

language, exposure to Welsh in public life is likely to be intermittent, as Rhys and Thomas state:

“[...] English is the dominant community language in many regions of Wales. Consequently, all public examples of language are expressed either in English only, or bilingually when in accordance with concurrent Welsh language governmental strategies. Children are therefore exposed to English written material in almost all domains of language use and are exposed to spoken English by a wide variety of individuals, and across numerous contexts of language use” (Rhys and Thomas 2013: 650).

The domains in which standard forms of Welsh seem most likely to be encountered are in literature, religious services, the public sector (i.e. local councils and other state-run institutions), the education system and Welsh-language media. I would also argue that speakers *can* be exposed to standard forms in other domains, although this may be more sporadic and dependent on factors such as geographical area and individual preference. In this category, I would include the private sector (encompassing certain businesses, supermarkets and banks), social networks (such as Facebook and Twitter), and associated technological domains. In the next sections, I will describe each of these domains and discuss the broad likelihood of exposure to standard Welsh that the majority of speakers may experience. In the last section, I will examine the Welsh media with a specific focus on *BBC Radio Cymru*, the primary Welsh language radio station, which is part of the focus of this thesis.

1.1.4.1 Literature and reading in Welsh

A.R. Thomas (1982) argued that Welsh speakers are particularly limited in their acquisition of standard Welsh forms by their lack of exposure to Welsh medium literature:

“Their situation is compounded by the fact that many speakers, outside the professional classes, do not read or write in Welsh with any regularity: they are effectively (by choice) non-literate, if not actually illiterate” (A.R. Thomas 1982: 216).

Similarly, although the authors do not provide precise numbers, Rhys and Thomas (2013) argue that the low “availability and (perceived) desirability of Welsh books” is a factor in limiting children’s exposure to Welsh, thereby implying that Welsh speakers are in many cases unlikely to read in Welsh frequently, and they add : “although there is a strong age-appropriate literacy tradition both in original book form and in the form of translations, this provision is in constant competition with the much broader and easily accessible English market” (Rhys and Thomas 2013: 649).

1.1.4.2 Religious services

Religious services have been described as one of the primary domains which sustained the existence of standard Welsh historically (Jenkins 2000). K. Jones (1995) described the use of Welsh in a Methodist chapel as primarily involving standard Welsh, although the majority of the congregation were only exposed to the code in a passive sense. Jones also distinguished the variety of Welsh being used as particularly notable for the marked absence of code-switching. The likelihood of a Welsh speaker attending a religious ceremony in Welsh of course depends on a number of factors, but the overall centrality of the chapel to Welsh life has diminished substantially since the nineteenth and early-twentieth centuries. A survey carried out by *Gweini* (2007:20) estimates that only around 7% of the Welsh population attend religious services frequently, with about 15% attending at least once a year. It is thus unlikely that religious services currently constitute a significant source of exposure to standard Welsh for the majority of modern Welsh speakers.

1.1.4.3 Welsh in the public sector

Political pressure from the middle of the twentieth century led to a series of legislative acts granting greater representation to Welsh in the public sphere, leading in turn to a greater level of exposure of higher forms of the language in institutions, which were compelled to produce Welsh language material (J. Davies 2014). Institutions affected ostensibly include any public bodies that are part of the government, including hospitals, local councils and public signage. Public signage is mandatorily produced in standard Welsh as well as in English in all areas of Wales, although the respective prominence of the Welsh or English message varies depending on the area, with Welsh appearing first in the North West and vice versa in the South East (Coupland 2012). A recent survey by C. Williams (2010) found substantial variation in the way with which the 1993 Welsh Language Act was complied in different areas in Wales. The county of Gwynedd in north-west Wales was consistently found to have complied with the Act to the greatest extent. For example, Williams describes the work culture of Gwynedd Council as operating internally almost entirely through the medium of Welsh:

“Gwynedd County Council represented the most constructive best practice found in the sample. The council operates internally through the medium of Welsh, all memoranda and internal communications are in Welsh, as is its correspondence with other public bodies in Wales. Gwynedd has adopted a proactive policy which goes far beyond the guidelines of the 1993 Welsh Language Act. In this respect Gwynedd is the closest we have to a successful model of bilingual operation.” (C. Williams 2010: 45)

Another body praised in the report was North Wales Police, who were also described as having engaged particularly proactively with the policy recommendations of the Welsh Language Act. Other Welsh regions were deemed to have avoided implementing policy recommendations however, with most areas providing an “incomplete service” (C. Williams 2010: 46). The

current reality thus describes a complicated pattern of language use, with some areas using Welsh in public sector domains, while others do not. Another relevant factor is the degree to which service users actually elect to use provisions in standard Welsh where they are available. Madoc-Jones et al (2013) provide the example of Welsh speaking service users in the criminal justice system choosing to undergo court hearings in English rather than Welsh, due to their lack of comprehension of the high register being used. Exposure to Welsh-medium services in high domains thus also involves individual speaker choice, which in turn may reflect social inequalities.

1.1.4.4 *The private sector and the private sphere*

Although the Welsh Language Measure (2011) oversees the provision of Welsh language services by utility providers (such as gas and electricity companies), other parts of the private sector currently remain exempt (as of 2015). In the majority of cases therefore, the use or non-use of Welsh is still left to the discretion of private businesses around the country (C. Williams 2010). However, recent years have seen an increase in the use of Welsh by private businesses, particularly in areas in which Welsh speakers are in the majority, such as Gwynedd in the North West of the country (C. Williams 2007). For example in the city of Bangor (where I live) the majority of supermarkets now have prominent bilingual signage, bilingual self-service machines and often also explicitly advertise the fact that they have Welsh-speaking staff. This trend may be at least partly due to political pressure, with social activism by language campaigners continuing to target the provision of Welsh at high profile businesses around the country even now. However it may also may be due to a growing change in positive attitudes towards Welsh and its potential as a language of business (C. Williams 2007).

1.1.4.5 Internet based social networks

Large international web-based companies such as Facebook and Twitter have a great deal of influence on people's daily linguistic input, and the relationship between Facebook and Welsh has been explored in a series of studies by D. Cunliffe and associates. In line with *social network theory* (Milroy 1980), Cunliffe (2010) has argued that Facebook could potentially serve to reinforce the local linguistic norms of Welsh-speaking internet users against those of standard or majority groups. A recent survey (Cunliffe et al 2013) found that many speakers do communicate in Welsh on Facebook, but that factors including the user's linguistic confidence and competence, the vitality of their speech community, the message topic and its intended audience, all played a role in determining the language chosen.

Studies so far have focused on language choice between English and Welsh. However, the role of the relatively recently added Welsh language interface is an interesting factor which involves stylistic considerations 'within' Welsh. Facebook introduced the option to use the website user interface entirely in Welsh in 2008, with translations being provided by users through a process of crowd-sourcing (Cunliffe et al 2013). Young people interviewed in the survey voiced a variety of attitudes towards the Welsh language interface, and the authors provide the following two quotes as examples:

“Dwi wedi newid o ddwywaith i Gymraeg, ond dw i ddim yn gallu handlo fo... roedd na eiriau dwi ddim yn dallt yna, geiriau massive”.

(I've changed it to Welsh twice, but I can't handle it . . . there were words that I didn't understand, massive words.)

Extract 1 Young participant describes Welsh-medium Facebook interface (Cunliffe et al 2013: 356)

“Mae’n llawn camgymeriadau. Mae nhw yn ceisio Cymreigio rhai o’r geiriau Saesneg, ond ddim yn gwneud ymgais i gael y Gymraeg yn gywir”.

(It’s full of mistakes. They try to Welshify some English words, without making an effort to see if the Welsh was correct.)

Extract 2 Young participant describes Welsh medium Facebook interface (Cunliffe et al 2013: 356)

These examples make it clear that some young Welsh speakers find the interface terminology - which generally reflects standard Welsh norms – alienating, either because it is perceived as *too* standard (particularly in its use of low frequency computing terms which avoid borrowings) as in the first quote, or because it is not standard enough, as in the second. These attitudes serve to highlight the fundamental difficulty involved in developing a variety of Welsh ideal for public consumption in a heterogonous community. The authors do suggest, however, that the use of the Welsh interface might normalize the use of Welsh equivalents for computing terms which have previously been predominantly used in English. The scale of this potential influence is largely unknown, as the option to use a Welsh-medium interface on Facebook is voluntary, and no numbers have been published concerning how many Welsh speakers choose to make use of such services. In summary, the use of social networking services has the potential to reinforce both vernacular norms and standard ones, although more research is needed.

1.1.4.6 The education system

One of the clearest vectors of delivery of standard Welsh is in the school system. The 1988 Education Act made Welsh a core subject in Welsh-medium schools, and a foundational subject in all other Welsh schools, meaning that every school in Wales is statutorily obliged to teach Welsh in some capacity (Welsh Government 2013). Despite this, provision varies from

school to school in what Baker refers to as a “kaleidoscopic variety of bilingual education practice” (Baker 1993: 15). As Jones (1998) notes, the provision depends to some extent on the demographic and sociolinguistic qualities of the area, with provision in Anglicised areas likely to be limited, while in places like Gwynedd, Welsh may frequently be the medium of instruction for every subject except for English and Modern Languages. Although the impact of standard Welsh on schoolchildren has not been widely studied, Jones found that exposure to standard Welsh among schoolchildren was leading to the loss of several local dialectal features, and the adoption of many features that were not part of the local dialect. This finding suggests that Welsh medium education is able to promote the acquisition of the standard variety of Welsh, although it is not entirely clear to what extent this occurs in less Anglicised contexts outside of Jones’ study, where the standard will come into competition with vernacular Welsh norms.

Support for Welsh-medium education at third level has lagged in comparison with that at secondary and primary level (Baker 2009), and only 43% of fluent Welsh speakers at university level reported taking either part or all of their course through the medium of Welsh (Centre for Welsh-Medium Higher Education 2009, quoted in Lewis and Andrews 2104: 173). The lack of progression in Welsh medium provision between secondary and tertiary levels of the education system has been described as a challenge to efforts to promote the renewal of Welsh language skills, because “diminishing provision and take-up at the older age, and the linguistic loss that occurs, has been a cause of general concern for many years” (Lewis and Andrews 2104: 172). However, Welsh-language provision in the third level is reported to be expanding at the current time, and a particularly important recent development in the field of Welsh education was the launch of *Y Coleg Cymraeg Cenedlaethol* (The National Welsh College) in 2011 (J. Davies 2014). The *Coleg* is intended to support the development of Welsh-medium academia, with funding made available for a significant number of scholarships, which must

be completed through Welsh. This is predicted to expand the availability of Welsh-medium provision learning in third-tier education, and consequently the use of Welsh in higher learning. Furthermore, this development exemplifies the fact that the extension of the prestige of Welsh is a continuing process.

1.1.4.7 Welsh in the media

Besides the school system and the government, the other main source of the H variety is the Welsh broadcast media, including television and radio. At the beginning of the twentieth century, all radio broadcasts were controlled by the state and restricted to English, with the head of the Cardiff station responding to protests by declaring that “it was natural to restrict broadcasting to the official language, and that it would be wrong to yield to extremists who were trying to force Welsh on listeners” (J. Davies 2007: 548). Following political pressure, including a license-fee boycott led by Saunders Lewis, the BBC’s stations in Wales began broadcasting Welsh language material in 1935, and won a separate Welsh Region wavelength in 1937 (J. Davies 2007). Demands from viewers for distinct services for both languages combined with political pressure from Welsh language activists, and this culminated in the expansion of *Radio Cymru* to fully Welsh language broadcasting in 1979 (Aitchison and Carter 2000). Today *Radio Cymru* is one of many radio stations broadcasting in Wales, broadcasting alongside its English language equivalent *Radio Wales* as well as well as numerous local and regional stations. Some of these more local stations do broadcast in Welsh – for example in North Wales the regional branch of pop-music service *Capital FM*² currently provides a mixture of Welsh and English programming.

Smith (2000) describes the ‘quality’ of spoken Welsh in the Welsh media as a long running and persistent source of controversy. Initially programmes seem to have followed the example

² <http://www.capitalfm.com/walescoast/on-air/schedule/>

of Anglophone traditions in that period, as well as the preference of language purists, of primarily reflecting standard varieties, particularly in formal programmes:

“Broadcasters were often regarded as the purveyors of standard Welsh and there can be no doubt that properly prepared news bulletins and scripted contributions made a large number of listeners conversant with a mode of expression in which traditional standards were respected.” (Smith 2000: 339)

However, Smith also indicates that this template held to some extent even in less formal domains: “many Welsh drama productions chose purity rather than authenticity [...]” (Smith 2000: 339). Not much work exists detailing the use of Welsh in the media during this period, but Ball et al’s (1988) study of Welsh on *Radio Cymru* found relatively formal patterns of Welsh in three radio programmes studied, with frequent use of prestigious aspirate and nasal mutation as well as standard ‘written’ forms of pronouns, which diverged markedly from the vernacular styles observed in Ball’s community speech study (Ball 1984). However, according to Smith, the tendency towards prioritization of standard language use began to change in the mid-nineties, with increasing use of vernacular language, including code-switching and English language music and interviews occurring from 1994 onwards (Smith 2000).

1.1.4.7.1 De-standardization of Radio Cymru?

The contemporary *Radio Cymru* is currently going through a series of important changes, with a recent change in chairwoman to Betsan Powys coinciding with a pledge to modernise the station. These changes aim to increase the station’s appeal towards younger speakers, which *Radio Cymru* has struggled to attract in the past. Powys promised to increase the appeal of the station through an inclusive approach, involving all kinds of people and crucially, all varieties of Welsh:

“Ehangu'r apêl amdani felly, a derbyn na fydd pob rhaglen yn apelio at bawb. Ond fe fydd pob un yno i bwrpas, i apelio at rywun. Fe fydd pob math o leisiau ar yr orsaf, pob un yn hoff lais i rywun, a phob un yn siarad Cymraeg rhywun”

“We will be broadening the appeal, and accepting that not every programme can appeal to everyone. But each one will be there for a purpose, to appeal to somebody. There will be all kinds of voices on the station, each one somebody’s favourite voice and each one speaking somebody’s Welsh”.

Extract 3 *Radio Cymru* chairwoman Betsan Powys describing policy changes (BBC 2013)

One of the most widely discussed aspects of this new format was the inclusion of Tommo, a popular personality who formerly worked at regional South West Wales radio station, and who is well known for his frequent use of vernacular and code-switching in his speech. The trend towards the inclusion of less standard forms on *Radio Cymru* is arguably part of the widely reported movement towards de-standardization in the media in Western countries (Coupland and Kristiansen 2011).

1.1.4.7.2 Reactions to de-standardization

It seems sensible that the directors of *Radio Cymru* might aim for a more inclusive range of speech styles, particularly in light of research which has shown that listeners often prefer to hear speech varieties that are similar to their own (Giles et al 1991, Bell 1984). Not all listeners agree with the need to change the station’s output in this way however. Multiple pressure groups have recently appeared, protesting against a perceived laxity of linguistic standards on *Radio Cymru*, including ‘overuse’ of English and ‘poor Welsh’ by individual presenters, as well as excessive airplay for English language songs. One such group, *Cylch yr iaith*, has carried out public protests, with one of its members, Geraint Jones, being imprisoned for ten days and fined £330 for refusing to pay his television license in 2011. He was reported in the

press as saying that his protest was intended to challenge: “haerllugrwydd a Seisnigrwydd BBC Radio Cymru” (“*the arrogance and Englishness of BBC Radio Cymru*”) and its content, which he described as “*sothach*” (junk) (BBC News 2011).

A more recently created, and perhaps more conventional pressure group, is *Dyfodol i’r Iaith* (A future for the language) which is composed of many well-known members of Welsh speaking public life, including university lecturers and former members of parliament. Although their policy aims are wider than those of *Cylch yr Iaith*, encompassing education, language planning and other issues, they have also challenged *Radio Cymru* on the use of language by presenters. In one blog post from 2014, the group claim to have received: “a number of concerns and complaints from its members regarding the presenters’ use of sub-standard Welsh and the number of English songs being broadcast on *Radio Cymru*” (Dyfodol i’r Iaith 2014). In a letter to the BBC, the pressure group questions the supervision of presenters - “without naming any individual presenters” - who are believed to be using ‘sub-standard’ Welsh:

“What advice and guidance is provided for programme presenters regarding the use of natural spoken Welsh in any dialect rather than ‘pidgin’ Welsh and English terms? Do they receive training in that regard, are they being supervised daily, and do they receive advice and guidance when they transgress?” (Dyfodol i’r Iaith 2014).

It seems clear then, that there is an influential segment of the population which opposes the process of de-standardization. What is less clear is to what degree this point of view reflects that of the wider Welsh-speaking community. In their recent discussion of de-standardization in the Anglophone media, Coupland and Kristiansen (2011) have presented excerpts of internet-based discourse posted on social media to illustrate the widening ideological field, with some English speakers now more likely to evaluate non-standard forms positively than in

the past. Before I move on to the next chapter, I would like to follow this strategy in order to present some similar examples in the Welsh-speaking context, which I argue reflect a wide range of attitudes. My first piece of evidence is taken from a debate that was broadcast on *Radio Cymru* in 2011 between *Cylch yr Iaith* member Peter Hughes Griffiths and well known musician and DJ Gareth Potter, who represent opposite sides in the discussion over language standards. The whole interview is currently available (as of mid-2015) in two YouTube videos on a blog (Morris 2011) at <http://ytwill.com/2011/06/gareth-potter-peter-hughes-griffiths/>, and I have transcribed parts which I think are particularly illuminating.

During the interview, *Cylch yr Iaith* spokesman Peter Hughes Griffiths states that he believes that the quality of *Radio Cymru* presenters' Welsh is too low, that they use too much English and misuse the Welsh gender system, and that some of the material covered in programmes is insufficiently 'Welsh'. Potter, meanwhile, argues for a more inclusive approach, saying that excessive purism impacts negatively on Welsh speakers' confidence, and prevents them from engaging with the language. He further argues that 'experts' in the media, who are well versed in a particular subject, should not need a grasp of standard Welsh in order to participate, and that Welsh speakers should be accepting of all linguistic 'errors'. He then uses himself as an example of a Welsh speaker who uses *bratiaith* (code-switching) and "doesn't mutate", but who still considers himself worthy of ingroup status:

Gareth Potter: "Oedd yn athrawon i wastad yn dweud bod yn Nghymraeg i yn fratiog, ond doedd dim ots gyda fi. Dw i isio siarad Cymraeg, a w i moyn bod yn rhan o'r Cymru Cymraeg diddorol a *fascinating*. A mae yna cymaint o bobl fel fi. A mae'n rhaid i fi *represento* y pobl yna".

“My teachers always said that my Welsh was poor³, but I didn’t care about that. I want to speak Welsh, and I want to be part of this fascinating and interesting community of Welsh speakers. There are so many people like me. And I have to represent these people”.

In response, Griffiths adopts a harsh tone, and states that the language is effectively doomed if Potter’s inclusive vision for the Welsh-speaking community is entertained any further:

“Os yw Gareth yn gweld mai dyna ydyw dyfodol y Gymraeg, fedra i ddweud fan hyn wrthy ch, does fawr o ddyfodol iddi hi, a waeth i ni roi y ffidil yn y to os dyna’r Gymraeg...”

“If Gareth thinks that this is the future of Welsh, I can tell you now, there won’t be much of a future for it, and we may just have to give up entirely if that’s the Welsh...”

This triggers a rather explosive response from Potter:

*“Rowch y ffidil yn y to! Wna i siarad fel y fi moyn siarad. A timod - a weda i hwn wrth Cylch yr Iaith: stick a website up there er mwyn i ni gweld beth yw eich amcanion chi! Sa i gallu ffeindio chi ar y we. Y chi fel ryw fath o *Illuminati* Cymraeg. Ryw *elite*. A nes y chi’n gwneud hynna, y chi ddim yn mynd i gael eich cymryd o ddifri o gwbl. A chi’n mynd i gasáu pobl fel fi. Chi ‘di gweud e nawr yn blwmp ac yn blaen fod fy math i o Gymraeg ddim gyda ddim math o ddyfodol. A chi *actually* yn casáu ni. Ewch i’r gorllewin a, timod, byw yn eich *bubble* chi. Da iawn. Wnawn ni gario mlaen fan hyn.”*

*“Give up!? I’ll speak any way I want to. And you know – and I’ll say this to Cylch yr Iaith: stick a website up there so that we can see what your aims are. I can’t find you on the internet. You’re like some kind of Welsh *Illuminati*. Some kind of elite. And until*

³ The word *bratiog* here has connotations of ‘fragmentation’, which suggest that Gareth is mainly referring to the incorporation of code-switching and other elements of English transfer into his Welsh.

you do that, you won't be taken seriously at all. And you'll hate people like me. You've made it clear that my kind of Welsh has no future. And you actually hate us. Go to the West and live in your bubble. Very good. We'll carry on here."

The vast majority of the commentary that I've been able to discover on social media about this explosive interview actually seems to reflect a relatively dim view of *Cylch Yr Iaith*'s purism, and support for Potter. The following tweets are typical of the response on Twitter:

"Syt fysa Cylch yr Iaith yn ymateb i rhywun yn deud, os chi'n cario mlaen siarad ffocing shit, dwi am wrthod siarad Cymraeg o hyn y mlaen?"⁴

("How would Cylch yr Iaith respond if someone were to say, if you keep on talking fucking shit, I'm going to refuse to speak Welsh from now on?")

"Ma Cylch yr Iaith yn gyrru fi i Nant Gwrtheyrn⁵ i gael "electric shock therapy" i ciwrio Cymraeg cachu fi"⁶

("Cylch yr Iaith are sending me to Nant Gwrtheyrn to get "electric shock therapy" to cure my shit Welsh")

"Diolch am blog gwych crynhoi teimladau cymaint ohonom at y ffasciaeth ieithyddol yma. Hen bryd datgelu'r snobyddiaeth afiach"⁷

("Thanks for the great blog which sums the feelings of so many of us about this linguistic fascism. It's high time for this disgusting snobbery to be revealed")

Extract 4 Welsh Twitter messages taking anti-purist stances

⁴ <https://twitter.com/dylmei/status/77716526068809728>

⁵ Nant Gwrtheyrn is a popular Welsh language and heritage centre to which learners go to improve their Welsh.

⁶ <https://twitter.com/marshallmedia/status/77854076125200384>

⁷ <https://twitter.com/RichardMRees/status/462283360707301376>

The balance in favour of anti-purist discourse may be a result of the fact that, as Gareth Potter suggests in his debate with Peter Hughes Griffiths, *Cylch yr Iaith*'s main base of activists are generally older, and therefore may be less likely to have a presence on social media. For instance, the member who was sent to prison mentioned above was in his late sixties, and this generally seems to reflect the group's demographic base. The ideological clash between purists and their opponents could perhaps then be seen as a generational confrontation, although interestingly there is also a geographic, as well as an attitudinal and cultural dimension – with the Welsh speaking, rural west situated in opposition to the implicitly more cosmopolitan and urban east of the country. More research is certainly needed for the full range of factors in the development of these competing ideologies and discourses to be clarified, but I believe that the contrasting views of Potter and Griffiths, extreme though they may be, provide a relatively good representation of the span of the ideological field among Welsh speakers at present.

Despite the strength of feeling concerning the issue, *Radio Cymru*'s own research indicates that the population in general does not share the outlook of purists such as *Cylch yr Iaith* and *Dyfodol Yr Iaith*. In a survey from 2011, the station found that 85% percent of respondents agreed that “the Welsh language is spoken to a high standard on BBC Radio Cymru” (Vagg et al 2011: 52). In contrast, the most significant source of negative appraisal of the station was on measures of creativity and innovation, with the researchers presenting the following free form responses as a sample of opinion:

“I think their programming could be a little less ‘safe’. They don’t seem to do new and interesting stuff. A bit more variety in some of the presenters perhaps.” Female, aged 63, DE.

“Don't imitate, be more creative.” Male, aged 53, C1.

“The format is old fashioned. It hasn't moved with the times.” Female, aged 56, C1.

Extract 5 Listener feedback on lack of innovation on *Radio Cymru* (Vagg et al 2011: 53)

Negative perceptions were also particularly common among younger respondents:

“Most evidently, negative performance gaps were generally larger for younger listeners aged 15 to 54 than listeners aged 55 or over. This reflected the over-representation of adults aged over 55 in the audience and a general feeling that the station caters better for older listeners” (Vagg et al 2011: 53).

These sentiments are reflected in the comments below:

“Some music is awful and not suitable for younger people. I know of a lot of people who don't listen to it due to this.” Female, aged 53, C1.

“It needs more easy going things for the younger people, news especially. Bulletin isn't suitable for young people.” Female, aged 18, C1.

Extract 6 Listener feedback on lack of innovation on *Radio Cymru* (Vagg et al 2011: 53)

It is not clear from the BBC Trust report whether younger speakers’ negative appraisals of the station might be linked in some way to the type of Welsh which presenters use, with other factors like the type of music or topic the most likely to be mentioned in the criticism. But the 2013 Beaufort report into the linguistic habits of Welsh speakers, which also found that “The TV and radio content available in the Welsh language was often believed to be aimed more at

older people” (McAllister et al 2013: 74) presents the following comment, which seems to allude to difficulty of comprehension of the Welsh used on the channel, even, in this case, for a self-reported fluent Welsh speaker:

Radio Cymru is like a Radio 4⁸: slow, boring, really, really difficult sometimes to understand. (M, 41-55, fluent, less often than weekly, RCT)

Extract 7 Beaufort report participant criticising *Radio Cymru* (McAllister et al 2013: 74)

The fact that a middle aged, self-reported fluent Welsh speaker would claim to find the variety spoken on *Radio Cymru* “hard to understand” is striking to say the least, and raises the possibility that some speakers may find the Welsh used *overly* standard in some cases, although to what extent this is true will remain unclear without more research. In general however, these research findings seem to suggest a situation where the majority of speakers are ostensibly happy with the speech varieties used on *Radio Cymru*, with little sign of the purist complaints of organizations such as *Cylch Yr Iaith*.

1.2 Study Design

My central intention in this thesis has been to explore how style works in Welsh. Drawing primarily from elements of variationist, audience design and speaker design methodologies, my first task has been to identify *sociolinguistic variables* appropriate for the setting. Sociolinguistic variables can be most simply defined as linguistic features with multiple variants that correlate with social and/or stylistic factors, which are ideal for quantitative sociolinguistic research (Tagliamonte 2012). For instance, one of the most widely studied sociolinguistic variables has been the variable (ing), which encompasses two distinct variants. The variant [iŋ], for example in the word “*running*” is associated with prestige, while its

⁸ BBC Radio 4 is the BBC’s talk radio channel, which is stereotypically associated with older and more middle-class listeners.

counterpart [in], as in “*runnin*”, is broadly stigmatized. In variationist research, [ɪŋ] has frequently been found to correlate with formal speech styles (involving more attention to speech) as well as with more prestigious speakers (Labov 1966).

This leads us to the first hurdle in applying variationist techniques to Welsh. Phonological sociolinguistic variables like (ɪŋ) have often been considered the most suitable type for analysis, for various reasons including their frequency and transparency (Tagliamonte 2012). However, according to Martin Ball, one of the few researchers who has tried to apply variationist methods in the Welsh context, “it is doubtful whether one form of pronunciation exists that is generally accepted as standard throughout the country” and previous attempts to describe such a variety have “described formal registers only likely to be used by a minority of Welsh speakers” (Ball and Williams 2001: 10). Some studies have described very local patterns of social and stylistic phonological variation, for example the devoicing of consonants in Glamorgan (E. Thomas 1988), where the loss of the feature in younger and more educated speakers reflects the influence of standard Welsh education and anglicization in rural areas. In general, however, we can say that phonetic variables primarily seem to encode regional differences in Welsh – there are ‘regional accents’, but not commonly used social or stylistic ones.

A more promising candidate as a sociolinguistic variable is mutation, the morphosyntactic process of sound change which affects the initial consonant of words (Ball and Müller 1992). Mutation is a marked feature of Welsh and the other Celtic languages, and which may be relatively salient because of this. It involves a phonological alternation, but crucially in many cases does not carry a functional load (P.W Thomas 1984), which means that the variation between the application and non-application of mutation arguably has the potential to carry a primarily social and stylistic meaning for hearers. Mutation has already been analysed as a sociolinguistic variable in Welsh in a series of studies by Ball and colleagues (Ball 1984; Ball

et al 1988) and Hatton (1988), in which the researchers claimed to find that the aspirate mutations and nasal mutations behaved like stylistic markers, reflecting social differences between speakers (including age, cultural orientation and gender) as well stylistic variation.

Another potential candidate is suggested by A.R. Thomas (1982), who as I mentioned above has written on the influence of English on Welsh. Among the most interesting of Thomas' predictions in this essay was that the attrition of Welsh's traditional systems for varying of style would lead to 'style attenuated' speakers using the very features which represent interference from English in Welsh for the purpose of stylistic variation. Thomas singled out code-switching, the admixture of English language elements into Welsh utterances, as one resource which speakers would draw upon if his predicted atrophy of native features in the 'dying language' were to continue (my italics):

“One productive area of research will be the role of code-switching in the usage of style-attenuated speakers, at first phrase and idiom bound but later necessary to furnish the grammar no longer available in the dying language. *Additionally, code-switching has an important function in that it compensates for the reduction of stylistic options in the dying language (and undoubtedly facilitates further loss by displacing stylistic variation)*” (A. R. Thomas 1982: 218).

More recently, Deuchar and Davies (2009) have corroborated this statement, stating that “[...] one could argue that code switching is nowadays a characteristic of informal speech, distinguishing it from formal speech where switching tends to be avoided” (2009: 19). Code-switching has been associated with informality in many studies, although attempts to quantitatively measure the stylistic variation of code-switching have been comparatively infrequent (Zentella 1990; Dorian 1994).

My research question can thus be refined to: can mutation and code-switching be analysed as sociolinguistic variables in Welsh? Answering this question requires some degree of falsifiability – what are the criteria for deciding whether a feature can be considered a legitimate sociolinguistic variable? Fortunately, the variationist paradigm offers a solution in Labov’s (1966) model of indicators, markers and stereotypes. This model considers linguistic features as having different levels of saliency in the minds of community members, and as being able to change their salience over time in response to social changes (Bell 1984). Indicators, the least salient type, encode information about speakers’ social group membership, e.g. class or ethnic group. Markers are more salient, and are under the control of speakers to some extent, such that speakers are able to adapt their use for stylistic purposes. Like indicators, markers also index social group membership. Stereotypes are features that have risen above the level of consciousness and become subject of comment, and they are often stigmatized. Because they are so salient, stereotypes can be easily avoided by speakers, such that they do not index group affiliations, but they are stylistically available. These categories provide a way of testing the potential status of linguistic features. Simply put, features that vary with social categories are indicators, those that vary with styles and social categories are markers and those that vary only with styles are stereotypes.

For my analysis I have developed a semi-novel methodology, partially based upon a study by Ball et al (1988), which used radio to examine the stylistic distribution of variables in programmes that were categorized by differing levels of formality. This study will consist of two parts, and will be organized around the analyses of two corpora. The first part will be devoted to understanding the social and language internal factors affecting the variables in a large naturalistic corpus of informal Welsh conversations collected from 2006 at Bangor University, the *Bangor Siarad* corpus (Deuchar 2014). This corpus will represent the vernacular (Labov 1966) – the most unmonitored, habitual speech style of Welsh speakers,

while they are engaged in interactions specifically with familiar interlocutors such as family and friends.

Previous analyses of the *Siarad* corpus by other researchers have found relatively frequent code-switching, with younger speakers using significantly more code-switching than older speakers, although other social factors such as level of education were not significant (Lloyd 2008). In contrast, the distribution of mutation in the *Siarad* corpus remains relatively unexplored, although Stammers (2010) has analysed patterns of syntactically triggered mutation. My analysis will involve eight lexical trigger variables, which I have coded as (o), (am), (ei.m), (ei.f), (a), (yn), (fy) and (â/gyda). Analysis of these variables involved an initial period of what Tagliamonte (2012) calls “circumscribing the variable context”, in which the grammatical system of the variables was explored in order to rule out any potential skewing effects. For mutation, this primarily involved removing tokens (i.e. instances of the relevant variants in the data) that were categorical, such as those in ‘fossilized’ formulaic constructions (A.E. Jones 1988), or which represented ambiguous contexts, such as target words which might be considered either borrowings from, or code switches into, English (Stammers 2010).

I also performed an RBRUL statistical analysis of some of the social and language internal factors affecting the variables, which revealed that mutation was in most cases positively correlated with the age and level of education of speakers, and that place names (and low frequency place names in particular) were strongly negatively correlated with standard mutation. I further performed an apparent time analysis of each mutation variable, with the participants split between seven age groups, in order to glean further contextual information about the status of each one synchronically within the speech community. These analyses confirmed that the use of mutation varies with social categories, and that it may be analysed as a sociolinguistic variable. Crucially, I was also able to calculate mean percentages and index

scores representing the community norm of standard mutation, allowing me to carry this data over as a baseline of informal language use to the next section of the study.

The second part of the thesis describes the collection of a small corpus of four Welsh language radio programmes, which I argue vary stylistically along a relatively linear continuum of formality. Programmes were first placed in a hierarchy of formality using quantitative methods, drawing on Labov's notion of channel cues (1966), with the frequency of laughter and overlapping speech used as an index of the relative level of formality in each programme. I complemented this with a more qualitative approach involving discourse analysis of the programme transcripts, to gain a deeper level of understanding of the data. These avenues of research revealed that the quantitative measures of formality coincided with institutional values of subversiveness versus authority, and revealed stance taking (Ochs 1992, Kiesling 1998), with presenters situating themselves in contrasting ways within an ideological framework of Welsh standard language ideology. In other words, contributors in programmes which had lower frequencies of channel cues indicating emotional engagement could also be described as taking 'authoritative' stances indexing values of sincerity, prestige and seriousness. By contrast, those with higher frequencies of laughter and overlapping speech were taking 'subversive' stances, indexing values of rebelliousness, irony and solidarity.

As well as the level of formality, demographic details about the contributors who make up each programme were also taken into account, including their relative ages and experience with standard Welsh. Unlike Ball et al (1988), I did not control for the contributors' attributes beyond dialect area and gender, partly because this would have been extremely restrictive in terms of the range of data available to use. Furthermore, I argue that taking this approach has allowed me to gain insight into the way in which each variable responds to social as well as stylistic factors. Some variables seemed to be distributed in a way which reflects style as a primary factor, and I argue such variables may be markers. The distributions of other variables

appear to be mainly shaped by social factors related to the programmes' presenters. I will make the case that such variables may be *indicators* – variables that are beneath the level of consciousness in the speech community, but which correlate strongly with speakers' social characteristics, such as their age. A final category of variable was most strongly affected by internal factors, such as the word frequency of certain kinds of target words as well as the gender system - or more specifically, the degraded status of the Welsh gender system in many speakers' repertoires (Thomas and Gathercole 2007) - making it difficult to measure whether such variables could be called markers or indicators.

1.3 Thesis structure

The structure of this thesis is as follows. **Chapter two** is a broad introduction to sociolinguistics and style research, discussing the key points of Labov's variationist methodology and the ways in which this has been adapted by subsequent researchers over the years. As well as classic variationism, I will also be discussing other approaches to style which contrast with that of the variationist paradigm, including the speaker design, audience design and register design models. **Chapter three** introduces mutation, first describing the system itself, and then outlining various studies that have attempted to survey variation in mutation among the Welsh speaking population either in terms of demographics or in terms of style. **Chapter four** reviews the literature on code-switching, paying particular attention to studies on code-switching in Wales. **Chapters five** and **six** comprise the methodology and results sections of the *Siarad* corpus analysis, while **chapters seven** and **eight** comprise the methodology and results sections of the *Radio Cymru* corpus analysis. **Chapter nine** describes the qualitative analysis of the *Radio Cymru* corpus data, while **chapter ten** is the discussion and **chapter eleven** is the conclusion, which brings together the threads of the whole thesis.

2. Style

2.1 Introduction

Approaches to style in sociolinguistics can be divided roughly between two schools – the *variationist school*, which uses community-wide large scale quantitative survey methods, and the *ethnographic school*, which uses *speaker-centred* qualitative methods. Within the scope of these broad paradigms I have identified five distinct approaches to style, which I will describe in the following order. Section (2.2) introduces the seminal *variationist* model along with the key sociolinguistic concepts of the speech community, vernacular, standard language, attention to speech and sociolinguistic variables. Section (2.3) discusses *accommodation theory*, which draws on insights from social psychology and explores style as a process of adapting speech to the preferences of the interlocutor, or alternatively as a device to emphasize group membership and distinctions (Giles 1972). Section (2.4) describes *audience design theory*, which combines the insights and methods of *accommodation theory* with those of classic variationism, in a model that considers the ways in which speakers design speech styles for different types of audiences (Bell 1984). Section (2.5) introduces *speaker design* approaches, which marshal observations from ethnographic and communications theory to foreground speaker initiative, creativity and the dynamic nature of style. Finally section (2.6) discusses *register design theory* (Finegan and Biber 1994), a controversial model which foregrounds social class and exposure to elaboration or economy forms as the primary factor in linguistic variation. One of the most significant trends in recent sociolinguistics has been the increasing combination of elements from these various approaches in order to triangulate a more effective theoretical and methodological approach to understanding style. This tendency has been referred to as the ‘third wave’ of sociolinguistics (Eckert 2005), and will be a substantial influence upon the theoretical approach of this thesis.

2.2 The Variationist model

One of the central aspects of variationist theory, which differentiates it from subsequent models, is its emphasis on the density and frequency of interaction between speakers within distinct social groups (but primarily in terms of class) as a constraint on the stylistic repertoires of speakers (Labov 2001.b), while giving less of a role to notions of speaker initiative, interactional strategy and the stylistic construction of self than later other models (Coupland 2007). Variationism also uses macro scale survey methods inherited from sociology (Labov 1972) to situate speakers within larger social and linguistic structures such as speech communities, in contrast to the ways in which subsequent models focused on smaller ‘communities of practice’ (Eckert 2000) or case studies of individuals (Rickford and McNair-Knox 1994).

Variationist approaches typically incorporate measures of intra-speaker variation (variation within the speech of the same speaker) and inter-speaker variation (variation between different groups of speaker – i.e. by age or gender etc.), which are argued to be interrelated (Romaine 1980). Labov’s seminal study *The Social Stratification of New York City* (1966) established a template for large scale sociolinguistic studies measuring stylistic and social variation using quantitative methods. The study used an easily replicable empirical methodology to measure variation in a group of sociolinguistic variables in the New York speech community. Labov introduced the concept of the *sociolinguistic variable* – a linguistic feature which encodes social and/or stylistic information about speakers and situations, and his analysis used *social class* and *attention to speech* as the central independent variables, which allowed the dependent sociolinguistic variable to be analysed along both social and stylistic lines. I will first describe the stylistic element of Labov’s method, focusing on attention to speech, before moving on to other aspects of variationist theory.

2.2.1 Labov's Attention to speech methodology

Labov considered *attention paid to speech* the crucial mechanism in modulating the formality of speech styles. The vernacular is considered to represent “the most systematic form of the language acquired by the speakers, prior to any subsequent efforts at (hyper) correction or style shifting” (Poplack 1993: 252). This is the variety that speakers will use most habitually during their lives, particularly with familiar peer members of the same social group, and in informal situations. Due to the influence of standard language ideology however, an interview with an unfamiliar outgroup member is expected to trigger at least some style shift towards a more prestigious style of speech, which provides the analysis of style shifting in variationism with its methodological basis. Labov argues that this process is modulated by *attention to speech*, as the majority of speakers must focus attention on speech to produce non-habitual prestigious styles. In order to measure such style shifting, his study methodology features a variety of tasks which allow researchers to manipulate the amount of attention a study participant is able to focus upon their language (Labov 1972).

The most informal (involving least attention to speech) context is the casual interview, which Labov defines as involving *careful speech* because “an interview which has as its professed object the language of the speaker will rate higher on the scale of formality than most conversation” (Labov 1972: 80). Once the interview section is complete, the researcher manipulates the interlocutors' level of audio monitoring with a series of reading tasks, intended to focus interviewees on the linguistic variables in increasing increments. This involves having participants read a series of carefully designed texts. Each increment of formality is labelled with an alphabetic letter, with B representing the casual interview segment (Context A is described below). Context C involves the participant being asked to read standard texts, with the target variables distributed in closely juxtaposed minimal pairs. The next task, context D, is a word list task which increases the level of formality by having participants read words in

isolation, thus increasing their focus on pronunciation. The most formal context, D', has participants reading a word list with minimal pairs, so that the pronunciation of each word receives maximum attention.

Labov describes manipulating speech styles in formal elicitation procedures as above to be “a simple matter” (Labov 2001.a: 88), but contrasts this with the difficulty of eliciting diminished attention to speech during the interview, context A. To achieve this, Labov uses a variety of techniques intended to disrupt the ability of interviewees to focus attention. These include discussing topics related to childhood – such as nursery rhymes – and questions intended to trigger an emotional state in participants, such as recollections of near death experiences (Labov 1966). Labov also retained any data recorded in serendipitous events where interviewees spoke to a third person (usually a family member or friend) when, for example, answering the phone during the conversation, as examples of speech more likely to reflect the actual vernacular of speakers.

Labov later refined these techniques using what he called *channel cues*: “changes in volume, pitch, tempo, breathing and laughter that are independent of the linguistic variables but signal overall shift of linguistic behaviour and emotional involvement” (Labov 2001.a: 89) which allowed him to better identify parts of an interview in which style shifting is really taking place. This method involves checking over parts of the interview where Labov believed style shifting might be occurring and looking for the presence of these kinds of paralinguistic elements. Although, as I will discuss later, many aspects of *attention to speech* have been criticised by theorists, the use of channel cues has received particularly strong criticism along practical and methodological lines. Wolfram (1969: 58) for example states that his experience of attempting to use channel cues to identify style shifting was essentially fruitless, particularly as cues such as laughter could be ambiguous, and might actually indicate nervousness rather than relaxation.

Perhaps for this reason, channel cues have not been used by mainstream variationists to any great extent (Meyerhoff 2011).

2.2.2 Sociolinguistic variables

Style shifting is measured in variationist approaches using sociolinguistic variables. These are linguistic features which can be split into two or more variants, and whose correlation with intra- and inter- speaker variation can be quantitatively measured. Labov (1966) used a specific system of notation to represent linguistic features used at different levels in a sociolinguistic analysis. A sociolinguistic variable is typically represented between rounded parentheses, so that the phonological sociolinguistic variable *postvocalic-r* as a whole is represented as (r). This variable has two possible realizations in the New York speech community that Labov studied, and these are represented between squared brackets as [r] and zero.

Vowels are more complex than consonants, with multiple possible realizations, so for example, the variable (eh) contains five separate variants. Each variant was provided with a numerical value, meaning that an average score could be calculated for individuals and groups and thus compared. In early variationist research, the focus of study was emphatically on phonological variables which were viewed as particularly suitable in terms of both method and theory (Labov 1966). However, subsequent studies have attempted to incorporate syntactic, morphological and lexical variables for analysis (Tagliamonte 2012).

Variationists typically do not accept all types of linguistic features as sociolinguistic variables, and are often more discerning in this area than other models of style which accept a broader pool of features (Schilling-Estes 2005). Labov described the ideal sociolinguistic variable in the following terms: “The most useful items are those that are high in frequency, have a certain immunity from conscious suppression⁹, are integral units of larger structures, and may be easily

⁹ Labov explains in a later footnote that this criterion applies only for those features which speakers could totally suppress, not those that can be controlled partially: “Immunity from conscious distortion is not required [...] but if an item can be completely suppressed by most informants (such as the use of *ain't*, or taboo words) it will give us a much more limited body of data for analysis” (Labov 1966: 61).

quantified on a linear scale” (Labov 1966: 49). This led Labov to decide that “by all criteria, phonological variables appear to be the most useful” (Labov 1966: 49). In order to work as valid sociolinguistic variables, different variants for the same variable must have the same function and ‘mean the same thing’, and be mainly differentiated by the kinds of social and stylistic factors which variationists wish to study (Tagliamonte 2012). While this is relatively unproblematic for the phonological features that were the main subject of early variationism, later studies added increasingly complex and multifaceted linguistic features, such as syntax and morphology, where different variants of a construction could be argued to have contrasting functional meanings (Lavandera 1978).

Ultimately, variationists have performed credible quantitative analyses on a broad range of variables, which includes morphosyntactic variables like (ly) (variable absence of suffix –ly in English adverbs) (Wolfram and Schilling-Estes 2006) and mutation in Welsh (Ball 1984, P.W. Thomas 1984, Hatton 1988), discourse features such as the quotative (be + like) (Tagliamonte and Hudson 1999) aspect and tense variables like (going to) (Tagliamonte 1996-1998) and auxiliary dropping (P. Davies and M. Deuchar 2014), as well as code-switching (Zentella 1990). Variationists typically consider it essential that all possible variants that can be circumscribed within the variable context must be accounted for in an analysis, a process known as *the principle of accountability* (Tagliamonte 2012).

Tagliamonte describes the process for analysing the relative pronoun *who*, for instance: “The principle of accountability dictates that in addition to examining *who* itself, the analyst must also take into account all the other potential variants within the relative pronoun system. Accountability requires that all the relevant forms in the subsystem of grammar that you have targeted for investigation, not simply the variant of interest, are included in the analysis” (Tagliamonte 2012: 10). The relative use of variables by speakers or groups is then understood by taking a proportion of the particular variant from all contexts in which it could have taken

place, having removed any contexts which are categorical. The potential application of the principle of accountability for my data, and the variables that I have chosen to study, will be discussed further in the methodology section in chapter five.

2.2.3 Markers, indicators and stereotypes

Labov (1972) divides sociolinguistic variables into three types; markers, indicators and stereotypes, based upon the level of salience they possess in the minds of speakers, and whether they encode social or stylistic information, or both. Markers, the most common type of sociolinguistic variable (Bell 1984), encode both social and stylistic information. In other words, they reflect social information about the speaker – for example, their age, gender and social class – as well as reflecting information about the situation, or the way in which they are being used, i.e. the level of formality. Previously studied markers include (th), (dh), (ing), and negative concord (Labov 1972). In contrast, indicators do not reflect stylistic information, only reflecting the social characteristics of the speaker. One example of an indicator is (a:) in Norwich, which varies according to speaker class, but is always stable across styles (Trudgill 1974).

The difference between these variable types is related to the salience of a variable in the minds of speakers in a given speech community. Although indicators are stratified in terms of use by different social groups, they do not have sufficient salience to be sensitive to changes in formality. Markers, on the other hand, are salient enough to be unconsciously adapted by speakers to changes in situation. While both markers and indicators are described as being below the level of consciousness¹⁰ a third type of sociolinguistic variable – stereotypes - have risen above the level of consciousness to the extent that members of a speech community are

¹⁰ Although both types of variable are defined as being below the level of consciousness, the crucial distinction between them lies in the fact that community members react to markers “in a uniform manner”, although “without necessarily being aware of it” (Labov 1972: 178). Kiesling (2009: 192) describes this kind of unconscious style-shifting as being an automatic response akin to catching a ball, as opposed to a planned and intentional decision. Stereotypes by contrast may involve more deliberate, conscious usage (see below).

fully aware of them¹¹. These variables encode styles more reliably than group membership¹², and are “the kinds of features that make it into the Letters to the Editor section of local papers” (Meyerhoff 2011: 26).

Labov presents the three variable types as part of a diachronic model of sound change, with the indicator being the earliest stage in the life of a sociolinguistic variable, and the other types forming later as their prevalence and use in the speech community increases. *Indicators* form as sounds which are associated with a “subgroup”, which come to operate as “a function of group membership” (Labov 1972: 178). The sound then spreads beyond the group where it began into the entire speech community where it “became one of the norms which defined the speech community, and all members of the speech community reacted in an uniform manner to its use [...] The variable is now a *marker* and begins to show stylistic variation” (Labov 1972: 179). The variable can finally become a *stereotype* when it becomes stigmatized enough to be the subject of “overt social comment” (Labov 1972: 180), and may eventually disappear. Unlike markers and indicators, stereotypes are not accepted by some sociolinguists. Bell (1984) for example, considers all stylistic variation to be a reflection of social variation, and thus stereotypes, being untethered from social variation, are redundant in his model.

2.2.4 Inter-speaker variation

One of variationism’s primary characteristics is that it locates style within broader social structures, including processes of change. This makes discussing the variationist approach to style difficult without also describing how variationists have studied the social axis – that is, variation between different kinds of speakers within a speech community. Labov (1972)

¹¹ One potential example of a Welsh stereotype is the vulgar address form *cont*, “a marker of solidarity among young Welsh-speaking males” (Williams 2009: 82) which is associated with the town of Caernarfon. Williams describes the apparently conscious use of the word by an outgroup non-Welsh speaking young man in order to try and “achieve acceptance of his identity as Welsh” in a rather tense inter-group context.

¹² Stereotypes are, as the previous footnote suggests, often associated with group affiliation, but in a less systematic, more abstract way than indicators and markers. Thus Honey (1997: 99) describes them as (my emphasis) “popular and conscious but *imprecise* general characterizations of the speech forms of particular social groups”.

considered the speech community to be the primary unit of study in variationist methodology, from which participants are drawn and in the context of which linguistic processes of variation and change are understood. An early definition of the speech community by Bloomfield defined it as “[...] a group of people who interact by speech” (Bloomfield 1933: 42). Labov adjusted the emphasis so that a speech community could be primarily defined as a social group which shares the same cultural evaluation of linguistic features: “the speech community is not defined by any marked agreement in the use of language elements, so much as by participation in a set of shared norms” (Labov 1972: 120-121). For example, New Yorkers showed predictable style shifting towards prestige variants of (r) in Labov’s (1966) study, because all members of the community shared a similar evaluation of the social meanings of both variants.

2.2.4.1 Change from above and below

Shift in the use of variants is typically described as occurring in two main ways – *from above* and *from below* the level of consciousness (Labov 2001.b). *Change from below* the level of consciousness involves the transmission of linguistic features from adults to children, where “children advance a linguistic change to a level beyond that of their caretakers in the same direction over successive generations” (Tagliamonte 2012: 58). *Change from above* the level of consciousness involves diffusion, where new variants are introduced into a community through face to face contact with speakers from outside that community.

One of the most important differences between transmission and diffusion is the gap in ability to acquire language between children and adults. Whereas children can effectively acquire any type of linguistic feature to which they are exposed in the vernacular, adults have difficulty in acquiring more complex phonological and grammatical constructions, which means that such features are more difficult to diffuse: “It follows that structural patterns are not as likely to be diffused because adults do not learn and reproduce linguistic forms, rules, and constraints with the accuracy and speed that children display” (Labov 2007: 349). Kerswill (1996) has proposed

a hierarchy of linguistic features, with the most difficult to acquire at the top, and the easiest at the bottom:

Rank	Linguistic feature	Age of Acquisition
1 (most difficult)	i. Lexically unpredictable phonological rules	By (3) ?
	ii. New phonological oppositions	By 3-13
	iii. Grammatical change; parameters	By 8 ?
2	iv. Prosodic systems	Adolescence? Lifespan?
3	v. Grammatical change	Not before 4-7; then lifespan
4	vi. Morphologically conditioned changes	Lifespan
5	vii. Reassignment of words or lexical sets to other morphological classes	Lifespan
6	viii. Mergers	Lifespan
7	ix. Neogrammarian changes	Lifespan
8	x. Lexical diffusion of phonological changes	Lifespan
9	xi. Borrowing , new lexical forms of old words new phonetic forms of existing morphological categories Borrowing, vocabulary	Lifespan

Table 2 Hierarchy ranking the difficulty of acquisition of linguistic features (Kerswill 1996, Table 4)

As the table indicates, simple lexical features like borrowings and vocabulary are diffused more easily than complex grammatical and phonological constructions, while the most complex features (i.e. those towards the top of Kerswill's table) can only be acquired by children before a certain age.

The necessity of face to face contact for diffusion to occur is generally accepted by most variationists. This means that the spread of variants is constrained by the physical movement

and interaction of speakers. This has also lead most variationists to strongly reject the idea that variants can be diffused between geographically discontinuous areas through the broadcast media or any other means which does not involve face to face interaction, although other researchers have recently begun to question this consensus (Stuart-Street 2011; Sayers 2014). An exemption is generally made for lexis however (Trudgill 2014), which as Table 2 shows is the most easily diffused class of linguistic feature.

2.2.4.2 Patterns of inter-speaker variation

Variationist studies have found patterns of inter-speaker variation in communities which tend to recur. These include, stability, generational change, lifespan change and age grading. Change in speech communities is usually measured using either apparent time, or less frequently, real time methods.

2.2.4.2.1 Apparent time and real time

The central method used by variationists to measure the process of change is *apparent time*, which allows researchers to measure diachronic change using synchronic data. The first to use the apparent time technique was William Labov in his seminal Martha's Vineyard study, where he argued that differences in the use of linguistic features between age groups indicate actual change in the use of these features over time (Labov 1963). This means that if the use of a non-standard form of a feature is more frequent amongst younger speakers in a community than among older speakers, a researcher can infer that a shift towards the non-standard form is occurring across the whole speech community. Apparent time methods can be contrasted with those of a real time analysis (e.g. Cukor-Avila and Bailey 1995), where change over time is measured with data actually taken from actual multiple sequential different points in time, rather than inferring change from differences between age groups within the same data set.

The use of apparent time as reliable reflection of real inter-speaker variation is reliant on the notion that there is a critical period (Lenneberg et al 1967) after which speakers are not able to

adapt their use of grammar and phonology easily, even if they move to other communities (Labov 2001.b, White and Genesee 1996). As speakers' use of such features are argued to remain relatively constant after adolescence, researchers argue that the language use of a given age group can be taken to reflect how they spoke during their formative years, thus allowing linguistic variation over time to be mapped.

2.2.4.2.2 Stability

Some variables are stable in a community, and will generally not show evidence of change over time. A well-known example in English is the variable (ing) which varies between [ɪŋ] and [ɪn]. According to Houston (1985), (ing) has existed in a stable condition in English speaking communities for a period of centuries. Although (ing) is considered stable, in the sense that both variants have persisted in use for long periods, the use of the variable by individuals shows strong variation according to social class and style, as (ing) is a marker.

2.2.4.2.3 Generational Change and Lifespan Change

Generational change involves change where one variant of a variable becomes more frequently used than another in each subsequent generation. Generational change generally does not involve a change in individual's way of speaking, but gradual change in the frequency of use of a particular variant in the vernaculars of speakers from each successive age group. This can be contrasted with lifespan change, which involves both gradual change across generations as well as change in individuals, as speakers "change their own speech after the critical period, but in a manner that is consistent with the gradual change taking place in the community as a whole" (Meyerhoff 2011: 152).

An example of generational change is the shift in the Montreal French speaking speech community in their use of the variable (r), from the alveolar variant [r] to the uvular variant [R] between 1971 and 1995, where speakers in each successive generation used greater frequencies of [R] but did not adapt their own speech significantly during their lives (Sankoff and Blondeau

2007). Meyerhoff (2011) argues that Martha's Vineyard - which was the focus of Labov's classic (1963) study - is an example of a community undergoing *lifespan change*, particularly in light of Pope et al's (2007) later reproduction, which provided a real time aspect to the study. Pope's analysis, which used the same methods as Labov's original study forty years after the original, showed that older speech community members who were alive at that time themselves had significantly increased their use of those variants, including centralized onsets, towards which the community as a whole had been moving.

2.2.4.2.4 Age Grading

Another pattern of variant distribution found in variationist studies is age grading, where a linguistic variant is used in a recurring way by specific age groups each generation, rather than gradually changing in use over time. This notion was introduced by Sankoff and Laberge (1978), who observed a recurring pattern where the standard use of certain variables reached a peak in speakers in their early twenties, and would begin to decline in late middle age. The researchers drew upon Bourdieu's (1991) theory of *the linguistic marketplace* to explain this pattern. According to this analysis, as middle class society is the gate keeper of power in capitalist society, the linguistic variants used by these speakers acquire symbolic capital. Because of this, when young people come into competition for work, they necessarily focus their attention upon these linguistic features in order to improve their chances in the market. Conversely, as they become older, the same speakers then cease their use of these variants as "social pressures reduce as people move out of the workforce and into a more relaxed phase of their life" (Tagliamonte 2012: 47).

An alternative pattern of age grading, which is the inverse of that mentioned above, is for adolescents to use stigmatized lexical forms such as swear-words and slang during their teen years, before ceasing their use as they get older, and these transient forms pass out of fashion (Wolfram and Fasold 1974). Despite the fact that they encompass both prestigious or

stigmatized forms, features which are involved in age grading phenomena have been described as having a specific linguistic profile. Labov (1994) states that a crucial attribute of age graded variables is their high level of salience in a speech community. Due to this they are considered more likely to be consciously controlled by speakers, hence their quick adoption and then abandonment by successive age groups. The classic profile of age grading in an apparent time study is an inverted *u* or *v* shape (Tagliamonte 2012).

2.2.4.3 Gender and Variation

Numerous sociolinguistic studies have discovered significant differences in the way that men and women use standard and non-standard variants. Labov describes this in terms of the *gender paradox*, where “while women conform more closely than men to sociolinguistic norms that are overtly prescribed, they conform less than men to variants that are not overtly prescribed (change from below), but which are innovative” (2001.b:291). In other words, in change from above, or in stable situations where prestigious forms are widely recognized by speakers, women usually lead in the use of standard variants. However in change from below, where change is taking place below the level of consciousness, women usually lead the adoption of innovative forms. One explanation for this is that men are simply less sensitive to variation in language than women, or have an inferior linguistic ability (Labov 2001.b). The phenomenon has also been tied to dynamics of the ‘linguistic market’ (Sankoff and Laberge 1978) and power dynamics between men and women. Eckert (2000) describes women as having less power than men in general, and thus being more reliant on mastering the use of linguistic capital as a compensatory strategy.

2.2.4.4 Class and variation

Not all speakers possess comparable repertoires, and crucial to Labovian variationism is the idea that *social class* mediates speakers’ access to and use of vernacular and standard varieties of a language: “the social class to which we belong imposes some norms of behaviour on us

and reinforces them by the strength of example of the people with whom we associate most closely” (Chambers 1995: 7). As mentioned above, variationist theory stresses Bloomfield’s (1933) principle of density, where frequency of interaction between speakers deterministically shapes speakers’ language through an automatic process of synchronization. As Labov states: “the more often people talk to each other, the more similar their speech will be” (Labov 2001.b: 228) and this reduces the question of the acquisition of linguistic features among a population to “a simple calculation” (Labov 2001.b: 19) of what kind of speakers interact most frequently. A speaker whose primary interactions are restricted to a defined social circle will thus predominantly reflect the vernacular of that group, forming a kind of “barrier” to the acquisition of other forms (Trudgill 1992). Thus, within a given speech community Labov argued that a speakers’ position in terms of class placed them passively within a social structure which determined their level of exposure to prestige (or standard) variants of certain meaningful linguistic features (Eckert 2005). Each speaker’s vernacular style – that is, their most unmonitored ‘natural’ style – varies in its relative starting position from idealized standard forms, as well as the stylistic distance to which they are ultimately able to travel.

2.2.4.5 The Social Network Approach

Milroy and Milroy (1992) have criticized the roots of Labov’s passive depiction of vernacular speakers as belonging to the functionalist view of sociologist Talcott Parsons’ consensus society (Parsons 1975), which presents society as being fundamentally cohesive. Subsequent theorists have instead argued for a view of social relations based around conflict, rather than consensus. In place of Bourdieu’s (1991) notion of a linguistic market, where all speakers in a community agree on the legitimacy of standard forms, and the vernacular only emerges in the absence of standard ideology pressures, Milroy and Milroy (1992) promote Woolard’s (1985) notion of alternative linguistic markets in which vernacular forms can be imbued with positive meaning by groups opposed to state authority. Milroy’s social network (1980) approach

provides one example of an attempt to balance the unidimensional focus of early variationism. Rather than conceptualizing speakers simply regressing to vernacular in the absence of state ideological pressure, Milroy's study found that tight-knit social networks played a strong role in actively maintaining non-standard forms against pressure from standard English.

Using a 'friend of a friend' method based around collecting data from participants within intersecting social networks, Milroy was able to show that dense multiplex networks in working class areas act to reinforce the use of vernacular variants, and that individuals' use of non-standard forms could be predicted by the density of these networks. This phenomenon reflects previous sociolinguistic insights (Labov 1966, Bloomfield 1933) on the importance of frequent face-to-face communication (and thus input) in shaping the variants that speakers use. Chambers (2003) has described the role of the social network as a "norm enforcement mechanism", which influences speakers to retain vernacular speech norms (Chambers 2003: 75). This conceptual approach provides an explanation for the way in which non-standard models of language can resist the influence of standard language ideology. The social network methodology was at least partly developed initially because of Milroy's dissatisfaction with social class as a grouping unit for speakers, as she had found that the variation in her data could not be accounted for by "corporate group membership" (Milroy 1980: 36). Milroy has stated that one of the main advantages of the social network concept is that it can explain data even in situations where class is not relevant in the social context. For this reason, many input oriented sociolinguistic Welsh studies (Ball 1984, R.O. Jones 1976) have preferred to draw on the notion of social networks as an alternative to class to explain linguistic variation.

2.2.5 Criticism of the Variationist approach to style

Although the variationist paradigm has clearly produced a large amount of seminal work, certain aspects have been questioned and problematized since the seventies, particularly by

researchers who take more speakers centred approaches (e.g. Coupland 2007, Eckert 2001). Criticism has centred on the central mechanism of attention to speech, which some argue ignores the dynamic and initiative driven nature of style. Although most theorists accept that attention to speech is a factor in variation, many have argued that it may not deserve the central position that it holds in Labov's approach. Eckert claims that attention paid to speech does not automatically lead to more prestigious styles, as she has "many examples of vernacular variants being used quite consciously and emphatically" (2001: 123) and this observation has been paralleled by other researchers (Coupland 2007, Bell 1984).

Another, related, criticism is that the attention to speech model reduces style shifting to a one-dimensional plane, depicting speakers as primarily shifting in one direction – always towards standard forms - in response to increasing saliency of standard language ideology norms, as personified by the outgroup academic researcher. Some researchers have pointed out that this model removes any agency on the part of speaker, as Labov's style shifting is fundamentally negative, non-creative phenomenon, and essentially represents a kind of instinctual response to oppression. Coupland (2007) evocatively sums up the discomfort that some sociolinguists feel towards this model:

"To me, Labov's style shifting speakers suggest a shoal of swimming fish, grouped together in a social bundle, who suddenly veer away together in a single new direction when they recognise the presence of a predator. If the shoal of fish is an aggregation of working-class speakers, their style-shifting might be a culturally predictable veering towards 'safer waters', when threatened by the ideological predator – the establishment's demand that public speech should be 'more correct'" (Coupland 2007: 39).

Although Labov (1966) himself proposed notion of *covert prestige* as a countervailing force to the influence of the standard language, which was explored further in Trudgill (1974), the emphasis in these early studies was still mainly on normatively driven, rather than creative, style shifting. This is perhaps partly because the attention to speech methodology itself has a tendency to reproduce this relatively narrow aspect of style. Because of this, developments towards other, less constrained conceptions of style arguably only developed when researchers experimented with the alternative, more flexible methodologies, which I will describe in the following sections.

2.2.6 Variationism – Summary

Variationism has played an important role in the development in the field of sociolinguistics, providing an easily replicable methodology and crucial insights into sociolinguistic stratification and the interaction of inter- and intra- speaker variation. Some of the defining attributes of the paradigm, which may render it distinct from other models are:

- the focus on attention to speech, and the use of a methodology based around this principle to elicit different styles
- the macro scale approach which demands large amounts of data from multiple speakers
- The initial focus on the speech community (and social networks) as the most appropriate setting for research
- The foregrounding of speaker input as the main factor which shapes speakers' use of language, subsequently minimizing speaker choice and initiative as factors
- the relatively stringent constraints upon what may be considered a sociolinguistic variable (Tagliamonte 2012)
- the use of written tests and materials to elicit styles, as opposed to using naturalistic data (Labov 1966)

2.3 Communication Accommodation theory

Communication Accommodation Theory (CAT) was developed by Giles and Powesland (1975) and associates in order to provide an alternative explanation model for intra-speaker variation as opposed to attention to speech (Labov 1972). CAT posits that rather than merely applying greater or lesser degrees of attention to their speech, speakers vary their speech strategically, and intentionally, in order to adapt to different kinds of interlocutors. Although CAT was initially conceived as Speech Accommodation Theory, it was rebranded as CAT in the late eighties to reflect the insight that accommodation is not restricted to language, but also involves paralinguistic features like gesture, posture and so on. The core of the theory involves *convergence* and *divergence* (Giles et al 1991). Convergence, the most typical interactional strategy, involves speakers adapting their speech to be more similar to that of their interlocutors. The driving force behind this behaviour is *similarity attraction* (Giles and Powesland 1975), a psychological mechanism which means that people generally evaluate those who speak similarly to themselves more positively. Effective convergence can often lead to compliance, in the sense that interlocutors may be more likely to respond positively to the speakers' requests. For this reason, convergence can be a particularly salient patterns in service contexts (Coupland 1984), where one person is trying to gain compliance from another. Another factor is *communicative efficacy* (Giles et al 1991), which involves speakers simplifying or otherwise adapting their speech to make it more comprehensible to interlocutors, and this phenomenon is brought to the fore in styles like foreigner talk (Ross and Berwick 1992) and baby talk (Greenbaum and Cooper 1988). CAT theorists generally consider convergence the default interactional strategy which is typical of the great majority of social interactions (Giles et al 1991).

One study which focused on convergence is Coupland's (1984) analysis of a travel agent's interactions with customers in the Welsh capital city, Cardiff. Coupland recorded fifty one

clients speaking with a female travel agent on different occasions, and the participants were categorized as belonging to one of six social classes. Four English phonological variables, considered sociolinguistic markers in the Cardiff area, were analysed in the speech of each speaker, as well as in the speech of the travel agent. Coupland found that the variants used by the travel agent in interactions with different speakers reflected relatively accurately both the social background and level of education of each interlocutor: “The percentage of variants in Sue’s speech provide a reasonably reliable index of the socioeconomic status and educational background of her interlocutors, just as the percentages of those forms in the clients’ own speech are able to do” (Coupland 1984: 61). Coupland argued that the travel agent was purposefully converging upon the speech of each client, within the limits of her ability, due to her strong motivation to succeed in her work and to please the customers. In other words, her behaviour effectively harnessed the psychological phenomenon of *similarity attraction* to gain *compliance*.

Divergence involves speakers intentionally adapting speech to become less similar to that used by an interlocutor. This behaviour is often associated with antagonistic situations involving intergroup tension, since divergence is normally evaluated very negatively by interlocutors (Putman and Street 1984). One well known study (Bourhis and Giles 1977) involved a group of Welsh learners who were asked to listen to a recording of an “ethnically threatening” RP speaking Englishman, who spoke in disparaging terms about the Welsh language. When asked to respond to the message, participants who were learning Welsh for integrative (i.e. emotional and cultural) reasons increased their use of Welsh phonological features, and even code-switched into Welsh, in order to distance themselves from the speaker. The same experiment was later repeated by the researchers in the Belgian context, with Flemish speakers diverging from an “ethnically threatening” Walloon speaker primarily through the use of code-switching from French into their native Flemish (Bourhis et al 1979). Besides convergence and

divergence, a third strategy is *complementarity* (Giles et al 1991), where speakers maintain a distinct linguistic pattern from their interlocutor due to the entrenchment of certain cultural roles. Unlike in the case of divergence, speakers can often evaluate complementarity positively. One example of this pattern is the medical context, where the power disparity between doctor and patient may lead to the adoption of highly distinct speech patterns. Interlocutors largely accept the divergent style used by doctors in such contexts because of the established role of medical professional as a prestigious expert (Street 1991).

2.3.1 Communication Accommodation Theory - Summary

CAT contrasts with variationism in some key ways - it is speaker centred, and emphasises the active role of the speaker in evaluating social situations (and particularly the attributes of interlocutors) and in adapting their speech strategically. The most notable features of CAT are listed below:

- A focus on identity and intentionality as central factors in stylistic variation
- The reduction of the role of *attention to speech* as a factor in style shifting
- The use of social psychology methodology in studies, with less of a linguistic focus
- Studies are typically focused on individuals and smaller groups
- The tendency to use naturalistic data, rather than elicitation tasks

Since CAT is not really a linguistic theory, but a social psychological one, it has been criticised for its linguistic naivety, particularly because the majority of early studies did not involve quantitative linguistic methodologies (Trudgill 1981). In the next section I will discuss Bell's audience design approach, which provides a useful 'bridge' between the variationist and CAT models.

2.4 Audience design theory

Like accommodation theory, *audience design* (Bell 1984) was purposefully drawn up as a response to issues seen as problematic in Labov's *attention to speech* paradigm. There are several important areas in which the model contrasts strongly with the variationist paradigm. Firstly, it offers an alternative fundamental mechanism as a driver for style shifting – strategic convergence towards interlocutors rather than attention to speech. Bell argues that the frequently discovered correlation of increased attention with formality is not causative, and that speakers can focus attention towards producing any style – whether or not it is a ‘prestige style’ or ‘vernacular style’ (an observation also made by Eckert (2001)) and that attention is a secondary mechanism. Secondly, audience design is dynamic - rather than adapting their speech unidimensionally towards a global notion of prestige, audience design presents the speaker as being able to ‘design’ their utterances for numerous different audiences. Thirdly, audience design differs from variationism in its explanation for the emergence of styles. Rather than speakers forming psychological associations between linguistic features and different kinds of emotional experiences during childhood (Labov 2001.b), Bell argues that style derives from inter-speakers variation, which is to say that styles are fragments of the vernaculars of different kinds of speakers within a given community.

2.4.1 Audience design

Bell splits the audience design model into two categories of style shift, depending on whether speakers are responding directly to an interlocutor's characteristics or to those of an idealised, absent ‘referee’. The former category of shift is *audience design*, which is generally analogous to convergence in the CAT model, in that it involves speakers adapting their speech to become more similar to their audience. In audience design, style shifting is responsive as opposed to initiative driven, and is based around an evaluation of the interlocutors attributes: “It assumes

that persons respond mainly to other persons, that speakers take most account of hearers in designing their talk” (Bell 1984: 159). The effect of an interlocutor on the speakers’ stylistic choices is dependent on their proximity and role in an exchange. Bell proposes a hierarchy of audience roles based upon whether individuals present in an interaction are “known, ratified, or addressed by the speaker” (Bell 1984: 159). *Addressees* are known, ratified and addressed, *auditors* are known and ratified (but not addressed), *overhearers* are known (but not addressed or ratified), and *eavesdroppers* are neither known, ratified nor addressed. The effect of each type of audience on the speaker varies according to their level on hierarchy, with *addressees* having the most influence on the speaker and *eavesdroppers* having the least.

According to audience design (as well as accommodation theory), speakers take cues for style shifting from the attributes of their interlocutor, although how this takes place is not entirely understood. Bell (1984) lists some of the ways in which this could work, including adapting speech in response to an appraisal of an interlocutor’s social attributes (i.e. their age, gender, clothing etc.) or actively observing and mirroring their objective speech patterns. There is evidence that an interlocutor’s social attributes, and the speaker’s stereotyped appraisal of those attributes, have a greater influence on style shifting than the interlocutor’s true linguistic patterns. A study by Beebe (1981) found that speakers converged on particular stereotypical ethnic variants that they believed their interlocutor was using - even when the interlocutor did not actually use these variants - solely because of their social attributes (e.g. their appearance). This phenomenon has been termed “subjective linguistic convergence” (Thakerar et al 1982: 238).

2.4.2 Referee design

The second category of style shifting in the audience design model is *referee design* (Bell 1984). Referee design contrasts with audience design in that it involves initiative style shifting, with speakers deliberately shifting towards the style of an absent ‘referee’ rather than

converging upon the speech of a physically present interlocutor: “The effect of referee design is to make the speaker style-shift as if actually talking to the referee rather than to the addressee” (Ball 1984: 186). Referee design is divided further into two subcategories, based upon whether the speaker considers themselves part of the group that uses the referee’s code or not. Ingroup referee design is similar to divergence in the CAT framework, with speakers shifting towards the speech of their ingroup, and thus purposefully declining to perform the normative act of convergence upon their interlocutor. Bell (1984) invokes the Welsh learners shifting towards Welsh phonology and lexis in Bourhis and Giles (1977), as well as a schoolchild in Cheshire (1982) who deliberately used vernacular forms when speaking to a teacher in order to signal displeasure with school norms, rather than converging on the teacher’s standard English as the other boys did.

In outgroup referee design, speakers shift towards a style of speech which belongs to some other group, but which holds substantial prestige for the speaker, usually for specific cultural and historical reasons. Outgroup referee design can be either short term or long term. In short term outgroup design, a speaker style shifts towards a prestigious code for short term benefit, such as winning an argument. Bell notes that this kind of style shift is necessarily rare, because it can alienate the interlocutors and “violate the norm of conversation between intimates, or in the extreme redefine the relationship as no longer intimate” (Bell 1984: 188). Long term outgroup referee design involves an institutionalised tendency for speakers to style shift towards an outgroup’s code, something which Bell specifically associates with diglossia (Ferguson 2007). In diglossic situations a prestigious linguistic variety is used in the society in high domains, but “is not native to any group in the speech community. Rather, it is the dialect of an external referee, distanced by either space or time” (Bell 1984: 189). As mentioned above, examples of diglossic societies are common in sociolinguistic literature, with Ferguson (2007) himself using the Swiss German and Haitian Creole speaking communities as examples of

communities using a geographically distant community's code, and Classic Arabic and Classic Greek speaking communities as examples of those who used 'fossilized' linguistic varieties of previous eras as their H code.

2.4.2.1 Referee design and the media

Bell identifies two broad situations in which speakers frequently use initiative and create a style which is based on idealized speech, rather than actual features of their audience. These are in the mass media, such as on radio or television, and in service encounters, such as in shopping centres. Bell claims that "all media language is initiative style design", and that this "holds supremely for radio" (Bell 1984: 192) because of the lack of feedback involved in presenting. Radio presenters have only their voices and patterns of speech at hand to create a symbolic relationship with their audiences, and this means that associations between speech style and identity are foregrounded. Bell (1984) categorises media language generally as belonging in the ingroup referee design category, as radio presenters (for instance) typically try and attune their language towards a particular audience to take advantage of the psychological mechanisms of similarity attraction and compliance. Indeed, Bell's seminal (1977) work, which I discuss further below, has shown that radio presenters often seem to converge upon the vernaculars of the social groups who are perceived as the programme's ideal target audience. But Bell also suggests that radio can involve outgroup referee design. One example of this is the use of received pronunciation on New Zealand radio in prestigious radio programmes (Bell 1982), where Bell describes New Zealanders looking to standard British English as a model of appropriate formal style for news reading.

2.4.3 Stylistic variation reflects social variation

One of Bell's most important theoretical contributions is his insight that variation on the social and stylistic axis in a speech community is always connected. While previous variationist

researchers had noted that both axes were connected (Romaine 1980), Bell was the first to state that this is because variation along the stylistic axis actually *derives* from variation along the social axis.

“In language acquisition, children learn the linguistic range they hear from speakers in their environment. Children first acquire the vernacular through contact with family and peers. As their range of interlocutors broadens to the wider speech community, they learn the wider range of styles there” (Bell 1984: 158).

According to this viewpoint, styles reflect fragments of different social groups’ vernaculars, which become associated with different values based on a process of social evaluation of those groups. Speakers thus base their notions of formal and informal speech on the social characteristics of the people who generally use particular variants. This principle is known as the *style axiom*: “variation on the style dimension within the speech of a single speaker derives from and echoes the variation which exists between speakers on the ‘social’ dimension” (Bell 1984: 151). For example, in a community where a given variable is stratified socio-economically, so that speakers of a higher social class use a particular variant more frequently than speakers from lower social classes, the axiom predicts that speakers in the community will use the ‘upper class’ variant more frequently in formal situations, because formal situations are associated with prestige, while the opposite will hold for the ‘lower class’ variant. Bell also claims that social variation in this case can involve variation in age, gender and other categories, so that speakers take their cues for the development of a sociolinguistic repertoire from observation of many different social categories, not only socio-economic ones. Thus, Bell speculates that “the very strong pattern of co-occurrent interspeaker and intraspeaker variation suggests a strong hypothesis: a sociolinguistic variable which is differentiated by certain speaker characteristics (e.g. by class or by gender or by age) tends to be differentiated in speech to addressees with the same characteristics” (Bell 1984: 167).

2.4.4 Empirical evidence for audience design

The initial empirical evidence for the audience design model came from Bell's work on New Zealand radio speech (1977). Bell found that news stations in the country operated a unique system where the same DJs performed broadcasts in multiple stations. A single DJ would frequently read news bulletins for two different stations in the same recording booth, sometimes with as little as ten minutes between them. Bell designed a study which collected seventeen hours of speech from five different stations and from between four and nine DJs for each one. His prediction was that the speech of the DJs would shift according to the imagined attributes of the target audience demographics for each station. Information about the audience for each station was collected through a random sample in the Auckland area, which confirmed Bell's initial characterization of the intended target audience of each one.

The survey questioned 586 respondents on their preferred radio stations and programmes, as well as for their age, education and occupation. An analysis of the data revealed that each radio station could be reliably differentiated into three groups according to the ages and occupational level of their listeners. In the first group was the National Radio station, which had an audience which was generally older and of a higher socioeconomic class. The second group had two stations with 'middle ranking' audiences, while the two stations in the third groups had the lowest socioeconomic levels and the youngest audiences. The most compelling results of Bell's study come when he examines the speech of five newscasters who were recorded on pairs of radio stations which had distinct audiences. The newscasters showed "a remarkable and consistent ability to make considerable style shifts to suit the audience" (Bell 1991: 120) in their usage of the variables /t/ voicing and /t,d/ deletion.

An early example of an attempt to apply an audience design methodology to face- to-face conversations was Rickford and McNair-Knox's (1994) 'Foxy Brown' study. The study was

part of a longitudinal research project on the language of a young African American teenager, Foxy Brown, who was eighteen at the time, but who had been recorded at intervals since she was thirteen. In order to test the theory that Foxy would style shift on the basis of the interlocutor's attributes, the researchers compared an interview in which she had spoken with an African American woman with whom she was already familiar, with one in which she spoke with a European American woman who she had never met. The second interview was recorded a year after the first, and the European American interviewer was able to bring up the same topics as the African American interviewer, so that the effect of topic could also be measured. The analysis compared the frequencies of several features which are considered typical of African American Vernacular English (AAVE), including zero copula, invariant be, Plural –s, third singular present –s, and possessive –s. The researchers claimed that their findings confirmed many of Bell's predictions about the working of Audience Design. Although there was significant variation between topics, particularly between 'school and career' and 'wives and slamming partners', Foxy used AAVE features consistently more often for each topic with the African American interlocutor than with European American interlocutor.

Bell has presented a template for a larger scale study based around audience design which others might replicate (Bell 2001). In Bell's study, four participants who varied in ethnicity and gender were interviewed successively by three different interviewers, who also varied in ethnicity and gender. The participants were a *Pakeha* (white New Zealanders of European origin) woman and man, and a Maori woman and man. Bell states that a tightly structured interview model was used in order to make the results comparable, with the same topics being asked of all respondents: "the attempt to elicit maximally informal speech was sacrificed to some extent by the need to ensure comparability across the interviews, e.g. by topic, and even by framing" (Bell 2001: 150). Although this sacrificed the possibility of recording completely 'natural' speech, Bell claims that it "paid dividends in enabling us to compare how different

informants presented their opinions and identities in response to the same questions” (Bell 2001: 150). Twelve interviews were collected which were each one hour in duration. As well as the interviews themselves, participants also read word lists which contained doublets – words that could be pronounced in either a Maori way or a *Pakeha* way - including a list of New Zealand place names.

One of the findings was that the two Maori speakers tended to use the interactional marker ‘eh’ significantly more frequently with other Maori interlocutors, which, according to Bell, they were using as a Maori marker of identity for solidarity purposes. In the word list section the participants pronunciation patterns seemed to reflect their attitudes towards the various interviewers, but also towards the wider ideological and social issues that they indexed. For example, a Pakeha male, who disliked the Maori, pronounced all doublets in Standard English, which reflected his stated antipathy towards Maori culture. The Pakeha female however, who had more positive feelings towards Maori culture, pronounced the Maori versions of the doublets more frequently, as well as making several meta-comments apologizing for her inability to pronounce certain words correctly. Bell considered this behaviour audience design, as the speaker was attempting to accommodate - within the limits of her ability - towards her Maori interlocutor in order to show cultural sensitivity.

2.4.5 Summary of Audience Design

Bell’s audience design model is credited with providing the insights of accommodation theory with more rigorous linguistic grounding. He has also proposed a potentially useful constraint on style with his style axiom – socially and linguistically diverse communities are likely to have a greater range of potential styles than those which are not, which has possible repercussions for communities that are undergoing declining ethnolinguistic vitality (Dorian 1994). However, while Labov’s classic (1966) methodology, for all its potential flaws,

provides a simple and easily replicable study design, there is no comparable approach for accommodation and audience design theories which can provide such effective and universally relevant data. Additionally, Bell's theory on the roots of stylistic variation is not universally accepted. Labov has argued, contra Bell (1984), that for younger speakers in the crucial age of style acquisition, "it is not likely that the distinction between less formal and more formal speech stems from an association with the way people of different social classes speak" as "their experience is focused on the home and pre-school play context" (Labov 2001.b: 421). The Audience Design model of style can be summarized in the following way:

- Portrays inter-speaker variation as the source of speakers' stylistic repertoires
- As in accommodation theory, stylistic variation is considered intentional and strategic, with a social psychological underpinning
- Style is considered primarily responsive, although the notion of referee design provides the model with an initiative driven component
- Uses both quantitative and qualitative methods, with a greater emphasis on naturalistic speech than elicitation

2.5 Speaker design approaches

Speaker design is a label applied by Schilling-Estes (2005) to encompass a variety of approaches which see speakers as able to actively create styles, rather than simply reflecting back other speakers' choices (Giles 1972, Bell 1984), or focusing attention to move along a linear prestige-stigma scale (Labov 1966). Such approaches have opened up the possibilities of style, challenging previous notions of authenticity, constrained repertoires and standard/non-standard dichotomies (Coupland 2007), and foregrounding the local, experienced reality of speakers as crucial to explaining how social meaning develops (Eckert 2005). In this section, I will explore some aspects of what might be considered two parallel streams of current research

in this paradigm. The first has grown out of Eckert's (2000) and Ochs' (1992) work, mixing insights from anthropology with variationist theory and methodology, and introducing the now crucial concepts of *stance*, *indexicality* and the *community of practice* to sociolinguistics. The second includes the work of Coupland (2007) and Rampton (1995), who draw on Bakhtin's (1981) notion of stylization to analyse performative speech, and critique notions of 'authentic' speech.

2.5.1 Indexicality and stances

A significant development in sociolinguistics has been the growth of a focus on indexicality and stance taking as central elements of style. Kiesling defines an index as "a type of linguistic (or other) sign that takes its meaning from the context of an utterance, with context understood fairly broadly, including aspects of the speaker, hearer, and speaking situation" (Kiesling 2009: 177). Stance can be defined as a "person's expression of their relationship to their talk (their epistemic stance—e.g., how certain they are about their assertions), and a person's expression of their relationship to their interlocutors (their interpersonal stance—e.g., friendly or dominating)" (Kiesling 2009: 172). The connection between indexicality and stances was developed by Ochs (1992), who emphasised the fact that linguistic features only indirectly index social groups. For instance, although a given feature might seem to index maleness or femininity, this is rarely an exclusive relationship. Ochs provides the example of the Japanese sentence final particles *wa* and *ze*, which are often seen as indexing female and male social categories respectively. In fact, Ochs argues, these features first index stances and social actions - *wa* indexes a stance of gentle intensity, while *ze* indexes coarse intensity. Both particles only become associated with a given gender because ideological frames of appropriate femininity and masculinity promote certain stances in terms of gender complementarity, i.e. women are taught to prefer politeness strategies, while men may be influenced by ideologies of hegemonic masculinity to use more confrontational and 'unrefined' language.

Following Ochs, Kiesling has argued that because stances and social actions come before the development of social group indexes, stance taking could be regarded as the fundamental primitive of stylistic variation. Furthermore, Kiesling, like Labov (2001.b), refutes Bell's (1984) idea that speakers learn styles from community variation, arguing that "the main differentiation a child is likely to hear is not among many different speakers, but different ways of interacting among the relatively few speakers (compared to adults) with whom she or he is in direct contact in early childhood" (Kiesling 2009: 176). By way of example, Kiesling discusses a study by Smith et al (2007), where the researchers found that young children's use of local Glasgow variant (ou) was conditioned by the salient variable of age, and was also sensitive to different activity styles (e.g. playing/teaching), but was - unlike in the case of adults - not differentiated by speaker gender. This leads Kiesling to argue that the stance indexicalities of linguistic features - in this case, 'serious' versus 'playful' language - develops prior to social group indexicalities such as gender.

2.5.2 Style as practice

Eckert (2000) emphasizes the fact that style is a practice rather than a 'thing'. It involves *bricolage*, where speakers combine features based upon their interpretation "of the meaning potential of the available resources" (Eckert 2003: 43). In other words speakers are able to pick and choose from among various linguistic features to create a personal style which conforms to their own value systems. Eckert (2008) provides the example of teenage girls in a North American high school who made sartorial choices combining different items of clothing from other social groups to create a distinct group style. The context involved two social groups in opposition- the preppy group who wore pastel colours and blue straight leg jeans, and the 'New Wave' group who wore all black and pegged trousers. Two teenaged girls who were in an intermediate position between both social groups explained a stylistic move they had made where they had pegged their blue jeans, thereby mixing both styles: "they segmented the new

wave style into meaningful elements, most saliently cut and color, identified the cut of one's jeans as indexing autonomy but not rebellion or sluttiness, and worked it into their otherwise preppy style, ultimately making a claim to being both preppy and independent" (Eckert 2008: 457).

In the same way, Eckert suggests individuals can combine linguistic variants associated with different groups, and with different social meanings, to build a style which reflects their preferences. For instance, Eckert (2012) describes the way in which the aspiration of intervocalic /t/ has come to be used as a stylistic resource by 'geek girls' (Bucholtz 1996), Orthodox Jews (Benor 2001) and gay men (Podesva 2004), with all three groups drawing on the perception of the variant as signifying "carefulness, precision, and general standardness, hence politeness, attention to detail, or education" (Eckert 2012: 97). Nerd girls (Bucholtz 1996), a white female subgroup, used /t/ releasing to index their values of intelligence and individuality, as well as their oppositional stance against ideas of 'cool' represented by jock and burnout type subcultures. Benor (2001) showed that Jewish boys in a Talmud study group were particularly likely to use /t/ releasing in order to make a point in an argument, while Podesva's (2004) work on the speech of gay men analysed the speech of a doctor who used /t/ releasing in contrasting ways - at work to index competence and education, and at a barbecue with friends to index a gay 'bitchy diva' persona. The indexical field for /t/ releasing thus encompasses multiple indexical meanings, such as articulate, elegant, angry, clear, exasperated and polite (Eckert 2008: 470).

Eckert's concept of an *indexical field* develops further the idea that variables do not have exclusive, fixed meanings "but rather constitute a field of potential meanings – an indexical field, or constellation of ideologically related meanings, any one of which can be activated in the situated use of the variable" (Eckert 2008: 454). Because the indexical meanings of variables are not fixed, Eckert suggests that speakers' evaluations of them can develop in very

localized and context specific ways. For example Eckert (2008) discusses the work of Campbell-Kibler (2007) on attitudes towards (ing), where she found that speaker attitudes towards the meanings of [iŋ] and [in] were affected by the context of delivery. Speakers who are associated with [in] use, such as Southerners or uneducated speakers, using [iŋ] were judged to be pretentious, while conversely educated or northern speakers using [in] were often evaluated as condescending or insincere. This leads Eckert to argue against the variationist orthodoxy that speech community members will tend to share the same evaluations of variables: “Since the same variable will be used to make ideological moves by different people, in different situations, and to different purposes, its meaning in practice will not be uniform across the population” (Eckert 2008: 465).

2.5.3 The community of practice

Speaker design adherents typically focus on the *community of practice* as a unit of study, in contrast to previous variationist studies’ focus on the speech community. As Eckert says: “What these studies could not provide is the link between broad, abstract patterns and the meanings that speakers are constructing in the concrete situated speech that underlies them”. A community of practice can be “A family, a linguistics class, a garage band, roommates, a sports team, even a small village” (Eckert 2005: 16). This focus reflects the ideal that a sociolinguistic study should examine the local creation of meaning rather than imposing large scale abstract categories: “The individual constructs an identity – a sense of place in the social world – in balancing participation in a variety of communities of practice, and in forms of participation in each of those communities” (Eckert 2005: 17). The seminal community of practice study of style is Eckert’s research on adolescent language in a Detroit school (Eckert 2000). Eckert reports that she arrived at the school initially intending to study social class in the traditional sense, but quickly discovered that this ‘adult’ concept had little immediate reality in the student’s conceptualization of their school cohort. Using ethnographic interview and

participant observation methods, she noted that the students categorized their community in roughly dichotomous terms as *Jocks* and *Burnouts*. One of the most salient aspects of this division was the distinct sartorial styles of both groups, but they were also characterized by differing attitudinal orientations and practices. Eckert describes Jocks as being more invested in the school as an institution, while the Burnouts are characterized by their rejection of the school and their orientation towards the urban locality (Eckert 2005). Although these are clearly categories which mirror class distinctions, they are locally constructed by the students from stylistic resources that are locally available.

Another study which powerfully illustrates the concepts of indexicality and stance taking in a community of practice setting is Kiesling's (1998) analysis of the use of (ing) in a male fraternity at an American university. Kiesling's study used a combination of quantitative and qualitative methods, and showed that speakers used both [in] and [iŋ] variants to project masculine values of power and dominance in different ways. The study was focused on the male fraternity as a community of practice, constituting a mixture of informal social activities with semi-formal meetings. Although statistical analysis showed a significant style shift between more and less formal contexts, there were also important differences between each individual fraternity member. Kiesling claims that "the usual suspects of dialect, race, or class differences" (Kiesling 2009: 180) could not explain these differences, leading him to focus on the contrasting stances of speakers during official fraternity meetings. Using discourse analysis of transcripts, Kiesling argues that the majority of participants took stances which aligned them with institutional power and authority. For example, one speaker, Mack, often took assertive epistemic stances with his statements (i.e. portraying his own discourse as factual, without hedging) and this was reflected in his use of the standard [iŋ] variant.

Some of the other fraternity members, however, projected masculinity in alternative way. For instance, the fraternity member Speed repeatedly took stances that were anti-authority in his

discourse, rejecting the organization's rules and power structure. This, according to Kiesling, was reflected in his frequent use of [in] variants in meetings, as it did for other participants who took similar anti-authority stances. Essentially then, the differences in the fraternity members' use of (ing) depended on whether they personally took on institutional or anti-institutional stances. Kiesling goes further to specify that personal styles, such as the laid back rebellious style of Speed, are the result of stances taken over a lifetime: "Over time, Speed has settled into habitual ways of taking stances in interaction, ways that agree with—or help him to continually recreate—his personality [...] at some point in his life (perhaps in adolescence) Speed tried out this stance, found it worked socially for him, and kept using it until it became a habit" (Kiesling 2009: 182). This reinforces the choice driven aspect of this model, since personal styles here are the result of stances taken consciously for the purpose of self-expression and identity work.

2.5.4 Authenticity and stylization

The search for 'authentic' speaker vernaculars has obviously been one of the central missions of classical variationism, as such approaches tend to paint the vernacular as somehow more real than other, later learned varieties. As Coupland states: "it is very likely that the low level of attention paid to style in variationist sociolinguistics reflects the feeling that style is where sociolinguistic authenticity starts to crumble, which might make it a less worthy topic for investigation" (Coupland 2007: 182). This has arguably led to an ideological aversion in variationist work towards the analysis of anything that could be perceived as 'inauthentic' speech. Schilling-Estes (2005) however argues that style researchers are now turning increasingly towards performative speech, since "the investigation of self-conscious speech, even overtly performative speech, seems essential in a research programme in which stylistic variation is viewed as a resource for creating and projecting one's persona – that is, with performing an identity" (Schilling-Estes 2005: 395).

Much of the research that has looked at performative speech has focused on the notion of stylization. Stylization is a concept associated with Soviet literary theorist Mikhail Bakhtin (1981), which foregrounds the way that speakers may draw from multiple personas and voices to project social meanings. Bakhtin described the ways in which performative speech styles allow speakers to subvert other people's speech varieties, by re-appropriating them in ironic playful ways. This relates to the productive (and often comic) tension that can be produced from juxtaposing different speech varieties together, with the speaker's true orientation or stance towards these varieties sometimes left deliberately unclear. As Coupland says "it is a subversive multi-voiced utterance, one that discredits hegemonic, monologic discourses by appropriating the voices of the powerful, and reworking them for new purposes" (Coupland 2007: 150). Bakhtin describes two kinds of multiple voicing - uni-directional double voicing can involve the use of another group's 'voice' where there is some element of approval of the voice's stance on the part of the speaker, while vari-directional double voicing involves a "clash between the speakers' stance and that of the voice s/he is appropriating" (Coupland 2007:102). Stylization is considered particularly characteristic of highly performative language, such as in radio speech. Coupland (2001) analyses the stylization in a *Radio Wales* programme, where two hosts exchange banter in a 'today in history' sequence. Coupland characterizes the segment as stylized because one of the presenters "produces what would usually be thought to be extravagant and chaotic variation between more 'standard' and more vernacular South Wales English values for a wide range of many phonetic variables" (Coupland 2007: 151). By varying sharply between RP and South Walian variants of phonological forms, the presenters reference familiar (for British listeners) stereotypes of gossiping housewives and gay comic characters, all the while cueing the audience to the "somewhat exaggerated" (2007:151) nature of the programme. Another radio study by Coupland (1985) focused on *Radio Wales* presenter Frank Hennessey, a Cardiff based DJ who drew from both referential and dialectal local forms to

construct a style evocative of the locality. Hennessey's prolific references to local shared cultural references, such as Cardiff Arms Park, *Brain's Dark Ale* and a local folk song performed the same function as phonological features from the local Cardiff vernacular. Coupland reports that Hennessey would often deliberately manufacture opportunities for characteristic vernacular variants like [æ:] into catchphrases and the show's introductory jingle. Although he was essentially speaking his own regional vernacular, the authenticity of Hennessey's speech was questioned by some Cardiff locals, who felt that his rendition of the Cardiff vernacular was "self-consciously exuberant and stylized" (Coupland 2007:124). Coupland notes that Hennessey was relatively unique in Welsh Radio at the time with Cardiff vernacular speech not having been heard in the media, and was able to draw upon highly localized stylistic resources to create a persona based around solidarity with vernacular speaking audiences – a potentially subversive anti-hegemonic act.

2.5.5 Speaker design approaches - Summary

One potential criticism of speaker design approaches is that they typically eschew the large scale macro dimensions of both variationism and audience design, losing in the process some of the predictive power that such approaches can provide (Schilling-Estes 2005). However, Coupland (2001) argues for a greater focus on the micro analyses of the styling of individuals embedded in local contexts, believing that generalizing from individuals' speech is fundamentally reductive, and that the creation of stylistic meaning is a local process. Other researchers have argued that finding a link between the dynamic and creative styling process in speaker centred analyses of intra-speaker variation and the orderly stratification found in larger scale inter-speaker variation should be one of the primary aims of sociolinguistics, as Eckert says:

“The challenge in the study of the social meaning of variation is to find the relation between the local and the global – to find the link between speakers’ linguistic ways of negotiating identity and relations in their day-to-day lives, and their place in the social stratification of linguistic variation that transcends local boundaries” (Eckert 2000: 69, quoted in Schilling-Estes 2005).

This suggests that speaker centred approaches can be most fruitful when they are combined with the methods and insights of other paradigms. Indeed, even though it has become increasingly a focus of modern sociolinguistic research into style, Eckert has said that she does not intend for the community of practice approach to replace the speech community, and that each one is “both necessary and complementary” as “the best analytic process would involve feedback between the two approaches” (Eckert 2006: 3). Although what I’m describing as speaker centred approaches here actually covers a broad range of approaches, this summary covers the general tendency of the paradigm:

- Moves the focus of style away from responding to interlocutors (as in CAT and audience design) towards the initiative process of stylization (Coupland 2007), bricolage, and stance taking
- With the notions of indexicality (Ochs 1992) and the indexical field (Eckert 2005), it sees sociolinguistic variables as flexible and as subject to continuous re-evaluation
- Sees the process of meaning creation as locally embedded in speakers’ everyday reality (Eckert 2000), rather than imposed from above (i.e. by academic researchers, or other elite social groups)
- Often focuses on the community of practice as the ideal study setting (Eckert 2000)

2.6 Register variation theory

The final model of style to be discussed here is probably the most controversial, Finegan and Biber's *register design theory* (Finegan and Biber 1994). Register is a term that the authors use to define styles that are differentiated situationally. Examples of registers according to this definition include telephone conversations, public speeches and journal articles. Biber and his associates use a unique methodology, which involves a multivariate analysis of a large corpus of different kinds of registers. Unlike most other sociolinguists, Finegan and Biber incorporate written texts into their analysis, and registers in past analyses have included formal interviews, phone conversations, personal letters and scientific journal articles. Variables are selected largely without making theoretical assumptions beforehand (making it a pre-theoretical approach) and are normally very numerous. Factor loading leads to the identification of distinct dimensions which reveal correlations between variables in certain situations, or registers. Registers are differentiated from the outset for solely functional reasons. For example, pronouns usually index colloquial registers both in speech and text registers, while their absence usually indexes abstract and de-contextualized registers, such as journal articles or news bulletins. As they argue: "they do so not by convention, but by the fact that the functional role of pronouns in highly contextualized situations characteristically produces more frequent pronouns" (Finegan and Biber, 2001: 240).

From their perspective, the distribution of sociolinguistic variables in the speech of social groups reflects the varying degrees of access which those groups have to the range of a language's registers. Linguistic features are divided between two poles of 'elaboration' and 'economy', which are based on Slobin's (1977) opposing mandates 'be clear' and 'be quick and easy'. Finegan and Biber claim that many features can be considered sociolinguistic variables, with elaboration and economy variants showing predictable sociolinguistic stratification (Finegan and Biber 1994). Whereas Bell's (1984) style axiom states that style is

the result of speakers' reflecting their perception of socially differentiated speech, Finegan and Biber's work has the reverse viewpoint, with style emerging from different kinds of speakers' (in terms of class, age and so on) degree of access to functionally differing variants (i.e. elaboration and economy variants). Groups from disadvantaged backgrounds are considered to have restricted access to registers which demand a high degree of clarity, including most written registers and formal verbal styles, while privileged groups tend to have greater access to such registers.

Finegan and Biber refer to the distribution of elaboration and economy variants in several past variationist studies which they argue confirm their thesis. Examples span the use of consonant cluster reduction (Wolfram, 1969) /h-/ omission (Trudgill 1979) and zero copula (Wolfram, 1969) in working class English speakers to the more frequent use of relative clauses (Kroch and Hindle 1982) and lexical variety (Macaulay 1991) by higher status speakers. In order to provide further empirical backing to their theory, Finegan and Biber (2001) conducted a study using the British National Corpus, which includes self-made recordings of participants speaking informally with friends. Speakers were divided into three categories (A/B, C, D/E) based on social class, with A being the highest and E the lowest. The analysis found significant differences for three variables. The upper class groups used the elaboration feature 'relative clauses' significantly more frequently, while the lower class speakers used the economy features 'third person pronouns' and 'third singular DO' significantly more frequently.

2.6.1 Criticism of Register Design

This approach has been strongly criticized by other style theorists (Schilling-Estes 2005, Coupland 2001). Coupland (2001) for instance, criticizes the functional and unidimensional focus of register design theory, and its lack of real explanatory power of for style: "Even if their claims about class, style, and elaboration were correct empirical generalizations (and there

are clear counterexamples), they would have no resources to interpret stylistic variation to which elaboration/simplicity is irrelevant. Their model does not allow for the unique indexical properties of dialect style variables as I have sought to describe them” (Coupland 2001: 195). Finegan and Biber’s work has also been criticized from a methodological standpoint. For example, in her review of the style literature, Schilling-Estes (2005) relegates register variation theory to a footnote at the end of the chapter, noting what she sees as two main methodological problems. The first is that Finegan and Biber present absolute counts of forms rather than proportional usage of different variants, going against standard practice of most variationist work (e.g. Tagliamonte 2012). The second is that the researchers use variables which carry functional meaning, instead of typical sociolinguistic variables which are functionally neutral (e.g. phonetic variables like vowels) “whose meaning derives solely from their association with social groups” (Schilling-Estes 2005: 396). Taken together, these differences make it “not surprising that Finegan and Biber’s findings differ from those of the classic variationist investigations [...]” (Schilling-Estes 2005: 396). The register design model can be summarized in the following way:

- Foregrounds speaker exposure to more ‘elaborate’ or ‘economical’ forms to explain the distribution of speaker styles (Finegan and Biber 1994)
- Style thus works deterministically, primarily reflecting a speaker’s experiences (for example, in terms of class), with little speaker choice or initiative involved
- Less stringent than other approaches about what is an appropriate variable for quantitative sociolinguistic analysis (Schilling-Estes 2005)
- Uses whole counts, rather than proportions, of variants, which is unusual in other approaches (Schilling-Estes 2005)

2.7 Style in minority language contexts

Before I move on to the summary of style, this section will briefly discuss how the theoretical approaches described above can be applied to minority language situations. Labov's claim that "as far as we can see there are no single style speakers" (Labov 1972: 208) has been problematized by studies showing that opportunities for stylistic expression may not be equal in all communities. Bell's audience design theory (1984) in particular makes specific predictions concerning development of style in minority language communities. If stylistic variation is derived from social variation within a community, this raises the possibility that the range of styles can decrease or increase in relation to the extent of social variation in a community: "The style axiom implies that there must be variation between speakers in a community for a variable to be subject to style shift in the speech of one speaker. If a variable has no interspeaker variation, it will have no intraspeaker variation" (Bell 1984: 158). This suggests that in minority language communities where the range of interlocutors are decreasing, we can also predict a reduction in the range of styles which are available to speakers.

Dressler and Wodak-Leodolter (1977), in their work on Breton - which they considered a dying language - described the participants of their study as monostylistic, with speakers' use of the language limited to a restricted set of interlocutors, topics and casual style. Dorian (1981) examined the same issues in the Northern Scottish community of East Sutherland, describing in detail the domain attrition resulting from the restriction of the local variety, East Sutherland Gaelic (ESG), to the home, and its gradual retreat from the public sphere. According to Dorian, the main domain in which standard Gaelic could be heard in this period was the church, where sermons could be heard in the language until the second half of the twentieth century. By the time of Dorian's research in the sixties, however, the use of East Sutherland Gaelic had largely retreated to the private sphere of the hearth and home: "There simply were no appropriate

public spheres in which their native form of Gaelic could appropriately be used” (Dorian 1994: 219).

However, unlike Dressler and Wodak-Leodolter (1977), Dorian did not believe that this lack of contexts of use had made speakers monostylistic: “my impression [...] was that there remained a useful range of styles available to speakers, though a limited range certainly by comparison with that available to speakers of languages which are used to more public and formal uses than ESG” (Dorian 1994: 220). Indeed, in a study looking at the speech of one female East Sutherland Gaelic speaker who was described as having almost no exposure to the standard language, Dorian (1994) found that she was able to style shift between more or less serious topics in a narrative by increasing or decreasing her use of code-switching, essentially using one of the very signs of language death in East Sutherland Gaelic – large scale one way borrowing from the majority language, English, into the vernacular – to supplement her lack of repertoire within ESG.

2.8 Summary

This chapter has covered five models of style, which I have argued can be broadly categorized according to the level of constraints that they place on the stylistic options available to speakers. *Register design theory* sees style differences between speakers as almost entirely determined by speakers’ exposure to different speech varieties, with no real concessions towards intra-speaker variation. The classic variationist *attention to speech* model meanwhile, portrays style as linear, reactive and normatively driven. As in register design, speakers are heavily constrained by social processes based around exposure, but speakers are able to style shift in response to normative ideological pressure (Preston 2001). *Audience design* (Bell 1984) and *Communication Accommodation Theory* represent speakers as being able to reflect the styles of multiple other speakers, although in *audience design theory* this is constrained by the level

of speaker variation that exists within a community, while CAT posits a variety of neurological and psychological constraints on accommodation (Giles et al 1991).

Speaker centred approaches place the fewest constraints on style, with the emphasis on speakers' creative potential. For example, in response to Bell's (1984) idea that styles are constrained by the extent of social variation in a community, Coupland reasons that "this assumption is increasingly out of touch with the circumstances of contemporary life". Instead, Coupland argues for "virtual repertoires – stylistic creations of the imagination" (Coupland 2007: 84), which rely on multiple discourse strategies allowing speakers to bypass 'physical' constraints. Factors which can empower speakers to circumvent constraints include the broad and potentially unlimited field of meaning potential of linguistic resources (which Eckert (2008) described for variables like /t/ release for example) and the Bakhtinian phenomenon of double voicing described by Coupland (2007) and Rampton (1995), which allows speakers to adapt and present features from other social groups' repertoires in novel, ambiguous and creative ways.

On balance, each model offers a variety of competing factors that need to be taken into account. An extreme position on either end of the debate is not really tenable – even Coupland (2007), who takes a robust speaker design approach, states that there must be some 'physical' constraints – i.e. if speakers are never exposed to linguistic features, they will not be able to acquire them. The opposite extreme – where speakers are more or less constrained by social category (e.g. Finegan and Biber 1994) – defies copious empirical evidence showing that identity and self-expression do play a role in style (e.g. Eckert 2000, Kiesling 1998). An analysis of style then, should balance an understanding of broader societal processes - including large scale language shifts (Labov 1966) and community wide ideological frameworks - with some insight into how individual speakers actually navigate these factors in their discourse. It should also proceed with a deep understanding of how the linguistic features under scrutiny

work, particularly in terms of language internal constraints (Tagliamonte 2012), as well as the particular social meanings attached to them by speakers (Eckert 2005). The next two chapters, on mutation and code-switching respectively, will discuss those linguistic features in the light of these concepts.

3. Mutation

3.1 Introduction

This chapter will outline the current literature on mutation, a process of phonological alternation which affects the initial consonants of words which occur in particular morphosyntactic environments. Researchers have studied mutation for a variety of reasons, with some focusing on syntactic and grammatical issues, others on sociolinguistic aspects of its use in the community, and some more recently focusing on psycholinguistic factors such as acquisition and the role of input. Here, after introducing the basic outline of the mutation system in Welsh, I will focus on sociolinguistic studies (Ball 1984; Ball, Griffiths and Jones 1988; Hatton 1988 and Jones 1988) but I will also discuss some psycholinguistic (Thomas and Gathercole 2007) and structural (Ball and Müller 1992) aspects which may provide further insight into the workings of mutation.

3.2 The Welsh mutation system

Consonant mutation in Welsh is divided between three categories of triggers, soft mutation (SM), nasal mutation (NM) and aspirate mutation (AM). Another category of mutation is pre-vocalic aspiration (PVA), although researchers seem to disagree whether it is a separate type of mutation or a subtype of aspirate mutation (Thomas 2001). Each mutation type triggers a different set of systematic initial consonant changes on the initial consonants of the lexical items they affect, and each category also affects a different set of consonants. As can be seen in tables 1-3, soft mutation affects the widest range of consonants, while aspirate mutation affects the fewest:

Radical form	Mutated form
/p/	/b/
/t/	/d/
/k/	/g/
/b/	/v/
/d/	/ð/
/g/	/ϕ/
/l/	/l/
/r/	/r/

Table 3 Consonant changes in soft mutation

Radical form	Mutated form
/p/	/m̥/
/t/	/n̥/
/k/	/ŋ̥/
/b/	/m/
/d/	/n/
/g/	/ŋ/

Table 4 Consonant changes in nasal mutation

Radical form	Mutated form
/p/	/f/
/t/	/θ/
/k/	/x/

Table 5 Consonant changes in aspirate mutation

3.2.1 Mutation triggers

Mutations are triggered in specific morphosyntactic environments, and can have both lexical and syntactic triggers. Lexical triggers are lexical items which trigger mutation in the initial consonant of the lexical item directly adjacent, which Ball and Müller (1992) refer to as the target word. Lexical triggers can be divided into three further subcategories. *Pure lexical triggers* are triggers which always trigger mutation on the adjacent word. *Categorical lexical triggers* involve a broad class of lexical items which consistently trigger mutation. For example, feminine nouns always trigger mutation on the adjacent target word. *Restricted lexical triggers* are single lexical triggers which only affect a certain class of lexical items. An example of this is the definite article *y*, which only triggers mutation on words which have feminine gender (Thomas 2001). Syntactic triggers involve mutation environments where there is no obvious lexical trigger causing the mutation. One frequently discussed example is the soft mutation of direct objects of inflected verbs (Tallerman 1990, Stammers 2010). Lexical triggers are far more numerous than syntactic triggers (Ball and Müller 1992).

3.2.2 Triggering environments

Most studies on mutation tend to illustrate the variety of triggers with a restricted set, partly for reasons of space, but also because some mutations are now archaic and rarely used, particularly

in non-standard, colloquial Welsh (Coupland and Ball 1989). The tables below shows a selection of mutation triggers which are taken from Ball and Müller (1992):

Trigger	Lexical trigger	Example	Type
Prepositions	<i>am</i> (for), <i>ar</i> (on), <i>at</i> (towards), <i>gan</i> (with, by), <i>dros</i> (over), <i>trwy</i> (through), <i>wrth</i> (by), <i>dan</i> (under), <i>heb</i> (without), <i>hyd</i> (till), <i>o</i> (from), <i>i</i> (to)	ceiniog -> am ^{°13} geiniog (<i>for a penny</i>) gwyliau -> ar [°] wyliau (<i>on holidays</i>) coeden -> wrth [°] goeden (<i>by a tree</i>) Bangor -> o [°] Fangor (<i>from Bangor</i>) tŷ -> i [°] dŷ (<i>to a house</i>)	Pure lexical trigger
3 rd person masculine possessive pronoun	ei	ci -> ei [°] gi (<i>his dog</i>)	Pure lexical trigger
Adjectives before nouns or adjectives	n/a	ci -> hen [°] gi (<i>an old dog</i>)	Categorical lexical trigger

¹³ For mutation examples in this thesis I will be following King (2008) in using the symbol ° to indicate that a word would trigger mutation on the adjacent word in standard Welsh. I then highlight the adjacent ‘target word’ in bold to indicate that its initial consonant should (prescriptively speaking) undergo mutation in standard Welsh. This seems a convenient and unobtrusive way to draw readers’ attention towards the relevant mutation environment in such examples.

Feminine nouns before nouns or adjectives	n/a	da -> noson° dda (<i>a good night</i>)	Categorical lexical trigger
Direct object of an inflected verb	n/a	ci -> gwelais gi (<i>I saw a dog</i>)	pure syntactic trigger

Table 6 Selected mutation triggers for the SM system

Trigger	Lexical trigger	Example	Type
Locative preposition	yn (<i>in</i>)	Dulyn (Dublin) -> yn° Nulyn (<i>in Dublin</i>)	Pure lexical trigger
1 st person possessive pronoun	fy (<i>my</i>)	cath (<i>cat</i>) -> fy° nghath (<i>my cat</i>)	Pure lexical trigger
Numerals	chwe (<i>six</i>) pum (<i>five</i>) saith (<i>seven</i>) wyth (<i>eight</i>) naw (<i>nine</i>) deg (<i>ten</i>) pymtheg (<i>fifteen</i>)	blynedd (<i>years</i>) -> chwe° mlynedd (<i>six years</i>) blwydd (<i>year</i>) -> pymtheg° mlwydd (<i>fifteen years</i>)	Restricted lexical triggers. Only affect: Blwydd (<i>years</i>) Blynedd (<i>years</i>) Diwrnod (<i>days</i>)

deunaw (<i>eighteen</i>) ugain (<i>twenty</i>) can (a <i>hundred</i>)	diwrnod (<i>day</i>) -> pum° niwrnod	
---	---	--

Table 7 Selected mutation triggers for the NM system

Trigger	Lexical trigger	Example	Type
Conjunction	a (<i>and</i>)	cath (<i>cat</i>) --> ci a° chath (<i>a cat and a dog</i>)	Pure lexical trigger
Preposition	â (<i>with/as</i>)	tywel (<i>towel</i>) --> dod â° thywel (<i>bringing a towel</i>)	Pure lexical trigger
3 rd person feminine possessive pronoun	ei (<i>her</i>)	pwrs (<i>purse</i>) --> ei° phwrs (<i>her purse</i>)	Pure lexical trigger

Table 8 Selected mutation triggers for the AM system

3.2.3 Internal factors affecting mutation

It has been widely noted that not all target words are equally likely to undergo mutation. Tallerman (1987) divides exceptions to mutation rules into two categories, *predictable exceptions* and *idiosyncratic exceptions*. *Predictable exceptions* are those which have systematic rules which exempt them from mutation. An example of this is that the consonants [f] and [h] render words in which they are the initial consonant systematically exempt from mutation in certain contexts – for example following *yn* as a progressive aspectual marker (Thorne 1993). *Idiosyncratic exceptions* cover a wider variety of features. These include a set

of borrowings from English that are g-initial, such as *gêm* (game) and *graff* (graph), place names, personal names, and some individual words such as *braf* (Tallerman 1987). These exceptions can be varied in the extent to which they apply. Personal names are categorically never mutated in modern Welsh – neither in vernacular nor standard varieties (Ball and Müller 1992). Place names, by contrast, have been described as being mutated variably. Both Ball (1984) and P.W. Thomas (1984) have noted that common nouns were more likely to undergo soft mutation than place names in their data, and Thomas (2001) cites earlier dialect survey studies by Williams (1959) and Evans (1930) that found place names to be resistant to nasal mutation. Ball (1984) and P.W. Thomas (1984) have also both found a pattern of verbs (as opposed to nouns) being resistant to mutation after the 3rd person masculine singular possessive pronoun *ei*.

Words that speakers consider English are not mutated in the standard language. In the case of place names, for example, the convention is to only mutate place names that have a Welsh name. So, for example, *yng Nghaerdydd* (in Cardiff) is standard but *yng Nghanada* (in Canada) is not, with the radical form *yn Canada* being used instead. There are some foreign place names (particularly within the British Isles) which have Welsh equivalents, and which are mutated, including *Caerwysg* (Essex), *Dulyn* (Dublin) and *Caeredin* (Edinburgh). In vernacular speech, however, the application of these rules is intermittent and unpredictable – there are certainly some apparently English words, including both common nouns and place names, in the *Siarad* corpus which are mutated by speakers (see section 5.7.1 below). The matter becomes particularly complex when the relative integration of English words into Welsh is considered, including the distinction between code-switches and established loans. For example, Stammers (2010) has found a correlation between the integration of a word into Welsh (indexed by its frequency) and its likelihood of being mutated standardly. I will address this topic in greater detail in the next chapter.

All researchers into mutation have reported a broad difference in rates of standard forms following the three main trigger categories. Soft mutation is typically deemed to be most ‘stable’, in the sense that multiple dialect survey studies have found that vernacular usage corresponds with standard use in diverse regions, from Davies’ (1934) study of the Ceinewydd dialect, to A. E. Jones’ (1988) study of Pwllheli Welsh, and more (see Davies (1955), Davies (1968), Samuel (1970), Bevan (1970) and Thorne (1971) for more). Ball (1984) found that the SM system was generally close to standard use, although he found that this notably varied according to some of the internal factors mentioned above, as well as the frequency of the trigger itself. For example, infrequent triggers like numerals were less likely to cause SM than frequent triggers like complementiser-*yn*, a finding which is echoed in P.W. Thomas (1984).

The only studies to have found an exception to this pattern are those which have looked at communities in which Welsh speakers are in a minority. One example is Jones’ (1998) survey of the Welsh of two eastern anglicised communities. Jones found a gradual decline in the use of SM following prepositions in all age groups, and a more abrupt decline in the use of 3rd person masculine possessive pronoun *ei* in the very youngest school age group. Non-standard use of AM and NM is far more widely attested. The widespread lack of standard mutation following AM triggers in the vernacular was described in Gwynedd as early as 1913 by Fynes-Clinton (1913). Studies have tended to report standard mutation following NM triggers as being more frequent than those following AM, with the pronoun *fy* more likely to be followed by standard mutation than the locative *yn* by most speakers (Hatton 1988, Jones 1998).

3.2.4 Functional load

The third person female pronoun *ei* has been found to be mutated standardly relatively frequently, even though it is an AM trigger, which is typically more resistant to standard mutation (Ball 1984, Jones 1998). Researchers have argued that this is because of the *functional*

load (Ball 1984) of this trigger, as the mutation of the target word can potentially be the only means of disambiguating the gender of the possessor. For example, in the constructions *ei gi* (his dog) and *ei chi* (her dog) it is clear that the possessor in the first example is male, and that the possessor in the second example is female, because of the contrasting application of SM and AM. If a speaker chooses to use the radical form of the target word, as in *ei ci* (?’s dog), the gender of the subject is harder to identify without contextual information. This may be the reason that the use aspirate mutation following the feminine possessive pronoun seems to have been maintained relatively strongly (particularly in comparison with other AM triggers), with Jones (1998) for example reporting 100% use of the standard form in speakers over 25 in a North Eastern dialect. In the same study, Jones found that mutation triggers that didn’t carry a comparable functional load, such as prepositions, were less likely to undergo standard mutation in younger age groups.

3.2.5 Variable zero triggers

Ball and Müller (1992) have described some mutation contexts which lack triggers, and thus may appear at first glance to be of the syntactic type, as actually having *zero lexical triggers* which are dropped, but understood by speakers, with their presence or absence dependant on stylistic considerations: “it is more insightful and economical to treat such examples as ‘zero lexical triggers’ in that all such examples in Welsh have a stylistic alternative in formal registers with an overt lexical trigger which we claim can be deleted by an optional (or variable) rule” (Ball and Müller 1992: 32). Among these *zero trigger variables* are negative sentences which have a dropped pre-verbal negative particle:

(ni)° **cherddodd** hi ddim (“*she didn’t walk*”)

Vocative mutation where the vocative particle *o* is dropped:

(o)° **dad** (“*oh father*”)

And positive statements where the affirmative particle *fe/mi* is dropped:

(*fe*)° **welodd** Rhodri adar (“Rhodri saw birds”)

Ball and Müller (1992) argue that these triggers’ presence or absence is largely dependent on the formality of the context, so that they will be retained in formal style but dropped in less formal situations. A related trend that has been observed in the speech of younger speaker is a tendency to drop the mutation triggering pronoun in certain genitive constructions (B.M. Jones 1990, Jones 1998), with some important implications for mutation. In standard Welsh, the established pattern is for the possessive pronoun to precede the target word, with the pronoun triggering nasal mutation, for example:

fy° **nhŷ**

1.S.POSS house

“*my house*”

However, some studies have reported the use of a construction in which the pronoun *fy/yn* is dropped and an object pronoun is used following the noun, as in: (noun) *fi*, with the noun subsequently often going unmutated (Jones and Thomas 1977; Watkins 1977, Hatton 1988), for example:

tŷ *fi*

house 1S.PRON

“*my house*”

Jones (1998), in a more recent study, found that younger speakers frequently used similar constructions, both to indicate first person possession and for third person possession instead of the standard pattern of “*ei (noun)*”, as in:

ei° dŷ

POSS.3SM house

“his house”

Or the slightly more informal “ei (noun) hi/e”:

ei° dŷ e

POSS.3SM SM.house PRON.3SM

“his house”

Jones found that many speakers dropped the initial pronoun for the construction “(noun) hi/e”. As with the (noun) + fi construction, in many cases this led to speakers not subsequently mutating the target word, as if the dropped zero trigger were really absent, for example:

tŷ e

house PRON.3SM

“his house”

Jones speculates that this process could be an example of what Trudgill calls “simplification with cost” (Trudgill 1983: 119), where the loss of an important functional mutation (which indicates the gender of the object) is made up for by the presence of the object pronoun. In other words, it involves the mutation system: “becoming simpler due to the lack of mutation but with the addition of the object pronoun e (‘him’) after the object possessed in order to assert the gender of the possessor” (Jones 1998: 62). Jones stresses that this pattern was largely confined to the school aged group, who had also shown simplification across a range of Welsh features, but goes on to state: “it will be interesting to see whether, in time, the usage of the younger group grows closer to the conservative norm or whether such historically inappropriate

forms are retained” (Jones 1998: 64). Oddly, the use of the (noun + fi) construction has been particularly harshly criticized by some prescriptive linguists as “baby talk and substandard” (P.W. Thomas 1984: 231) although more recent scholarship reports that it is common in vernacular speech (Borsley et al 2007).

3.2.6 Mutation in Gwynedd

Studies on mutation set in the North West of the country, one of the areas of Wales where predominantly Welsh speaking communities are at this point arguably most likely to be found, are “surprisingly sparse” according to Ball (Ball 1988: 23), and with the exception of laboratory based projects (e.g. Gathercole and Thomas 2007) this remains true. A very early study set in Gwynedd was Fynes-Clinton’s (1913) study of the Welsh of the Bangor area, who stated even at that early stage that aspirate mutation following the triggers *â* and *chwe* were routinely avoided by local speakers. A more recent study set in Gwynedd, at the far end of the Llŷn Peninsula, is A.E. Jones’ (1988) survey of the Pwllheli dialect. Jones states that in general the NM and AM mutations are heard infrequently in the area, although in line with the majority of Welsh dialect surveys he does not provide statistical information. The pattern of NM mutation is variable, with the trigger (preposition *yn* + place name) typically not mutated (i.e. *yn Bangor* rather than the standard *ym Mangor*) while the trigger (possessive pronoun *fy/yn*) usually does trigger the standard nasal mutation (A.E. Jones 1988). The AM mutations are rarer than the NM, surviving generally only in fossilised formulaic or idiomatic constructions such as *ceffyl a chert* (“horse and cart”). The SM mutations are again reported to be reasonably stable, with the only exception being variability in mutations which are triggered by gender. According to Jones, younger and older speakers differ in their ability to recognize gender in certain nouns, for example *cwch* (boat), which young speakers consider feminine, leading younger speakers to say *y gwch* while older speakers will use the standard *y cwch*.

3.3 Variationist studies of mutation

This section looks at a series of studies which have attempted a sociolinguistic analysis of the distribution of standard and non-standard mutation forms in different communities throughout Wales. Many of these studies share common themes, including a focus on diglossia (P.W Thomas 1984) and acculturation towards either Welsh or English (Ball 1984; 1989) and the impact of this on the amount of input of standard forms that speakers are able to receive. As I discussed in the introduction, no studies on mutation have engaged with class as a sociolinguistic variable, and studies have generally prioritized the social (inter-speaker) over stylistic (intra-speaker) aspects of mutation. Of the studies discussed below, three have attempted to measure style shifting - with instructively contrasting approaches.

Ball (1984) used a variety of methods to elicit different styles, including word lists with pictures and numerals instead of texts, a practical imitation task called the *reporter's test* (Ball 1985), and a more casual interview context to provide a continuum of attention to speech, while Hatton (1988) used translation and completion tasks to elicit different styles. Ball et al (1988) on the other hand, used a corpus based approach which ranked radio programmes in order of scriptedness and spontaneity to map a linear stylistic axis. Sociolinguistic work on mutation so far then, has generally fallen within the Labovian tradition of prioritising inter- over intra-speaker variation. The studies that I discuss here are predated by a series of earlier dialectal surveys (see Ball and Müller 1992 for a full description), whose main aims were to establish differences in use between various geographical areas. Since the rationale behind these studies was not rooted in more modern sociolinguistic concerns about variation between speakers in the same community (Ball 1988), and as they rarely report empirical data, I will not describe them in any detail.

3.3.1 P.W. Thomas' mutation study

The first study to attempt a variationist analysis of social variation in mutation is P.W. Thomas' (1984) study of South Glamorgan Welsh. The data consisted of sixteen tape recorded interviews with members of the local rural community, and were collected by two different interviewers. The interviews were part of an archive kept at St Fagan's, and they spanned a period from 1953 to 1980. Thomas describes the participants as 'non-mobile' and 'of a non-industrial background' (P.W. Thomas 1984: 208). Since the study essentially consisted of single style sociolinguistic interviews, Thomas conceded that it could not provide data on the style shifting behaviour of the participants. He argued, however, that the nature of rural communities precludes the use of Labovian style methodologies:

“Most of the available quantitative descriptions of sociolinguistic phenomena deal with industrialized English-speaking societies in which may be found a hierarchy of social classes, the highest of which may include individuals who maintain certain prestige variants” (P.W. Thomas: 210)”.

He also argued that there were aspects of the study setting which were distinct from those traditionally studied in the Labovian paradigm. Without specifically mentioning diglossia, he outlines a situation where Welsh occupies the L position used for “home, work, and recreation” and English occupies the H position as “the language of officialdom” (P.W. Thomas 1984: 210). The only situation where he argued a register of Welsh occupied the H position was the chapel, where most of the congregation would experience it passively and sporadically. In terms of style then, it seems that Thomas' argument is that in this type of community (with bilingual diglossia, and a relatively non-hierarchical social make-up) most Welsh speakers will be monostylistic, with the majority unable to style shift in more formal situations. According to the author, only a minority active in the chapel or in Welsh literary circles would thus be

habituated towards the use of standard Welsh, as a result of taking part in religious readings, preaching (lay or otherwise) and/or engagement with Welsh medium literature.

Thomas uses this notion of relative degrees of exposure to categorize the speakers in his corpus according to their likelihood of having encountered formal registers of Welsh. Group (a) consists of elite members of the community, who performed leading roles in central aspects of community life, including leading roles in prestigious chapel based activities. This group was considered to have the highest level of input from standard Welsh. Like group (a), group (b) also lived their day to day lives through the medium of Welsh, although they did not have as much access to formal Welsh due to their lower status within the community. Group (c) consists of members of anglicised communities, who had fewer opportunities to receive input through the medium of Welsh, both in terms of formal domains and the vernacular. They also would mainly interact in Welsh only with older members of their communities, which Thomas argued meant that they may be likely to inherit some more formal or archaic elements in their speech. Group (d) were similar to group (c) but were younger, and were thus presumed to have even less input from Welsh speakers.

As linguistic variables, P.W. Thomas focused on eight triggers from the three main category of mutation triggers; SM, NM and AM. The triggers he selected were as follows:

Variable	Trigger meaning	Mutation Type Triggered
(ei.f) ¹⁴	<i>her</i> (female third person possessive pronoun)	AM
(a)	<i>and</i> (conjunction)	AM
(negative verb)	-	AM
(fy)	<i>my</i> (first person possessive pronoun)	NM
(yn)	<i>in</i> (locative preposition)	NM
(ei.m)	<i>his</i> (male third person possessive pronoun)	SM
(prepositions)	-	SM
(negative verb)	-	SM

Table 9 Variables analysed in P.W Thomas' (1984) study of mutation

The results of the analysis found that the SM was essentially stable, while the standard use of AM and NM were declining overall. One exception to this trend was that the possessive triggers were maintained relatively strongly across all categories of mutation, including AM (after S3F pronoun 'ei') and NM (after first person possessive pronoun 'fy'). Thomas proposed that these triggers were functioning in a similar way to Labovian sociolinguistic variables, indicating social differences between members of a community.

The stratification of standard mutation followed the division of each group in order of prestige and acculturation, as group (a) used the most frequent realizations of all mutations and group

¹⁴ For consistency's sake I am representing the mutation triggers discussed in P.W. Thomas and others' studies using an adapted version of Labov's notation, placing a symbol representing the trigger between parentheses. P.W. Thomas and most other authors have not represented mutation variables in this way.

(d) used the least. Thomas' basic explanation for these differences focuses on participants' varying degrees of exposure to Welsh, and to formal Welsh in particular, with the extremely low rates of mutation in group (d) driven by "rule loss associated with decreasing use of language" (P.W. Thomas 1984: 234). One discrepancy to this pattern noted by the author was that group (c) used standard mutations more frequently than group (b) on certain highly formal or archaic triggers, which Thomas explained as being due to these speakers using Welsh almost exclusively with older speakers in their community. This implies that level of exposure, and the nature of the social network are some of the central factors in shaping a participant's use of standard versus non-standard mutation.

3.3.2 Ball's Cwmtawe Study

One of the most ambitious attempt to study the social and stylistic distribution of mutation has been a series of studies by Ball (1981; 1984; 1985; 1988). Ball's aim was to apply the insights of Labovian variationist research to the study of mutation, in order to test the hypothesis that Welsh speakers style shift in their realization of standard and non-standard mutations. In order to apply variationist techniques to the Welsh context, Ball faced a challenge in adapting Labovian techniques to create a methodology appropriate for the language and the cultural situation. As I mentioned in the previous chapter, Welsh communities are generally heterogonous (Ball and Coupland 1989), with Welsh speakers of various abilities and cultural acculturation living alongside monolingual English speakers. Like P.W. Thomas (1984), Ball thus breaks with traditional variationist approaches in his lack of a class based variable analysis, reasoning that it is too difficult to recruit participants with "an even class mix" in small minority language communities (Ball and Müller 1992: 248). In order to further justify his lack of focus on class, Ball also quotes Milroy (1980), who has critiqued the notion of class on the basis that it lacks "any kind of objective, or even intersubjective reality" (Milroy 1980:14). Ball's solution to this problem was to draw on Milroy's notion of social networks, where the

researcher analyses networks consisting of families and friends, and to replace class with the notion of ‘cultural acculturation’ towards either Welsh or English, an approach which R.O. Jones (1984) had previously used to analyse phonetic variables among Welsh speakers in Patagonia.

Participants were placed into one of three groups; group A, group B and group C. Group A were those who were most ‘Welsh oriented’, and who had standard Welsh as their model of mutation due to their Welsh medium education. Group B were an intermediate group who had not received Welsh medium education, and for whom the vernacular of the local Cwmtawe dialect served as their only model of mutation. Group C were the most English acculturated participants, being predominantly younger speakers who had “no model of mutation usage”. Ball noted that the acculturation groups generally overlapped with age groups, with groups A and B generally having an older age profile than group C. This reflected the decline in Welsh acculturation amongst the young people in the Swansea Valley at the time of the study. Other non-linguistic variables that Ball included for analysis were the traditional sociolinguistic variables of age and gender.

In terms of strictly linguistic variables used for analysis, Ball restricted his main focus to a group of eight variables including seven AM triggers and one NM trigger, relegating SM to a sub-study.

The triggers for the main study are the following:

Variable	Trigger meaning	Mutation type triggered
(a)	<i>and</i> (conjunction)	AM
(â/gyda)	<i>with/as</i> (preposition)	AM
(tri)	<i>three</i> (number)	AM
(chwe)	<i>six</i> (number)	AM
(na/tua)	<i>nor, than, towards</i>	AM
(yn)	<i>in</i> (locative preposition)	NM
(negative verb)	<i>ni, na</i> or \emptyset + inflected verb	AM
(ei.f)	her (female third person possessive pronoun)	AM

Table 10 Variables analysed in Ball’s (1984) study

Ball relates his decision to focus on the AM and NM triggers partly to a desire not to overextend his study, but he also considered AM and NM to be the mutation triggers “most often reported as being subject to variable usage” (Ball 1988), and thus more likely to reflect social and stylistic differences between participants. As previously mentioned, many writers have already noted a clear shift away from common usage of standard NM and AM mutations, while seeing the SM system as largely stable (Awbery 1986). However, Ball did report on the distribution of ten SM variables in a smaller subsequent study, which I will describe following the AM and NM results below.

3.3.2.1 Elicitation Methodology of Bell’s Study

Ball’s data collection methodology follows Labov’s (1972) approach reasonably closely, using a traditional sociolinguistic interview followed by a series of elicitation tasks intended to collect

distinct styles of speech. The method is adapted slightly however to adapt to the specific demands of mutation as a variable. Ball points out that Welsh has a clear orthography, in contrast to the opaque orthography of English, which means that a “greater approximation to standard forms will be found in Welsh, due to the stricter correspondence between grapheme and phoneme” (Ball 1984: 115). This leads to methodological problems: “The mutation system is reflected in the orthography, therefore if techniques are used which utilise written material (such as the minimal pair word list and reading passage styles) the results are going to be biased depending upon whether the mutation is shown in the written text or not” (Ball 1984: 117). Ball argued that this bias could be particularly strong in cases where a relatively prestigious mutation such as AM is being used, because its presence in a text could be particularly salient to participants.

In developing a way around the orthography problem, Ball tested multiple pilot studies in order to ascertain the most effective approach. These are listed below (taken from Ball 1984: 122):

1. Translating sentences from English

- a) spoken

- b) written

2. Sentences including pictures for the mutable word.

3. Prose passage containing pictures for the mutable word.

4. Description of items in a composite picture.

5. Memory test (Kim’s game) - informant attempts to list a number of items previously committed to memory.

6. Reporter's Test - where the informant must describe verbally as if to a third person actions undertaken by the investigator.

Ball considered the translation task to be “at a disadvantage in that two languages are involved”, and identified further problems with the memory test (5) and descriptions of items in a composite picture (4), which were “subject to too many extra-linguistic effects to be very useful in the future” (Ball 1984: 130). Task 3, which used pictures in an extended prose passage, was described as being overly affected by syntactic factors. The author thus decided to focus upon task 2 - the mutation of picture items in simple sentences - for the most formal style, and task 6 - the reporters’ test - as an intermediate style situated between the casual interview and the sentence reading task.

The final interview schedule was as follows. Section A consists of the sociolinguistic interview, with ten questions on informant’s background to “gain necessary information about the subject but also interview style speech samples” (Ball 1988: 205). The second section, B, involves the reading task with 16 sentences containing pictures, including dummy sentences to mislead participants about the purpose of the study. Section C involves more background questions; specifically local area, education background (including the medium of education) as well as emotive questions intended to provide more casual examples of speech. Section D involves the reporter’s test, where the interviewee is asked to describe actions performed by the interviewer (such as moving coloured shapes and body movements) which are designed to elicit specific mutation triggers. Section E involved specific language questions: e.g. whether the interviewees participate in local ‘linguistic networks’, their opinion of the local dialect, how well they speak Welsh and their opinion on the future of the language. The final sections were F and G, which both involved testing the linguistic awareness of speakers by asking them about specific variables, and subsequently comparing this to their actual use of the variable.

3.3.2.2 Results of Ball's study

Variable scores were calculated in one of two ways, either as a simple percentage of non-standard versus standard occurrence, or as an index score where variables had more than two possible variants. Index scores were used for the NM and AM triggering variables (6-8), to account for the fact that participants sometimes substituted the traditionally expected variants for SM. Ball calculates index scores on a scale of 0 to 200 following Chambers and Trudgill's (1980) method. A general statistical value for each variable was produced by averaging their use across all styles.

Variable	Trigger meaning	Percentage of standard forms	Mutation Triggered
(a)	<i>and</i> (conjunction)	30%	AM
(â gyda)	<i>with/as</i> (preposition)	22%	AM
(tri)	three (number)	15%	AM
(chwe)	six (number)	15%	AM
(na/tua)	nor, than, towards	18%	AM

Table 11 Standard mutation of percentage scored variables in Ball (1984)

For the percentage scores variables, Ball interprets the results as falling into three distinct scoring groups – the most standard group, which contains (a), a medium scoring group which contains: (na/ tua) and (â/gyda), and a least standard group which contains the numerals (tri) and (chwe). Ball argues that (a) was the aspirate variable mutated most standardly due to its relatively high frequency in speech, leading to more opportunities for input. Another explanation that Ball drew upon was the presence of (a) in well-known idioms such as *ceffyl a chert* (“a horse and cart”), where, as A.E. Jones (1988) has argued, the aspirate mutation is

preserved in a ‘fossilized’ form. Ball seems to be arguing here that the use of (a) in these kinds of idioms serves to reinforce the rule of AM mutation following the trigger, cueing speakers to the application of the rule. In terms of the numeral triggers, which scored particularly lowly, Ball argues that this is due to their being far less frequent in speech than (a), but also because numerals as a class do not mutate categorically in Welsh, which may make acquisition of the rules more difficult.

3.3.2.3 Index scores in Bell’s study

The index scored variables were presented on a scale from zero to two hundred.

Variable	Trigger meaning	Index score	Mutation Triggered
(ei.f)	<i>her</i> (female third person possessive pronoun)	168	AM
(negative verb)	<i>ni, na</i> or 0 + inflected verb	76	AM
(yn)	<i>in</i> (locative preposition)	82	NM

Table 12 Standard mutation of index scored variables in Ball (1984)

Ball relates the high standardness of (ei.f) to the fact that it carries an important functional load, allowing speakers to make crucial distinctions between the genders of subjects. The lower scores for (negative verb) and (yn) in comparison thus may be due to the fact that they do not carry a similar functional importance. Ball claimed that speakers were particularly aware of (ei.f), because teachers had placed a particular emphasis on it at school. The locative preposition (yn) was described as being a mutation trigger of which participants were also relatively more aware than others. Participants did not show awareness of any of the other variables, which indicates that each variable has a different level of saliency for speakers.

3.3.2.4 Inter-speaker variation in Ball's study

Ball found strong differences between the acculturation groups, with a general pattern of the most Welsh acculturated group (group A) using more standard forms than the more English acculturated groups. The differences that Ball ascribes to the groups usage of mutation largely mirrors psycholinguistic studies into mutation (Thomas and Gathercole 2007) which have found that level of input is an important factor in a speakers' likelihood of using standard mutation. Ball also found a difference between genders, with male participants using more standard mutation than the female participants for each variable - although this difference was only statistically significant for (chwe). As with the acculturation groups, Ball's explanation for the difference is again based around exposure. Like Dorian's description of East Sutherland Gaelic speakers (1994), Ball argues that the men in the Cwmtawe community tended to be those who played a more active role in domains where formal Welsh was likely to be used, including in chapels and literary societies.

3.3.2.5 Style Shifting in the Cwmtawe study

Ball found a significant difference in the extent to which the different acculturation groups modified their language use from one style to the next. Ball calculated the average of the mean between two styles (f and n-f) for all participants:

Acculturation Group	Style shift
A	15.15
B	14.95
C	0

Table 13 Mean style shifting for three acculturation groups in Ball (1984)

Ball argued that these results illustrate the fact that the groups differed in their ability to style shift, with the reason being once again tied to exposure to standard forms. The most English acculturated group (group C) did not style shift at all, as they had no control over the variables, due to their very low level of exposure to standard Welsh. Although Group A's score was not significantly higher than group B, there was an important difference in the way that they style shifted some variables. Whereas group B maintained a consistency between (chwe) and (tri) in both styles, group A style shifted much more strongly. Once again, Ball emphasizes exposure and input to standard Welsh in his explanation for this phenomenon, arguing that Group A had more awareness of the prestige of the standard forms than the other groups due to their Welsh medium education.

3.3.2.6 Markers, indicators and stereotypes in Ball's study

From his findings, Ball attempted a tentative categorization of each AM and NM variable as markers, indicators and stereotypes. In line with Bell's (1984) observation on the infrequency of indicators, only (ei.f) was placed in this category, with evidence that the variable correlated with the social acculturation of speakers, but not with changes in style. In the marker category, Ball placed (yn), (negative verb), (a) and (â/gyda), with the suggestion that these variables both varied with social and stylistic differences. The remaining variables; (tri), (chwe) and (na/tua) were described as stereotypes – varying exclusively with style - but only for the most Welsh acculturated group A, with the other groups described as lacking these variables in their repertoire. Although Ball did not describe the use of SM variables in terms of markers, indicators and stereotypes, he did describe some differences in social group use as well as internal factors, which I discuss in the next sub-section.

3.3.2.7 The SM variables in Ball's study

As mentioned above, Ball (1984) considered that the SM variables largely reflected standard Welsh norms. Although Ball reported that the use of standard variants was common in most cases, he also described substantial variation between the different SM triggers:

SM Trigger	Standard mutation mean
'complementiser-yn' before verb or adjective	96.97%
(ei.m)	95.65%
all gender triggered SM	94.56%
preposition + noun	93.67%
mutation of direct object of inflected verb	90%
negative verb	85%
pre-modifier triggered SM of nouns (excluding the definite article)	79.41%
preposition + other (i.e. verbs, numerals and adjectives)	76.15%
preposition + place	69.23%
(ei.m) + verb	62.5%

Table 14 Mean percentage standard use of SM triggers in Ball (1984: 284)

Ball identified two main reasons for the lower standard use of some variables. The first was that triggers which appeared less frequently in speech were less likely to be mutated due to lack

of input. This included numeral triggers, verbs following prepositions and verbs following masculine pronoun *ei*, which were all relatively rare in the data. The second was that certain features resisted mutation for language internal reasons. These include place names following prepositions (e.g. *i° Fangor* “to Bangor”), which underwent SM far less frequently than in common nouns following prepositions (e.g. *i° dŷ* “to a house”), and verbs after the masculine third person possessive pronoun (*ei° ddarllen* “to read it”) as opposed to nouns following masculine pronouns (*ei° gath* “his cat”). Ball argues that the constraint on mutating place names may be a similar phenomenon to that on personal names in modern Welsh, and speculates that the constraint on verbs following pronouns may be a result from the kind of syntactic change discussed in Jones (1998), although this is not pursued further. As Ball notes, both these constructions were also noted as being mutated less frequently in P.W. Thomas’ (1984) study.

3.3.2.8 Criticism of Ball’s study

Ball’s study is important in that it provides a model for applying a Labovian variationist methodology in order to analyse social and stylistic variation mutation in the Welsh context. Ball’s (1984) explanation for variation in mutation is restricted to the relative degrees of exposure that different social groups have to standard language, but this is a persuasive argument, particularly in a community where some groups’ access to standard forms are highly restricted. On the other hand, it ignores the possible contribution of other factors such as identity and attitude which could be brought to the fore with the addition of more qualitative aspects to the study. Ball briefly does mention audience design in a discussion of the reporters’ test in (Ball 1985), but claims that the style shifts in his data did not seem to be affected by this principle, as he was consistently the only interviewer present. Conversely it might be argued that audience design theory does not rule out attention as a potential factor – only that it is secondary to other processes (e.g. identity) which should also be explored.

3.3.3 Hatton's school based mutation study

Another study which has examined social and stylistic variation in mutation is Hatton (1988), which examined the use of mutation amongst schoolchildren at a primary school in Swansea between 1978 and 1980. The purpose of the study was to look at the relation between mutation and social factors like age and parental input, but also at the relationship between mutation and style. Unlike Ball's work, which focused on aspirate mutation, Hatton's analysis centred on the NM system, encompassing all of its reported triggers: locative preposition (yn), possessive pronoun (fy) and a series of restricted triggers involving mutation of a specific set of time related words (*blwydd* – year, *blynedd* – years, *diwrnod* – day) following a set of specific numerals (including *pum* – five, *saith* – seven, and *wyth* – eight). Many of the variants in the third category, such as the standard mutation in the expression *pum ° niwrnod* (“five days”), are highly archaic, as Ball and Müller (1992) note. Participants were divided into four age groups, which reflected different class cohorts, with 1 representing the youngest class and 4 the oldest. They were also divided into four groups based on their home language situation and birthplace of their parents: A - both parents are L1 Welsh and local B - at least one parent is L1 Welsh and local C - both parents English monoglots D - at least one parent is L1 Welsh but not local. This factor was primarily included so that Hatton could compare children whose main model was the local vernacular, with others, whose model may be another dialect (due to having parents from another area) or that of the school (due to having monolingual English speaking parents). To measure style, children were recorded in two conditions: a formal questionnaire condition, where they were asked to translate English sentences containing the target triggers to Welsh, and answer prompts in a completion task, and a free conversation condition intended to represent informal style. Hatton predicted that the mutations in the first condition would be more standard, while in the second they were would be less standard.

As can be seen in table 15, there was clear correlation between age group and the standard use of (fy), with the older students using generally more of standard variant. The majority of children across all age groups showed a clear pattern of shifting stylistically between the informal and formal conditions. However, in terms of the ‘linguistic background’ groups, Hatton found that group C, who did not have home exposure to Welsh, showed almost no style shifting between conditions for (fy). This was because: “the formal patterns of the school are the only influence on the speech of the children of this group and as a result they are monostylistic; they do not have the same variety of register as the children of other groups” (Hatton 1988: 251).

Class	Informal style
4	62.16%
3	81.08%
2	11.11%
1	33.33%

Table 15 Percentage standard use of (fy) by age in free conversation. Hatton (1988: 248)

The analysis of (yn) revealed that, unlike in the case of (fy), the standard variant was hardly used across age groups in the informal conversation section. The variable did show an increase in use in the careful style however, particularly in the two oldest age groups. Hatton claims that this means that children “were aware of formalizing” (Hatton 1988: 253) and were able to adapt the use of the variable for a more formal context.

Class	Informal style	Careful style
4	0%	37.03%
3	8.89%	19.84%
2	0%	3.33%
1	4.17%	4.94%

Table 16 Percentage standard use of (yn) by age. Hatton (1988: 253)

The distribution of standard use of (yn) by linguistic background again shows style shifting being affected by the home situation of the speakers. Speakers from group A and B showed the highest range in style shifting, probably due to their high level of exposure to both school and vernacular models. In contrast to this, Hatton notes that group C, who had no vernacular model at home, had the highest standard use of (yn), although they showed no style shifting.

Linguistic background	Informal style	Careful style
A	2.5%	17.28%
B	0%	7.14%
C	16.67%	14.81%
CH	8.33%	12.22%

Table 17 Percentage standard use of (yn) by background. Hatton (1988: 253)

Hatton's explanation for this is that the group C children's sole exposure to standard variants of (yn) was through school. This means that they were thus 'better able' to apply the standard mutation, but were also not aware of the stylistic significance of (yn) through comparison

between vernacular and standard models of use, and were thus monostylistic. This finding is particularly interesting in that it suggests that vernacular community norms of non-mutation can ‘push back’ against the standard variants promoted in H domains. In terms of the numeral triggers, Hatton found that, as in Ball’s (1984) study, their usage was highly restricted in comparison to the other variables. There was further variation depending on the specific pairing of numeral and period of time. For example, the standard realization of *pum mlwyddyn* occurred in around fifty percent of the questionnaire responses, while the more archaic *pum niwrnod* occurred only a handful of times.

3.3.4 Ball, Griffith and Jones’ Welsh Broadcasting study

Ball et al’s (1988) study on mutation is the third attempt to study stylistic variation in mutation. Unlike the other studies discussed here, this project did not attempt to analyse inter-speaker variation, concentrating instead on stylistic variation. The study is based partly on an elaboration of methods used by Bell (1977) and Morin and Kaye (1982). Morin and Kaye (1982) developed a study design that categorized radio programmes by order of formality in order to analyse the stylistic variation of elevated liaison in French. Ball et al adapted and extended this methodology in order to analyse some of the same mutation variables examined in Ball’s (1984) Cwmtawe study. Categorization of the programmes followed the principle that the level of formality of a radio programme could be measured using the level of ‘scriptedness’ versus ‘spontaneity’ as an index. Three programmes were selected, and then categorized into three incrementally varying levels of formality.

The first and most formal programme was a conventional mid-day news bulletin, with headlines being read by a newsreader. This was considered the most formal programme due to the lack of spontaneity and its completely scripted format. The second level of formality was to be a programme where a script was involved, but where material was read ‘as if spontaneous’.

For this category the researchers selected the religious programme *Oedfa'r Bore*. This programme involved the broadcast of a Welsh Presbyterian sermon, within which the authors identified multiple stylistic situations. Completely scripted contexts, including hymns and readings, were omitted from the analysis. The two main remaining contexts were a sermon and a 'free prayer'. The sermon was described as being particularly interesting due to the ostensibly deliberate stylistic variation within it, with one section including a story in which the minister performed the voices of two different characters. The final and most informal category was a programme where no script was used, and speech was considered completely spontaneous. This category was represented by an entertainment programme, *Stondin Ddyddiol*. The programmes were also differentiated by the apparent level of formality belonging to the subject matter, as perceived by the researchers. In this way, the music and entertainment programme was considered less formal than the religious service and the newscast. The final categorization of each programme by level of formality was as follows:

1. News bulletin (most formal)
2. Religious service ' *Oedfa'r Bore* ' (semi-formal)
3. Entertainment programme *Stondin Ddyddiol* (Least Formal)

The researchers state that other (non-mutation related) linguistic features were used to support the categorization of formality for each programme. The only variable which is specifically described in this capacity is the shift from the standard diphthong (OI) to vernacular monophthong (O:) which the minister employs when 'performing' characters' voices during a story, which the researchers argue signals a stylistic shift from more formal to a less formal style. Similar variables are said to signal the informality of *Stondin Ddyddiol*, but are not described in any detail. This appears to represent an attempt to test the formality of each programme in a similar way to Labov's channel cues such as laughter or voice pitch (Labov 1972). The researchers also attempted to control other non-linguistic variables, particularly the

social and regional backgrounds of the participants. All of the presenters were from the South West of the country, and were described as being ‘educated’.

One of the study’s most interesting findings is that NM usage was standard in all cases but one, with one instance of a participant using a soft mutation following preposition (yn), and this in the least formal programme. The authors emphasized the contrast with this in comparison with the findings of Ball (1984) which found 37% to be the average for the same variable across all styles in the Cwmtawe community. The AM mutations, in contrast, vary with the level of formality of each programme, as categorised by the authors.

Programme	Standard	Non-standard	Percentage standard
News	a x 1 ; tri x 2 ; â x 2; ei x 1	-	100%
<i>Oedfa’r Bore</i>	a x 6 ; gyda x 1 ; na x 1	a x 2	80%
<i>Stondin Ddyddiol</i>	a x 3; chwe x 1; na (neg) x 1	a x 5 ; â x 3 ; gyda x 1 ; na x 2	31%

Table 18 Standard use of AM mutations by programme (Ball et al 1988: 188)

The authors argue that the difference between the AM variables, which showed stylistic variation, and (yn) which exhibited nearly categorical standard use by contributors, may have been due to the higher salience of (yn). This relates to Ball’s (1985) examination of speaker awareness of different mutations, which found that the nasal mutation following (yn) had higher community awareness than the majority of AM triggers, particularly relatively archaic ones such as the numeral triggers (tri) and (chwe).

The study also refers back to data from Ball (1984) in to provide a comparison between community use of AM in different styles, and variation between different programmes:

Cwmtawe Study	Standard AM	Radio Study	Standard AM
Casual Style	0%	-----	-----
Reporter's Test	48%	<i>Stondin Ddyddiol</i>	31%
Interview Style	66%	<i>Oedfa'r Bore</i>	80%
Sentence List	94%	<i>News</i>	100%

Table 19 Mean standard use of all AM variables in community and radio styles (Ball et al 1988: 189)

The authors argue that the stylistic variation of individuals between the different elicitation tasks in the Cwmtawe study broadly reflects that of the three programmes, because “the same trends are obviously present” (Ball et al 1988: 189). Significantly, the authors have not compared the most informal programme *Stondin Ddyddiol* with the most informal *casual* style, in which AM was absent, because “completely casual speech is not to be expected from the broadcasts” (Ball et al 1988: 189). This finding allows the authors to make the following defence of the mutation patterns of presenters on *Radio Cymru*:

“The evidence presented here, then, suggests that radio Welsh is not, as some would claim, an externally imposed variety that native speakers cannot understand; nor, as others would charge is it guilty of promoting ‘corrupt’ or bad Welsh. Just as spoken Welsh varies according to style, so does radio Welsh, and reflects the usage of the speech community to a remarkable extent in the above examples” (Ball et al 1988: 190).

Conversely, I would argue that this is not clear at all from the data in table 19. The vernacular style in the Cwmtawe study has a mean standard use of AM of 0%, while the informal *Stondin Ddyddiol* has a mean of 30% - arguably quite a substantial difference. The argument that *Radio Cymru*'s use of AM reflects the norms of the speech community is thus to suggest that speakers do not want informal programmes to reflect vernacular unmonitored use of language – but

prefer the more formal style reflected in *Stondin Ddyddiol*'s use of AM. In Bell's (1984) terms this would reflect a *referee design* situation, where speakers prefer linguistic forms divergent to their own speech in certain domains. According to Bell, such situations are typical of diglossic contexts, and he describes one example as the implicit preference of New Zealanders for the British 'received pronunciation' variety of English for certain prestigious domains, such as news broadcasting. What isn't clear is whether such a situation really reflects community norms (particularly in terms of the consent of non-elite listeners to participate in such an arrangement), and whether, particularly in the case of the modern *Radio Cymru*, the situation is now different, given the process of de-standardization that I described in chapter 1.

3.3.5 Jones' language obsolescence study

The most recent sociolinguistic work on mutation in Welsh was carried out by Jones (1998) as part of her research into language obsolescence in two increasingly anglicized Welsh communities. Her study was set in the town of Rhymney in South East Wales and the village of Rhosllannerchrugog in North East Wales. Both communities are distinct both in terms of dialect and Welsh speaking population, with Welsh speakers more prominent in the latter than the former. Compared to parts of the West of the country however, Welsh speakers are a minority in both communities, and Jones' study primarily focuses on the erosion of the local dialects in the face of Anglicisation and the state school system, which is a vehicle for the diffusion of standard Welsh (Jones 1998). Mutation is only one variable among many which Jones uses in her study, and it is primarily measured in relation to the simplification of the Welsh grammatical system as opposed to as a sociolinguistic variable. There is, nevertheless, detailed data in the study showing relative use of standard mutation according to age group, which essentially represents an apparent time study. The variables Jones uses are a mixture of all three categories of mutation triggers, including SM, which Ball (1984) did not use. The variables Jones included are as follows:

Trigger	Meaning	Trigger type
(prepositions)	<i>am</i> (for), <i>ar</i> (on), <i>at</i> (towards), <i>gan</i> (with, by), <i>dros</i> (over), <i>trwy</i> (through), <i>wrth</i> (by), <i>dan</i> (under), <i>heb</i> (without), <i>hyd</i> (till), <i>o</i> (from), <i>i</i> (to)	SM
(Feminine nouns)	--	SM
(ei.m)	<i>ei</i> (his/its)	SM
(fy)	<i>fy</i> (my)	NM
(yn)	<i>yn</i> (in)	NM
(ei.f)	<i>ei</i> (her/its)	AM

Table 20 Variables used in Jones' (1998) study

Jones' inclusion of SM variables contrasts with Ball's (1984) decision to omit them, which was mainly based on the fact that he believed the SM system was largely stable in his dialect area. Jones' decision to use only a single AM variable, feminine pronoun (ei.f), is also an important contrast with Ball's study. Additionally, unlike Ball, Jones did not attempt to measure style shifting with a formal elicitation task. Thus, while Ball focused on AM triggers as the most likely markers of style shifting and social variation, Jones took the opposite decision based on the general lack of AM in the vernacular of the speakers in her study.

The results of the study do seem to vindicate Jones' decision to use SM variables, with some SM triggers showing strong variation, particularly in terms of age. One of the most significant differences found were between 'school aged' young people under twenty, and those who were older. Jones reports that standard SM mutations following preposition triggers were gradually declining in use among all speakers, with every subsequent age group using it less than their elders. Non-mutation following masculine possessive pronoun *ei* and *fy* however was described as an "exclusively school based phenomenon" (Jones 1998: 69). Jones found that school age children were increasingly using the (noun + *fe/fo*) pattern, which she argued represents a strategy developed to avoid the mutation environment. A similar pattern was found in terms of (*fy*). Similarly to Hatton's (1988) study, Jones found that NM following possessive pronoun *fy* was no longer a categorical rule for all speakers, with children using the radical in over half the possible environments, often using a (noun + *fi*) pattern.

Jones concludes that the drastic decrease usage of mutation amongst younger speakers could be a prelude to further decline in mutation in the future. She argues that one cause for this trend is the imbalance in Welsh schools between greater numbers of L2 learners of Welsh and progressively fewer L1 speakers, with both groups ultimately having fewer chances to hear standard forms of mutation. As Jones says: "if learners are less accustomed to hearing *fy* unaccompanied by a nasal mutation they will be less likely to produce it themselves" (Jones 1998: 63). The reduced presence of standard forms in the speech community would then be predicted to continue, as the younger generation grows and passes on its 'reduced' model of mutation to other speakers. Jones work highlights the fact that the Welsh mutation system has been increasingly adapted and simplified over the past decades, with the author linking this to the influence of L2 learners of Welsh on the speech community as a whole.

3.4 Psycholinguistics and Mutation

Thomas and Gathercole (2007) examined the ability of children and adults to apply AM in anaphoric reference – i.e. their ability to match the historically appropriate form of mutation to the gender of possessor nouns. The intention of the authors was both to see at what stage children acquired mutation, but also to confirm earlier theories on the ‘piecemeal’ nature of acquisition of the mutation system in Welsh. The researchers predicted that if children used an abstract system of learning to acquire the mutation system, providing them with cues as to a mutation’s gender would aid in completing an elicitation task. If the cue did not aid participants in completing the task, this could be considered evidence that acquisition of mutation is a process of piecemeal acquisition reliant on input.

For the cue condition, participants were cued with a sentence which contained the determiner *y* or *yr* (which triggers mutation in adjacent feminine nouns) preceding the possessor noun. In the non-cued condition, participants saw the possessor noun in an indefinite NP with no mutation triggered, and thus had no cue as to the gender of the possessor. In both conditions the participants then had to complete a *cloze* elicitation task, where a missing noun preceded by *ei* had to be provided. The results found that both children and adults were better able to apply the standard mutation for masculine possessors, which require SM, than for feminine ones, which require AM. Further, when participants applied a non-standard mutation in the case of feminine possessors, this typically involved overextending the soft mutation, rather than using the radical. Thomas and Gathercole explained this overextension of SM with reference to SM’s larger range of triggers and initial consonants that can undergo the mutation, as well as the fact that previous studies have found that *ei* often only undergoes AM when a possessor noun is animate. The results were also significant in terms of age, with older children performing better than younger ones, and adults performing better again. In terms of the cue conditions, the findings showed that cues to the gender of the possessor did not help

participants in using the correct gender. The authors claim that this provides evidence that Welsh speakers do not acquire the mutation system through an abstract system of rules, but piecemeal, on a word by word basis.

For further context on the role of speakers' knowledge of the mutation system, the authors decided to add two further experiments to the study, both of which tested participants' knowledge of mutation in non-gendered contexts. The first of these experiments tested speakers' use of standard mutation following the SM triggers prepositions *o* and *ar*, while the second tested standard mutation of AM following the triggers *na* and *a*. The results for these experiments showed that the children's application of mutation following the prepositions was less standard than following *ei* in the first experiment, but it also showed that the children gradually progressed towards adult usage in older age groups. This pattern did not occur for the non-gendered AM triggers, where the children applied the AM 2% and 4% of the time for *a* and *na* respectively, whereas the adults did so 42% and 57% of the time. The authors claim that this is evidence that at nine years old, children have still not acquired the AM system.

Overall, the authors suggest the results provide evidence that the acquisition of mutation is not a rule driven process, but is piecemeal, and based on the varying types of input that children receive. Children acquire the SM system first because of the large amount of opportunities for input, while acquiring the AM system may take substantially longer due to the scarcity and varying quality of the input. This finding seems to imply that the level of input a speaker receives for a particular mutation type should be a strong factor affecting their ability to control that variable stylistically, which perhaps vindicates previous sociolinguistic studies on mutation's tendency to focus on exposure and input as a central factor.

3.5 Mutation and attitudes

Little attention has been paid to mutation in terms of attitudes. This may be because as A.R. Thomas (1982) claims, variable use of mutation may be largely below the level of consciousness for many Welsh speakers. On the other hand, there is some evidence that mutation has been discussed in the media, and subject to “complaint culture” (Milroy and Milroy 1985). Ball and Müller (1992) provide the example of a Welsh language rock band named *Y Treigladd Pherffaith* who ignited a series of heated exchanges in the Welsh language media in the late eighties with their provocative use of hypercorrect AM mutation both in the title of their group and in the name of one of their albums. The title of the band translates into English as *The Perfect Mutation*, and involves an AM mutation whimsically applied to the adjective *perffaith*, even though there is no legitimate mutation trigger present. According to Ball and Müller, this led to some bad tempered correspondence in a local Welsh paper, with one individual accusing the group of warping young people’s grasp of Welsh, while others came to their defence.

3.6 Summary

Studies have offered various possibilities where stylistic markers are concerned. Because of its apparent stability (although cf. Jones 1998) the SM system has not been considered, since any candidate for the status of a marker will typically involve at least some social variation, and Ball and others see the SM system as generally stable. Ball (1984) emphasizes the style shifting of AM and NM triggers (yn), (a), (negative verb) and (â/gyda), while Hatton (1988) suggests that the NM triggers (fy) and (yn) are markers, even convincingly showing the acquisition of social competence in the style shifting of these variables by children of primary school age.

One of the main threads in variationist research on mutation is the centrality of input as a factor in driving variation, and most of the social categories which researchers use as replacements

for class in their studies are based around the concept of relative speaker exposure to standard forms of Welsh. This is justified to some extent by the findings of Thomas and Gathercole (2007), which has shown that children's acquisition of AM is likely to be strongly affected by level of input. However, it is noteworthy that no attempt has been made to analyse social meaning in mutation, with all studies assuming the universal prestige of 'standard' forms of mutation for all Welsh speakers, and using a conception of style similar to Labov's - where speakers respond to standard language pressure by moving along a linear scale. In summary there is some evidence that mutation varies in terms of both social categories and style, particularly for some AM and NM trigger variables.

4.0 Code-Switching

The study of code-switching is a diverse field which is characterized by cross-discipline research. Section (4.1) focuses on research into the structural constraints upon code-switching, which is where most code-switching studies on Welsh have focused. Section (4.2) looks at social factors which affect the use of code-switching, including the level of input that speakers receive, their age and their level of education. Section (4.3) moves on to consider style and code-switching, both in terms of discourse strategies such as flagging, retracing, direct speech and humour, and in variationist studies which have considered code-switching something like a sociolinguistic variable. Section (4.4) moves on to discuss attitudes towards code-switching, including what is currently known about the Welsh context. Finally, section (4.5) examines the use of code-switching in the Welsh media, particularly in terms of the normative pressures on presenters in the Welsh language television of the late eighties.

4.1 Structural Factors in code-switching

The definition of code-switching remains somewhat controversial, with some researchers preferring the alternative term code-mixing (Muysken 2000) for overlapping concepts. Broadly, code-switching is used here to mean cases where lexical items and grammar from two languages are used together in the same sentence. Some researchers (Sankoff et al 1990) advocate restricting the term code-switch to multi-word stretches, while others (Myers-Scotton 1993) accept single words as switches. Historically, code-switching research has developed from the view which saw it as essentially random 'free variation' (Labov 1971) to a focus from the eighties onwards on identifying structural constraints. In terms of language internal constraints, two of the most important models have been Myers-Scotton's MLF and the alternation based approach of Poplack and her associates.

Muysken (2000) takes a descriptive and synthesizing approach to code-switching, attempting to bring together different models for appraisal. Muysken views the seemingly disparate patterns of code-switching observed as different ways of surmounting the fundamental ‘impossibility’ of code-switching: “code-mixing is impossible in principle, but that there are numerous ways that this fundamental impossibility can be circumvented” (Muysken 2000: 30). In other words, speakers use varying strategies to circumvent the pressures that work against code-switching. Following Chomsky, these constraints are broadly divided by Muysken into two categories: e-language (social pressures) and i-language (language internal pressures). Among the e-language factors are the needs of the individual to support and maintain a single cohesive identity, and to conform to societal norms. The i-language factors involve pressures to form cohesive sentences which do not violate (or which circumvent) the grammars of both languages.

Muysken’s model divides the three main code switching patterns between three general societal situations. Insertion is described as being prevalent in contexts where two languages exist in an asymmetrical power relationship, for example in a post-colonial country. The asymmetrical relationship is thus reflected in the asymmetry of the syntax, with one language acting as a matrix into which the other language is inserted. Alternation patterns occur in communities with a “long tradition of language separation” (Muysken 2000: 8), and this is reflected by a pattern where the structure of both languages are kept basically separate. The final pattern, congruent lexicalization, involves a situation where the syntax of both codes are close enough for switching to potentially happen at any point in a clause. This is why Muysken associates congruent lexicalization predominantly with the tradition of style variation research evidenced in Labov (1972) and Trudgill (1974), which often focuses on variation from standard to non-standard codes, with more of a focus on external rather than internal constraints. However Muysken also cites the work of Clyne (1967) on German and Dutch migrants in Australia as

an example of a strictly bilingual structural analysis of this kind. Muysken's categories of insertion and alternation resemble relatively closely some of the most important recent models of structural constraints on code-switching. Myers-Scotton's Matrix Language Frame theory (2002) is an insertional model, where in 'classical' cases of code-switching one language takes the role of the linguistic frame into which another language can be inserted. Poplack's (1981) paradigm involves the alternation of material from two languages, and focuses on potential switch points between them. Both models differ in that constraints come from the need to obey the rules of both grammars in Poplack's model, while the focus is on the requirements of the dominant matrix language in Myers-Scotton's model.

Deuchar et al (2007) used a corpus linguistic study to identify whether language pairs could be associated with particular models of code-switching more than others. The researchers developed a point based method of categorizing each language pair as either primarily insertional, alternational or indicative of congruent lexicalization. The points based system also allowed for a secondary pattern to be established. The three languages pairs studied were Welsh-English, Tsou-Mandarin, and Taiwanese-Mandarin. These pairs were representative of different social and typological combinations. The analysis found that Welsh-English bilinguals typically use the insertional pattern of code-switching, with congruent lexicalization as a significant secondary pattern. The authors note that the Welsh-English bilingual community can be considered typical of insertional patterns in some ways, arguably representing a post-colonial situation (E. Williams 2009) as well as having a typologically dissimilar language pair. The fact that the majority of work so far on Welsh code-switching has adopted the MLF can be argued to justify Muysken's observation that asymmetrical social systems tend to lead to asymmetry in the type of switching that generally takes place.

Poplack (1981) formulated two linguistic constraints on code-switching, the *equivalence constraint* and the *bound morpheme constraint*. The *equivalence constraint* posits that

bilinguals code-switch in such a way that obeys the grammar of both languages. This means that code-switching should only occur at switch points, which are parts of a clause where both grammars are similar enough to allow an exchange of material. The *bound morpheme constraint* predicts that switched material cannot occur between a lexical stem and bound morphemes. In an analysis of sixty six hours of recorded Spanish-English conversations, containing 1,835 switches, none of the speakers violated either constraint. Poplack (1980) also found evidence that competence is linked to the type of code-switching that speakers produce. The more fluent bilinguals' knowledge of grammar allowed them to switch at points in each clause which did not violate the equivalence constraint, while less competent bilinguals tended to avoid this difficulty by switching between clauses.

4.1.1 Defining borrowings and switches

One of the most important contentious issues in the study of code-switching is the distinguishing of code-switching from borrowing. Poplack and associates (Poplack & Meechan 1998) consider borrowings and switches to be totally different phenomena, while Myers-Scotton's approach considers them to be essentially different points along the same continuum (1993). Those who support Myers-Scotton's model tend to advocate the idea of gradualism, where loanwords begin as momentary code-switches before becoming established borrowings after increasing use over time. According to the MLF model, borrowings are distinguished from switches by their 'predictability' which Deuchar (2006) relates to Muysken's notion of *listedness*, or the "degree to which a particular element or structure is part of a memorized list" (Muysken, 2000:71). Myers-Scotton thus would argue that one could use frequency as a way of measuring a lexical item's status as either code-switch or borrowing. If a word's frequency of use could be positively correlated with its integration (whether morphological, syntactic or phonological) into the other language, this would provide evidence for the gradualist perspective. Poplack and her associates however deny this correlation, arguing that frequency

does not play a role in categorizing words as either nonce borrowings or established borrowings (Poplack & Meechan 1998). This argument was formalized with the nonce hypothesis (Sankoff et al 1990) which states that borrowings and nonce-borrowings do not differ in terms of frequency.

A gradualist perspective was adopted by Stammers (2010) and Stammers and Deuchar (2012) who looked at the degree of integration of morphologically mixed verbs into Welsh. Their analysis used mutation of a word's initial consonant (as opposed to non-mutation) as an index of its integration into Welsh. The analysis found that speakers' tendency to mutate the target word was positively correlated with the word's frequency in the corpus. The authors argued that this represented evidence of the process of integration of English language verbs into Welsh, with lexical items being differentially absorbed into Welsh according to speakers' acceptance of them as either Welsh or English words, reflected in the application of mutation rules. Poplack (2012) has rejected these findings, arguing that mutation is a phonemic feature and is thus not covered by the nonce borrowing hypothesis, which only considers syntactic and morphological integration as representative of established borrowing. In their rejoinder, Deuchar and Stammers (2012) argue that mutation in fact is a phenomenon driven by syntactic and morphological processes.

The debate over the status of switches and borrowings is important because it has practical implications for any code-switching study. Despite disagreements over the exact process, most researchers would agree that transferred lexical items need to be differentiated in terms of their level of integration into the host language. In work on the *Siarad* corpus, Deuchar and associates have used the *dictionary criterion* (Deuchar 2006), which considers a word's listing in a dictionary as being indexical of its level of established usage in Welsh. If a word is only listed in Welsh dictionaries, it may be considered a Welsh word. If a word is listed in both Welsh and English dictionaries however, it is considered ambiguous. This criterion aims to

reflect the fact that if a word has become integrated enough to be included in dictionaries of the host language, it can be considered an established borrowing. Conversely, if a word is not included it is considered a code-switch, since it has not been integrated into the language.

4.1.2 The Matrix Language Frame and Welsh

Myers-Scotton's Matrix Language Frame is an approach to code-switching which focuses on the fundamental asymmetry of the phenomenon. According to Myers-Scotton's Matrix Language Frame approach (2002) two languages within the same clause will normally not be equal, with one language taking the dominant role as the matrix language, while the other language contributes only certain types of morphemes. The asymmetrical nature of code-switching is referred to as the *asymmetry principle*. In a *classic code-switching* situation this asymmetry is reflected by the fact that the grammatical frame is provided by the matrix language, which supplies the word order and a particular class of morphemes. The other language is then called the Embedded Language. The matrix language can thus be identified by with two diagnostic principles, the system morpheme principle (SMP) and the morpheme order principle (MOP). The MOP states that the matrix language provides the word order for all clauses, while the SMP states that system morphemes come from the matrix language. The SMP relates to the division of morphemes in bilingual clauses into content and system morphemes. Content morphemes are open class morphemes including verbs, nouns and adjectives, while system morphemes are closed class items which vary in their roles. For the SMP, the relevant type are *late outside system morphemes* which are morphemes with grammatical relations external to the head: "In Matrix Language + Embedded Language constituents, all system morphemes which have grammatical relations external to their head constituent ... will come from the Matrix Language" (Myers-Scotton, 1993:83; 2002:59, quoted in Deuchar 2006).

Deuchar (2006) attempted to establish whether Welsh-English code-switching was an example of *classic code-switching*, as defined by Myers-Scotton in her MLF Framework (2002). Deuchar utilized both the SMP and MOP to classify the matrix language of bilingual clauses in a corpus of informal speech between Welsh-English bilinguals. Both principles were adapted for Welsh-English bilingual clauses. In terms of MOP, a testable criterion was identified in the different word order of verbs and subjects – (English SV, Welsh VS) as well as head modifier phrases (English MH, Welsh HM). For the SMP the analysis focused on subject verb agreement. If agreement used Welsh morphology the matrix language was considered Welsh, while if English morphology was used the matrix language was coded as English. Analysis used the *dictionary criterion* as a practical method of distinguishing between established borrowings and genuine instances of code-switching. The analysis focused on bilingual clauses, which were those including code-switches. Results showed that the majority of bilingual clauses consistently used Welsh as a matrix language, leading Deuchar to claim that the Welsh situation was indeed a case of classic code-switching.

Davies and Deuchar (2009) used the MLF to explore a prediction by Alan Thomas (1982) concerning the future of Welsh. As I briefly mentioned in the first chapter, Thomas wrote an article predicting changes in Welsh due to the influence of English. As part of this model, Thomas predicted that demographic decline in the numbers of Welsh speakers and Welsh speaking communities would lead to a kind of degradation in the Welsh used by Welsh speakers. Thomas was writing about language death, and argued that transfer from English was stripping Welsh speakers of the ability to exhibit stylistic variation in their own language. He argued that code-switching in particular, which he defined as the “extensive use of borrowed lexical items” (1982: 218), was being employed by speakers as a compensatory mechanism to replace the monolingual stylistic variation presumably belonging to Welsh in the past.

The researchers inferred a three step process of language death from Thomas' statements. In the first, speakers would continue to use a Welsh grammatical frame, but would insert English words and phrases. In the second speakers would begin to use grammar from both languages. In the third and final stage, speakers would only use English. This model of language change was compared with Myers-Scotton's matrix language turnover hypothesis (1998), which posits a similar three stage shift. The methodology again used the MOP and SMP in order to identify the matrix language of clauses. Although it was expected that in most cases both criteria would point to the same matrix language for each clause, in very few cases the MOP and SMP actually indicated different matrix languages. For these instances the researchers added the concept of a *dichotomous matrix clause* to the model. This is a type of clause where the matrix language is shared between both languages, and it is considered indicative of language convergence. According to the researchers, this innovation to the MLF makes it possible to measure convergence between language pairs quantitatively, by measuring the prevalence of such clauses in a corpus. The results of the analysis found that there was only one dichotomous clause in the dataset, which was taken to mean that Welsh-English convergence was not prevalent.

The analysis found so few examples of shared grammar in clauses that the authors state "there is really no evidence for any progression to the next stage of their models" (2009: 34). The authors thus situate Welsh at stage one of their model, or as a case of 'classic code-switching' in Myers-Scotton's terms. They cite Myers-Scotton's claim (1998) that major social changes, such as foreign occupation, would be prerequisite for progressing further through the turnover. They also refer to Thomas' (1982) prediction that the grammatical turnover would coincide with further demographic decline in the proportion of Welsh speakers. At the time when the article was written in 2009, the most recent census results had indicated an increase in proportion of Welsh speakers, which allowed the researchers to predict a stable outlook for the

foreseeable future. The more recent results however (ONS 2011), which show a decline occurring particularly in heartland areas, make it clear that the status quo cannot be taken for granted.

4.2 Extra-linguistic factors and code-switching

It has long been known that code-switching does not occur in exactly the same way in every community, or even between social groups and individuals within the same community. Various factors which may shape the way in which speakers code-switch include the level of exposure to each language, age, level of education and identity. For her MA thesis Lloyd (2008) analysed the *Siarad* corpus for correlations between the frequency of code-switching, indexed by the proportion of English words used by each speaker, and a variety of extra-linguistic factors, including age, gender and level of linguistic input. Lloyd found that there was a strong significant relationship between speakers' ages and their frequency of code-switching. Younger speakers code-switched more frequently than older speakers, although speakers in the oldest age band code-switched slightly more often than the middle-aged speakers. Surprisingly, the analysis also found that participants who had gone to either bilingual or English only schools were less likely to code-switch than those who had gone to Welsh medium schools. Although this result is difficult to interpret, it may be related to the distinction between formal language acquisition in the classroom and naturalistic acquisition in the community, where code-switching is relatively widespread.

In another analysis drawing on the *Siarad* corpus, Parafita et al (2011) found a correlation between bilingual speakers' identity and their choice of language for the grammatical frame of clauses. The analysis compared Welsh-English bilinguals in Wales with Spanish-English bilinguals in Miami, and employed Myers-Scotton's Matrix Language Framework (Myers-Scotton 2002). The analysis found that 100% of the *Siarad* clauses were Welsh matrix language

clauses, while the Miami clauses were 66% Spanish Matrix Language and 34% English Matrix Language. The researchers explained the difference as being due to a number of both social and linguistic structural factors. In terms of social factors, the Siarad corpus speakers were far more likely to identify themselves as Welsh (90%), while the Miami speakers were from a varieties of backgrounds, identifying themselves as Cuban, American, Venezuelan and so on. Another factor cited was the fact that the Welsh-English bilinguals tended to have mainly Welsh speaking social networks, while the Miami bilinguals were more likely to have mixed English and Spanish speaking networks.

Another branch of relevant research has looked at speakers' knowledge of Welsh words of varying frequencies. It is important to note that expecting the breadth of Welsh speakers' vocabulary to shape their use of code-switching in natural speech could be misleading, as this ignores the extent to which code-switching may be stylistically driven (as I discuss in section 4.3 below). However, speakers' ability to draw from a broad lexicon of Welsh language equivalents could be a factor in their ability to avoid *crutching* (Zentella 1990), or lexical gaps in their speech. While not directly addressing code-switching, psycholinguistic work by Gathercole and Thomas (2009) has examined the influence of Welsh language exposure on the relative sizes of speakers' lexical repertoires. They collected data on the receptive vocabularies of children of various ages, and of different linguistic backgrounds; children from households with two Welsh speaking parents (only Welsh at home - OWH), two English speaking parents (only English at home - OEH) and one of each (Welsh and English at home - WEH). The researchers used the million word CEG corpus of Welsh texts to divide lexical items into the following eight bands of frequency (Gathercole et al 2008: 686):

1. 600-1000 occurrences per million words
2. 200-599 occurrences per million words
3. 150-199 occurrences per million words
4. 100-149 occurrences per million words
5. 75-99 occurrences per million words
6. 50-74 occurrences per million words
7. 25-49 occurrences per million words
8. <25 occurrences per million words

The results indicated that vocabulary knowledge increased with each progressive age group, and with words of greater frequency. The OWH children also knew significantly more Welsh words than both OEH and WEH groups at every level of frequency. For frequency bands 1-5, the WEH and OEH performed similarly to each other, but at lower frequencies (bands 6-8) the WEH children outperformed the OEH children. The researchers explained the connection between frequency and Home Language as being due to speakers requiring a set amount of input before being able to acquire low frequency words. This implies that speakers' likelihood of code-switching and using English equivalents because of lexical gap for some words will be increased if their exposure to Welsh is infrequent. Rhys and Thomas (2013) also relate the potential difficulties faced by some children in acquiring vocabulary in Welsh to problems of access to Welsh language reading materials. The Welsh Government has targets for all children in bilingual schools to reach the reading age in Welsh, although it concedes that "we are clearly falling short of that at the present time" (Welsh Government 2011, quoted in Rhys and Thomas 2013). As I discussed in the first chapter, it is implied that L1 speakers of Welsh are also amongst those suffering from a lack of exposure to literary Welsh, which is due to both the paucity of material available and the perceived unattractiveness of what is on offer (Rhys and Thomas 2013: 649).

One topic that has not been truly resolved is whether socioeconomic class may be correlated with any particular pattern of code-switching. Jilg (2003) recently conducted a study on the linguistic competence of Welsh speakers in the village of Blaenau Ffestiniog, and found that

two different communities of practice, *pobl y petha* and *pobl y dafarn*, differed in the way in which they used certain lexical equivalents. The latter tended to substitute Welsh words for English lexical equivalents and borrowings, while the former were more likely to maintain Welsh equivalents. *Pobl y petha* could perhaps be loosely translated as “the artsy people”, and *pobl y dafarn* “the pub-goers”, and Robert (2011) has argued that these groups could map on to class distinctions. Given the mounting suggestions that class in Welsh is becoming a salient factor for Welsh speakers (G. Williams 1987; Madoc-Jones et al 2013) Robert calls the influence of class on speakers’ lexical competence in Welsh “a question that needs to be asked” (Robert 2011: 138). Once again, it may not be the case that low competence in using standard Welsh lexical forms necessarily translates to more frequent use of code-switching, but it may be an important factor.

4.3 Style and code-switching

Studies into what might be considered the stylistic use of code-switching have largely taken place apart from mainstream of sociolinguistic research (which I described in chapter 2), and can be divided into two main approaches – *discourse analysis* and *conversational analysis*. Discourse analysis includes the work of researchers who consider codes to be discrete entities, which signal different indexical qualities through their associations with different social groups (e.g. Gumperz 1982, Myers-Scotton 1993). Conversation analysis on the other hand, is a theoretical and methodological approach which views such associations as less fixed, and more contingent on speakers’ aims in a conversation (Auer 1998). Alfonzetti, for example, has argued that speakers do not necessarily always “call into play the social and symbolic values of the codes in the repertoire” (Alfonzetti 1998: 180), but often use code-switching for more immediate and pragmatic reasons, such as drawing attention or managing turn allocation. As I am primarily approaching my data in terms of more traditional sociolinguistic concerns with

indexicality and social meaning (e.g. Eckert 2008), I will henceforth focus mainly on discourse analysis studies.

Like other kinds of stylistic variable, researchers have found that stylistic use of code-switching is tethered in communal norms. For example, Blom and Gumperz (2007) have defined code-switching situations as falling into two broad categories, *situational* and *metaphorical* switching. Situational code-switching refers to code-switching which is determined by the context, to the extent that one code will be considered exclusively appropriate for particular situations:

“The linguistic forms employed are critical features of the event in the sense that any violation of selection rules change members’ perception of the event. A person who uses the standard where only dialect is appropriate violates commonly accepted norms. His action may terminate the conversation or bring about other social sanctions” (Blom and Gumperz 2007: 88).

Situational switching is thus associated with the concept of diglossia, and the perception within a society that certain domains are more or less exclusively connected with certain codes. An example from the Welsh context is K. Jones’ (1995) analysis of code-switching in different domains, where she found that Welsh was considered the exclusively appropriate code in Welsh religious services. Situational switching also takes place between different interlocutors, for instance where speakers might switch from the dialect to the standard variety in order to engage with a tourist who does not speak the local code (Blom and Gumperz 2007). Metaphorical switching by contrast, involves switching codes for “particular kinds of topics or subject matter” (Blom and Gumperz 2007: 88) even where interactional participants and the general situation stay the same. This kind of switching depends on the pre-existing semantic

relationships between each code in the minds of speakers, which allows alternation between them to convey specific indexical meanings.

Another crucial concept introduced in Gumperz' (1982) work is that of 'we' and 'they' codes. As Gumperz states: "the tendency is for the ethnically specific, minority language to be regarded as the 'we code' and become associated with in-group and informal activities, and for majority language to serve as the 'they code' with the more formal, stiffer and less personal and out-group relations" (Gumperz 1982: 66). Myers-Scotton's markedness model of code-switching (1993) proposes that linguistic codes are indexical of different rights and obligation sets for speakers in particular communities. Codes are thus resources upon which speakers can draw strategically, either to impose authority or to claim solidarity with other speakers. Again, the particular linguistic norms of each community are crucial, since in some communities code-switched speech in itself might be the unmarked choice, with any movement to monolingual speech subsequently constituting a marked choice. In the next sub-sections, I will discuss specific discourse strategies of *direct speech* (4.3.1), *flagging* (4.3.2), *repairing speech* (4.3.3) and *humour* (4.3.4) as examples of the way in which code-switching analysed at the level of discourse can provide useful insight into underlying ideologically driven evaluations of both codes in a speech community.

4.3.1 Direct speech

Gumperz (1982) states that direct-speech, the quotation of another speakers words, is one of the most frequent discourse functions with which code-switching appears. Some researchers have considered this kind of code-switching unremarkable, with Gal (1979) claiming: "[...] quoting is relatively predictable. All one needs to know to predict the language in which most quotes will be spoken is the language in which the original utterance was spoken" (Gal 1979: 109). In contrast to this however, other researchers have argued that direct speech can reveal

evidence about “both how identity is attributed by speakers and of the symbolic value of the different codes” (Sebba and Wootton 1998: 282). For example, in Álvarez-Cáccamo’s (1996) analysis of Galician-Spanish code-switching, the codes are put into contrasting positions through the use of Galician as the ‘we-‘ code frame to narrate an event, and the use of Spanish as the ‘they-‘ code reserved for reported speech, marking the traditional perception of Spanish as the outgroup language, while Galician is emblematic of ingroup solidarity. However this pattern was contrasted in an interaction in which the patronizing words of an elite Galician speaker, who represented for the speaker the Galician speaking establishment, is quoted in Galician, while the narration is framed in Spanish, in a reversal of the roles ascribed for the ‘we-‘ ‘they-‘ codes in the previously described (switch to Galician highlighted in bold) :

Estudiante: negociar cómo ?

[hi] como siempre ---> presentando nuestras quejas ?

para que nos digan --->

bueno tranquilinhos

...que ao me*lhor

.. dentro de três ou quatro anos fazemos caso ?

English version

Student: Negotiate, how?

As usual, by presenting our complaints so that they, tell us.

"Well,

you take it easy there,

and maybe

in three or four years we'll listen to you" ?

Extract 7 Direct speech indexing power roles in Galicia (Álvarez-Cáccamo 1996: 38)

This use of direct speech to attribute Galician to the establishment reflects the sociolinguistic change which has taken place in Galicia recently, where, as in Wales, after an extended period of bilingual diglossia: “Bilingual elites have emerged or re-emerged, and have taken partial control of the administrative, political, and cultural resources of the newly decentralized state” (Álvarez-Cáccamo 1996: 35). Here then we see that the decision to report another speakers’ words is not constrained by the original source of an utterance, since the quote involves symbolic speech that never necessarily took place, and the switch to Galician is intended to convey a stylistic effect through its indexical associations. As Álvarez-Cáccamo states in response to Gal’s (1979) conception of direct speech as an objective reproduction of speech that has already taken place: “This emphasis on the 'triggering' effect of context for code-switching underplays the speaker's power to renew context itself” (Álvarez-Cáccamo 1996: 42). In other words, direct speech should be considered a stylistic phenomenon.

4.3.2 Flagging

Another practice associated with code-switching is flagging, where speakers draw attention to the use of code-switching with meta-commentary, hesitations and other paralinguistic features (Rosignoli 2011). One interesting aspect of flagging is that the prevalence of the strategy can reflect the extent of normative attitudes towards code-switching in a community. For example, Muysken (2000:106) has stated that flagging often reflects social situations in which there is social stigma against code-switching, and a preference for monolingual speech. Flagging was first analysed by Poplack (1987), who compared code-switching patterns among French English bilinguals in Quebec with those of the Spanish-English speaking Puerto Rican community of New York. Whereas the Puerto Ricans code-switched smoothly and frequently,

showing little awareness of switching, the Canadian participants were far more likely to flag switches, drawing attention to their use of other language material with meta-commentary, hesitations and other paralinguistic features. For instance in the following example, a French speaker flags the use of the English term ‘cloud nine’ both with hesitations and an explicit justification through the distancing phrase *qu’ils appellent* (“as they say”) (Poplack 1987:61):

J’ai accepté le Seigneur là, ben... j’étais comme sur un... cloud nine, cloud nine qu’ils appellent.

“I accepted the Lord then, well... I was like on a... cloud nine, cloud nine, as they say.”

Poplack argued that the distinction between both communities was due to different community norms and cultural attitudes towards code-switching formed in separate historical contexts. For the Puerto Ricans, code-switching is “emblematic of their dual identity” and “smooth, skilled switching [...] the domain of highly fluent bilinguals” (1987: 71). This contrasted with the French Canadian community, who were more likely to have negative views of code-switching, as well as lower competence in English. Flagging can involve numerous different strategies to attract attention, ranging from fillers such as ‘uh’ and ‘um’ to explicit metacommentary (Muysken 2000). Rosignoli (2011: 109) uses the following typology to code flags in his Italian-English data:

- i. Paralinguistic markers (e.g. laughter, nervous coughing)
- ii. Unfilled pauses (silence)
- iii. Lexicalised filled pauses (like, you know)
- iv. Non-lexicalised filled pauses (uh, uhm)
- v. Other-language equivalents
- vi. Explicit pre-empting (pre switch only)
- vii. Explicit justification (post switch only)

- viii. Disfluencies, false starts, retracing
- ix. Vowel/syllable lengthening

Flagging has also been considered an index of the level of integration switches into a host language, with both Jones (2005) and Rosignoli (2011) finding low frequency switches more likely to be flagged than higher frequency ones. This provides further evidence that flagging may be indicative of bilingual speakers' desires to avoid code-switching for normative reasons.

4.3.3 Repairing Speech

Speech repair involves speakers reformulating utterances from a situationally inappropriate code that they have “almost inadvertently used” (Alfonzetti 1998: 185) to another. This normally either occurs in situations where speakers' dominance in one code outstrips another, such as in migrant communities (Auer 1984) and among Erasmus students (Torres and Gafaranga 2002), or in diglossic situations where the highly conventionalized use of a vernacular must be avoided in formal situations (Alfonzetti 1998). In the first two situations the repair is arguably taking place because of lack of fluency in the target language, while in the latter it may be driven by self-censorship, and the need to respect the community norms for situational switching. Alfonzetti highlights the direction of repairs as being particularly meaningful, as this provides insight into the social evaluation of each code: “Indeed the switching [...] highlights a conflict between norms of situational appropriateness and spontaneity of linguistic usage. It therefore provides indirect cues about the speaker's individual preferences and competences and also about the sociolinguistic evaluation of the two languages” (Alfonzetti 1998:185). Alfonzetti provides the following example of this process in action, where a speaker inadvertently begins using a dialect word before swiftly reformulating it to standard Italian (my emphasis):

“Forse sarebbe sarebbe più giusto invece di non fare, cioè di non fare trattenute dallo stipendio e fare un’assicurazione a vita cioè su di noi. Se abbiamo *spis*- cioè spese per il medico, per i medicinali che ce li paghiamo da noi stessi”.

*“Perhaps it would be it would be fairer instead to make, that is to make no deductions from our salary and to take out life insurance, that is on us. If we have **expense**- that is expenses for the doctor, for medicines, we’ll pay for them ourselves”* (Alfonzetti 1998:186).

Extract 8 Example of speaker repairing the use of Sicilian in their speech, reformulating to standard Italian

Torres and Gafaranga (2002) describes the same process as *medium repair*, in the following terms: “(i) medium repair is a strategy used by bilingual speakers to solve a word problem (not juste, lapse, etc.); and (ii) in medium repair, participants draw on a language other than the current medium while orienting to this other language as a repairable matter” (Torres and Gafaranga 2002). In Torres and Gafaranga’s research, the repairs take place in the context of second language speakers communicating with one another. In this context, the interlocutor can ‘aid’ the speaker in repairing the other language material, by supplying the word most appropriate for the current medium. The example below takes place in an interaction between speakers at an Erasmus office:

AAA: no (.) I'm going to give this mmm (.) eh today (.) maybe today or tomorrow you will be inscribed

BBB: uh

AAA: matriculated (.) and after this eh it has to wait (.) four five six **jours** eh six ...

BBB: **days**

AAA: days (.) after being...

Extract 9 Other-repair by interlocutor in conversation between L1 and L2 English speakers (Torres and Gafaranga 2002: 533) Emphasis added.

In the example, although English has been adopted as the medium of interaction, speaker AAA has slipped into French inadvertently, using the word *jours*. This leads to a hesitation and an attempt at repair, until the interlocutor BBB provides the English equivalent 'days'. These examples taken together suggest that analysing the direction of speech repair can reveal underlying tensions related to the sanctioned use of codes in particular contexts, particularly in terms of illustrating which code is considered most appropriate for a particular setting.

4.3.4 Code-switching and humour

A connection between code-switching and humour has been attested in many bilingual communities (Woolard 1988, Gardner-Chloros and Finnis 2004). Gardner-Chloros and Finnis (2004) described humour as one of the primary uses of code-switching in their study on women in the Greek-Cypriot community. Speakers in the study used humour to channel different kinds of personas and values associated with each code to humorous effect. For instance, in the following example, one speaker switches to Greek to convey a Greek persona "the laid-back type who won't bother with too much detail" (Gardner-Chloros 2009: 85):

MI: ...??? Happen to know anyone that has like a colour laser jet

FI: I know a place where they do???

MI: yeah

FI: ???

MI: What make are they?

FI: **En iksero, en leptomerias**

I don't know, these are details

[General laughter]

Extract 10 Humorous use of code-switching between Greek-English Bilinguals (Gardner-Chloros and Finnis 2004: 524)

This extract suggests a potential difficulty for analyses of humour and code-switching. Understanding why such a code-switch might be humorous demands an emic (Brend and Pike 1976) understanding of the indexical associations of linguistic forms from the perspective of the speech community itself, which may not be immediately available to outsiders. For example, Siegel (1995) describes Fijian code-switching into Hindi, where Hindi insertions are often met with hilarity by community members for reasons that can seem baffling to outgroup members. Siegel illustrates this using the following extract from his data, where a family member is initially addressed by another in Fijian, and in a respectful manner:

yadra vinaka saka.

'Good morning, sir.'

This is immediately followed by switch to Hindi, which elicits laughter from those present:

Ao, baito.

come sit

'Come in, sit down.' [general laughter]

Extract 11 Humorous code-switching from Fijian to Hindi (Siegel 1995: 97)

As Siegel comments: “What's so funny? The joke would not be clear at all to an outsider, as there is nothing humorous in the referential meaning and no visual or paralinguistic clues. The humor is provided simply by the switch from Fijian to Hindi” (Siegel 1995: 97). Siegel's explanation centres on incongruity. In Fijian society, Hindi is a negatively evaluated ‘they-‘ code, with strong norms against use in most situations, which Siegel links to the mild ethnic tensions which exist between ethnic Fijians and their Indian neighbours. These restrictive norms then lead to tension surrounding the marked use of Hindi in social interactions. As Goldstein argues, this kind of mild breaking of social rules can be in itself amusing: “Violation of a rule of language results in one or other kind of incongruity, and incongruity is often amusing.” (Goldstein 1990: 103, quoted in Siegel 1995: 103). A similar situation can be seen in the work of Woolard (1988), who has described the stand-up comedy of a Barcelona comedian who uses code-switching to elicit humour. In Siegel's analysis “it is not creative language play which is the source of humor but rather the ‘promiscuous’ use of code-switching itself in situations where strict separation of language is the norm” and this use allows speakers of languages in ethnically tense situations “relief from the rigid separation of codes” (Siegel 1995: 104).

4.4 Variationist approaches to code-switching and style

The previous section covered analyses of code-switching which resemble speaker design approaches, i.e. where speakers are revealed to be using particular discourse strategies in ways which draw from the indexical meanings of each code. This section will focus on the application of variationist theories of style to analyses of code-switching. Variationist studies attempting to measure quantitatively the link between code-switching and style seem to have been relatively infrequent (Zentella 1990). Despite this, multiple researchers have commented on a general link between code-switching and informality. For example, Gumperz states (emphasis added) : “Code switching is perhaps most frequently found in *the informal speech* of those members of cohesive minority groups in modern urbanizing regions who speak the native tongue at home, while using the majority language at work and when dealing with members” (Gumperz 1982: 65). Labov (1971) himself however notoriously dismissed code-switching as a sociolinguistic variable, as he considered participants’ code choice to be essentially idiosyncratic - in free variation rather than correlation with social or stylistic factors. Dorian (1994) used a broadly variationist approach to measure shifting use of code-switching in an East Sutherland Gaelic speaker’s narratives. Dorian claims that the speaker varied her mean proportional use of English insertions from one narrative segment to the next, which she argues represents style shifting according to topic. Zentella (1990) provides one of the most intriguing attempts to study code-switching and style, with a methodology that combines both quantitative and qualitative approaches. The ethnographic component of Zentella’s study involved participant observation in a Puerto Rican immigrant community in New York. Over 100 hours of speech was recorded using recording devices placed in local children’s backpacks, and from this data the author developed a paradigm of eight distinct styles which are listed in table 21.

Style	Paca	Linda	Maria	Eli	Sara	Total N / hours of recording	Switches per hour
Casual conversation	256	386	524	116	54	1336/70	19.1
Games	6	0	2	8	-	16/10	1.6
Narratives	6	34	79	32	91	242/9	26.9
Buying	0	0	9	-	-	9/7	1.3
Language/lessons	12	-	10	24	-	46/3	15.3
Taping/interviewing	0	4	5	11	0	20/2	10.0
Jokes	-	-	-	14	0	14/1	14.0
Egocentric speech	1	-	1	-	0	2/1	2.0

Table 21 Code-switching patterns in different styles (Zentella 1990: 83)

To calculate scores representing individuals' use of code-switching in each distinct style, Zentella used a method of counting the frequency of switches per hour. According to the author, the patterns roughly reflect a tendency for speakers to code-switch more in informal situations than formal ones, with the highest frequencies in the casual conversation and narrative segments: "In style shifting from formal to more casual, all children demonstrated their ability to employ code-switching to greater effect in their most relaxed social situations, such as those likely to occur in the street or in the home" (Zentella 1990: 83). The author makes an explicit link between her conception of formality and Labov's attention to speech model, with reference to Labov's (1972) claim that narratives elicit the most informal speech because "they are the most unmonitored part of discourse" (Zentella 1990: 83). Similarly to Zentella, Poplack (1981) describes the speech of one Puerto Rican informant who's use of code-switching was four times more frequent (per hour) in informal than formal situations. However, Poplack also states that the identity of the interlocutor was crucial in modulating this kind of style shift – in informal

settings where she, an outgroup member, was also present, the code-switching use of the informant remained at the same low level as in formal situations.

4.4 Attitudes toward Code-Switching

Gardner-Chloros (2009) states that there are “surprisingly few” systematic studies of attitudes towards code-switching, although she notes that there are recurring tendencies. One tendency is for speakers to associate code-switching with laziness, as it is seen as “an easy way out when people cannot be bothered to search for the words they need in a single language” (Gardner-Chloros 2009: 14). Speakers also are typically likely to associate code-switching with a lack of linguistic skill, rather than stylistic choice, with participants often assuming that it mainly involves lexical gaps or ‘crutching’ rather than style (Zentella 1990). Code-switching is also often seen as indicative of a lack of intelligence and education (Gumperz 1991).

There are also studies which have found positive evaluations of code-switching. Gardner-Chloros notes that positive evaluation of code-switching “tends to coincide with a laid-back attitude towards authority; for example within the same community it tends to be more common in younger rather than older generations” (Gardner-Chloros 2009: 15). Another split is between more and less ‘developed’ societies, and Lepage and Tabouret-Keller (1985) have linked negative attitudes with the Western model of centralized state systems, which typically involve the development of strong standard language ideologies. In this sense attitudes towards code-switching are clearly not in any sense ‘natural’, but learned, and vary depending on context. For example, in some African contexts code-switching is seen as a positive social convention which allows speakers to show respect towards speakers of other languages by “meeting halfway” (Finlayson and Myers-Scotton 1998: 396). Various studies on Spanish-English speakers in the USA have found that speakers often regard code-switching as a positive marker of their dual identity and proficiency in both languages. The next section moves on to discuss

an important study that attempted to measure attitudes towards code-switching and borrowing in Welsh using empirical methods.

4.4.1 Empirical Evidence on Attitudes in Welsh

Robert (2009) attempted to measure Welsh speakers' attitudes towards different varieties of Welsh, including regional varieties and levels of competence. Robert used an adapted version of Lambert's (1967) Matched Guise Test (MGT), which usually involves the use of an actor able to produce different variants of sociolinguistic variables. Sample utterances containing these variants are then interspersed with fillers, with the aim of convincing listeners that all the voices are in fact different people. Participants are then asked to rate the voices they hear on various scales, including social prestige, attractiveness and ethnicity. The aim of the MGT is to elicit indirect evaluations which are intended to reveal the participants genuine underlying attitudes towards linguistic features.

Rather than evaluating others as more or less standard speakers along a classical socioeconomic scale, Robert had speakers evaluating each other along a continuum of perceived competence in Welsh, measured by the frequency of use of "innovative features through the influence of English (real or presumed)" (Robert 2009: 97). The MGT was adapted for the study by recording female teenagers discussing comic strips with the text removed. Six recordings were selected, with speakers selected to represent different regions, early or late acquisition of Welsh and their level of competence in Welsh. Two L1 speakers were chosen, one from North West Wales and the other from South East Wales. The four L2 speakers were selected for their varying competencies in Welsh. The recordings were then evaluated by eighty five speakers at the national Eisteddfod (a Welsh cultural festival) who were stratified by age, gender and geographical region. The participants were given nine questions which asked them to evaluate

different aspects of the voices on the recordings, which were related to social attractiveness, prestige and ‘Welshness’.

The results indicated that linguistic features showing English influence on speech had a strong impact on evaluations, and the speakers who exhibited the most transfer from English were also the most poorly evaluated on the social dimensions. Although both L1 speakers were evaluated relatively highly in terms of the ability and ‘Welshness’ measures compared to the L2 group, any use of non-traditional features by these speakers was still negatively evaluated. For instance, Robert argues that the North West L1 speaker was rated relatively poorly on the social attractiveness and prestige measures directly because of her frequent use of English origin words. Some of the free-response data refers directly to her use of *bratiath* (literally ‘broken language’), the colloquial and pejorative Welsh term for language mixing. Another free-response comment describes her as possessing “*meddwl a siarad llac*” (“Loose thinking and speech”) (Robert 2009: 111) which Robert associates with prevalent attitudes towards what code-switching represents in individuals, namely laziness and a lack of good self-presentation. Conversely, the speakers who were evaluated most positively in Robert’s study were those who used fewer elements of transfer and those who exhibited either local dialect or standard Welsh instead. The L1 speaker from South West Wales used predominantly colloquial regional forms from that area. Meanwhile, an L2 speaker whose speech contained more standard forms was also evaluated very positively, with some indication of a socioeconomic appraisal in one free response comment “*acen posh Cymraeg, cyfryngol*” (‘posh Welsh accent, media-like’). The implication here is that standard Welsh, along with its presumed lack of English influence, is in fact associated with a class of prestigious and relatively higher socioeconomic category of speakers. Ultimately then, dialectal or standard forms of Welsh were evaluated more positively than forms which reflected the influence of English.

In her conclusion, Robert warns that speakers (and particularly learners) who use English influenced innovations in their speech, and who are thus evaluated negatively in terms of social prestige and attractiveness, may find trouble integrating into Welsh speaking communities. It might be significant that part of this study was carried out at the annual Eisteddfod, which has an emphasis on purely Welsh language activity, and might be considered a bastion of standard language ideology. It is possible that a broader recruitment from outside this particular context, particularly perhaps with younger speakers, might have found more positive attitudes towards these kinds of features. Indeed, Robert herself suggests that there is a general change in attitudes towards code-switching currently in process, particularly among younger Welsh speakers. One example she uses is the work of Welsh language author Llwyd Owen, who writes gritty novels set in the Welsh capital of Cardiff. The characters of Owen's novels often incorporate code-switching, as well as other elements of English influence into their speech. A recent study (Orwig 2013) also found frequent code-switching in two Welsh language novels. If this can be taken to be part of a trend, such a development may herald a wider shift in attitudes.

4.5 Code-switching and the Welsh media

In Ball et al's (1988) study on Welsh in broadcasting, the authors discussed the controversial use of code-switching on Welsh medium television. The authors describe the output of S4C at the time as subject to a strong diglossic separation of code use, according to the subject of each programme: "The factual programmes – news, current affairs, documentaries – all employ a high variety of Welsh" while "low varieties of Welsh occur, as would be expected, in the light entertainment programmes" (Ball et al 1988: 193). The authors also explicitly make a connection between these categories and the degree of English transfer involved: "Low varieties of Welsh tend to exhibit interference in the lexicon and on the phonological, morphological and syntactical levels. High varieties of Welsh, on the other hand, resist

interference on all these levels” (Ball et al 1988: 192). The lengths that ‘high’ programmes went to maintain the distinction between codes at the time is striking - in order to avoid the use of English language material, programme makers would go so far as using Welsh language voice-overs which would be played over interviews with non-Welsh speakers. This strategy is described as being used as “the standard response to the occurrence of English in a programme” (Ball et al 1988: 195), although in a restricted set of situations the original English was allowed to remain without Welsh dubbing.

As discussed in the section on the discourse functions of code-switching (Gumperz 1982), this strategy reflects ideological norms in terms of the appropriateness of both codes, with the use of English strongly dispreferred in formal contexts. This norm could be broken on some occasions however, particularly when taking into account the possible low comprehension of standard Welsh features by some speakers (A. Thomas 1982). In such cases the researchers describe a strategy where presenters translate low frequency Welsh words to English: “If a vocabulary item is likely be unfamiliar to the audience is deployed, as for example, in programmes which take the language into domains usually exclusively English, the commentator/presenter will frequently gloss the Welsh with the equivalent English” (Ball et al 1988: 193). This strategy illustrates that the ideological use of standard Welsh can be balanced on some occasions by pragmatic considerations involving facilitating audience understanding of the actual content.

In the ‘low’ programmes, the authors describe the kind of switching predominantly involved as metaphorical switching: “being used to express irony, satire, humour and also anger and disagreement” (Ball et al 1988: 193), although they unfortunately do not provide examples of these strategies being used. The strategic use of code-switching to create humour, and to increase the social distance between interlocutors has of course been already attested in other language contexts by many researchers, as discussed above. What is interesting here is that the

authors are claiming that contributors generally use code-switching stylistically and artistically, rather than as a result of lexical gap. This suggests that the use of code-switching is likely to have the qualities of a sociolinguistic marker, although diglossic mechanisms of high and low domains may also play a role.

4.6 Summary

Code-switching emerges as a promising candidate for an index of formality, given the very strong association between it and behaviours which typically index informal contexts, such as humour (Siegel 1995), and the powerful social stigma levelled against it in Wales (Robert 2009). These factors suggest that its use is likely to be highly salient to speakers. Research by Zentella (1990) has also provided evidence that code-switching can be used for analysis in a way akin to a traditional variationist sociolinguistic variable, with informal contexts associated with low audio monitoring correlating with increased use of code-switching, while situations associated with higher self-monitoring like teaching are correlated with less frequent use of code-switching. However, as Zentella herself suggests, this kind of method is best supported by a mixed methodology approach. The contributions of researchers like Gumperz (1982) and Auer (1998) have convincingly highlighted the extent to which style in code-switching use is constructed through particular discourse strategies, which a simple count of switches versus Welsh words in the data would fail to account for. For this reason, this thesis will combine quantitative analyses of vernacular and radio data (chapters 6 and 8) with a qualitative discourse analysis of code-switching in the radio programmes (chapter 9).

5. *Siarad Corpus Analysis Methodology*

5.0 Introduction

This chapter will focus on a quantitative analysis of the mutation and code-switching in the *Siarad* corpus, a natural speech corpus using which I intend to represent a baseline of informal speech for a broad range of Welsh speakers from various demographic backgrounds. The main aim of this analysis is to provide contextual information to support the *Radio Cymru* corpus analysis. The *Siarad* corpus here is intended to serve as an unmonitored naturalistic contrast to the performative and ritualized contexts of the *Radio Cymru* corpus. The analysis centres on two categories of variables, mutation and code-switching. The mutation category can be broken down further, covering eight different variables. A further goal of this aspect of the study is to attempt a variationist analysis of these variables which is accountable to the data (Tagliamonte 2012) and which may provide further information concerning the interplay of external social factors like age, sex and cultural orientation, as well as internal factors such as initial consonant and word type.

5.1 *The Siarad Corpus*

The *Siarad* corpus is a large corpus of around 40 hours of informal conversations between 151 participants, who are all Welsh-English bilinguals. The corpus was collected by a group of researchers working at the ESRC Bilingualism Centre for research at Bangor University between 2005 and 2008 (Deuchar 2014). Much of the detail concerning the collection and transcription of the *Siarad* corpus which I describe below comes from the comprehensive documentation file for the project, the *Documentation File for the Bangor Siarad Corpus* (Deuchar 2014) which is freely available at the BangorTalk website: <http://bangortalk.org.uk/>. Although I was not part of the team that originally collected, transcribed and coded the corpus, I did perform some maintenance and correcting work in 2009-2010, and am therefore relatively

familiar with the corpus. Before I describe my methodology for coding and analysing the data, I will briefly describe the way the original research team collected (5.1.1) and transcribed (5.1.2) the corpus. I will then describe the demographic details of the *Siarad* corpus participants (5.1.3) before moving on to a brief discussion of the level of formality that the corpus represents, and why I believe it to be representative of informal Welsh (5.2). From section (5.3) I will begin discussing my own analysis of the *Siarad* data, mainly focusing on how I found mutation variables in the data, and then identified and coded them before subjecting them to statistical analysis using RBRUL. The final section explains my analysis of code-switching (5.11).

5.1.1 Collection of the *Siarad* Corpus

The interviews were recorded in various places, including in participants' homes, at Bangor University itself and at music festivals, depending on the preferences of the interviewees. The typical procedure of recording interviews involved the researcher setting up recording equipment at the selected location, and then leaving the participants alone to converse for a fixed period of half an hour. The researcher would then return and inform the participants that the interview could come to an end. There were a few cases where the researcher had to return in order to deal with additional questions or technical difficulties. The majority of the recordings were made using a Marantz hard disk recording in stereo, along with radio microphones for the participants. In a few cases where the Marantz was unavailable a Minidisk recorder was used, which was only capable of recording in mono.

Once the recording was finished, the participants were given questionnaires to fill in with metadata for sociolinguistic analysis. The questions included demographic details such as age and gender, as well as questions designed to measure language attitudes, level of linguistic input received in both of their languages during childhood, language proficiency and so on. As

well as filling the questionnaire, participants were asked to sign a consent form providing permission for the recording and transcripts to be used for research purposes, and for the data to be uploaded to internet under an open license. The consent form also assured participants that the data would be anonymized with the use of pseudonyms for all participants. For children under 16, the consent form was given to a parent or guardian to complete.

5.1.2 Transcription of the Siarad Corpus

Transcription was performed by a team of five trained transcribers who were fully bilingual in Welsh and English. The software chosen for the process was CLAN, a programme developed to facilitate bilingual transcription and coding. The transcription system was CHAT, also developed by Brian MacWhinney (2000), and all transcription used the CHAT manual as a standard. The transcription system uses four (and sometimes five) tiers; a main tier, a free translation tier, two gloss tiers and an optional comments tier. The aim of the main tier is to represent speakers' language as accurately as possible. Since the original projects' aims were focused on the syntax of code-switching, the transcription structure was based around dividing the material into clauses:

“Where an utterance contains two main clauses, each clause in that utterance is written on a separate main tier. Complex clauses are treated as one clause and therefore subordinate clauses are included in the same tier as their main clauses. Adverbial clauses are also written on the same main tier as their related main clause” (Deuchar 2014: 5).

5.1.2.1 Transcription of non-standard forms

One challenge for the researchers was how to represent colloquial Welsh, particularly since there has been so little work on informal varieties of the language. Contractions which were not in Welsh reference dictionaries were usually represented using parentheses to enclose the

elided sounds, so for example, the informal variant of *fel yna* (“like that”) which is pronounced [vela] is transcribed as *fel (yn)a*. The researchers have also transcribed a variety of lexical items which do not appear in standard Welsh reference dictionaries. These include dialectal terms like *jaman* (an expression of consternation) which is widely used in Caernarfon Welsh, as well as relative neologisms such as *cyfryngi* (“media person”) and playful ad hoc nonce words created by the participants, like *cyn rodidenaidd* (“pre-rhodonedric”). Such words are transcribed in an orthographic form agreed upon by the transcription team, and compiled in a list which is available in the documentation file (Deuchar 2014). The researchers also transcribed non-standard synthetic prepositions and verb endings using a similar method of consultation and compromise, and the agreed forms can also be found in the documentation file.

In terms of mutation, the researchers deliberately took an approach which embraces non-standard mutation, and in some respects makes this thesis possible. As this extract from the documentation file shows, variation between standard, non-standard and intermediate forms of mutation are recorded with high fidelity:

“We have represented mutation (sound change to initial consonants) or its absence without following prescriptive rules as to where mutation might or might not be expected. Thus the Welsh form of ‘in Cardiff’ may be transcribed *yn Caerdydd* (with an initial [k] on the place name) and *yn Gaerdydd* (with initial [g]), as well as the standard form *yng Nghaerdydd* (with initial [ŋ], according to what is heard (Deuchar 2014: 10))”.

The fact that the researchers transcribed the mutation patterns in audio recordings objectively in this way means that a corpus based analysis of mutation is facilitated, since tokens of

different mutation variables can be easily searched for and coded. This process will be described in detail starting from section (5.4) below.

5.1.2.2 Language Classification

Since the *Siarad* corpus was initially collected with the main aim of analysing code-switching, it should not be surprising that significant proportion of the researchers' efforts have been focused upon developing an effective system for classifying language, and for distinguishing code-switches from borrowings and Welsh words. As was mentioned in the previous chapter, classification of words operated from the dictionary criterion, which stipulates that 'listedness' or the degree of integration of a word into the speech community is the crucial factor for this classification process (Deuchar 2014). In the Welsh context, where there are no monolingual speakers who could help clarify the distinction between integrated and non-integrated words, the researchers have drawn upon dictionaries as an alternative method of establishing the extent to which words were integrated into Welsh. Words were considered Welsh if they were listed in any Welsh reference dictionaries, or in grammar reference books King (2003) or Thomas (1996). English words, conversely, were those that could only be found in the Oxford English Dictionary Online (2008). A final class of words were tagged as 'undetermined', for a variety of reasons including being listed in both Welsh and English dictionaries. Using this classification system, all words are tagged as belonging to one of these three categories:

“Welsh words are unmarked but English words are identified with the tag *@s:eng*. Words which could come from both Welsh and English are considered to be of 'undetermined' language and are marked *@s:cym&eng*, where *cym* represents Welsh and *eng* English (Deuchar 2014: 7)”.

An additional type of category is for mixed morpheme words – words that include one or more morphemes from both languages. These words appear particularly frequently with an English

verb root and a Welsh ending e.g. *concentrate-io@s:eng+cym* (to concentrate), but also can appear in other combinations of morphemes, which are reflected in the coding of the language tag.

5.1.2.3 Other Tiers

The translation tier is a free translation into English of the main tier, and uses the code (%eng). There are two gloss tiers available for each utterance. The manual gloss tier (%gls) was typed by hand by the original researchers when the data was originally transcribed. An automatic gloss tier (%aut), which has some expanded features such as the tagging of mutation, was added later. The comment tier (%com) was used when transcribers wished to draw readers' attention or explain any aspect of the data, including the meanings of colloquial expressions or the rationale behind technical transcription choices.

5.1.2.4 The Quality of Transcription

As has already been mentioned, the transcription team were trained transcribers with a thorough knowledge of linguistic principles. To check that transcription was of a high quality, around a quarter of the interviews had a randomly chosen minute of recording re-transcribed by an independent transcriber. This transcript was then compared with the original using plagiarism detection software. The average reliability score was calculated to be at 75%, implying that the general standard of transcription in the corpus is high (Deuchar 2014).

5.1.3 The make-up of the *Siarad* corpus

In the context of this study, the second function of the *Siarad* corpus is to serve as a representation of natural everyday conversation between Welsh-Speakers from a variety of backgrounds, ages and genders. In this section I will describe the variation in region, age, gender as well as other social factors amongst the *Siarad* participants.

5.1.3.1 Region

The majority of participants are from the North West of Wales, although participants from other parts of the country are present in the corpus, including the South East, South West and North East. The composition of the corpus largely reflects the headquarters of the original project at Bangor University, and the use of the friend of a friend (Milroy 1980) method in the data collection phase. The majority of the subjects are from areas proximal to Bangor, including the counties of Gwynedd and Anglesey. Both areas are notable for two reasons, the relatively unique high level of vitality of the Welsh speaking community and the North Welsh dialect (sometimes called *Venedotian* – Ball 1988) shared across both areas. This is one aspect that makes this study relatively unique, as most previous large scale quantitative mutation studies have (deliberately in some cases) looked at communities where the linguistic vitality of Welsh was comparatively lower, such as Cwmtawe (Ball 1984) and Rhosllanerchrugog (Jones 1998).

It's open to debate whether this corpus represents a single speech community in the way that many sociolinguistic studies strive to do. Previous studies have often stressed the diffuse nature of Welsh speaking areas, with different communities exhibiting diverse mutation patterns and possibly different linguistic norms. I believe that these differences have been somewhat over-emphasized in past studies. Nation-wide institutions, like the school system, Welsh government and media bodies (including the BBC and S4C) have now been influencing Welsh speakers for decades, so it can be argued that speakers from different regions do share many of the same attitudes and linguistic norms at this point in time. Also, the majority of speakers from other regions in the country were living in the Gwynedd area when they were recorded, meaning that they could be considered part of the speech community of North West Wales to some extent.

5.1.3.2 Age and Gender

With these demographic variables, the researchers attempted to achieve an even distribution when recruiting participants. Gender is split more or less evenly, with only slightly more women than men in the makeup of the corpus:

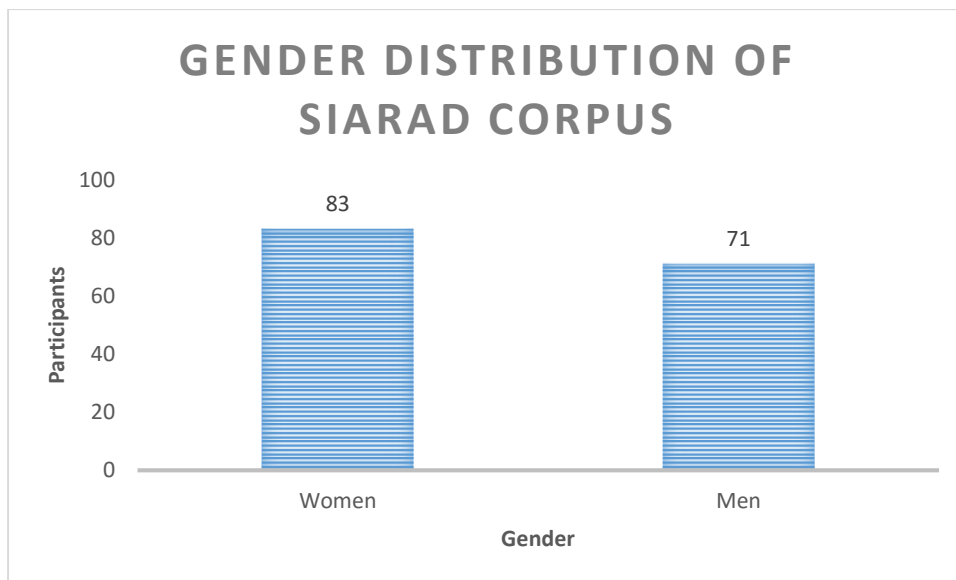


Figure 1 The distribution of gender among participants in the *Siarad* corpus

In terms of age, the researchers reported some problems in terms of collecting older interviewees in particular. This may have been partly due to the ‘friend of a friend’ method, with the relatively young researchers making connections more easily with younger participants. This may explain why the 20-29 age group has more participants than the 70+ group, for example.

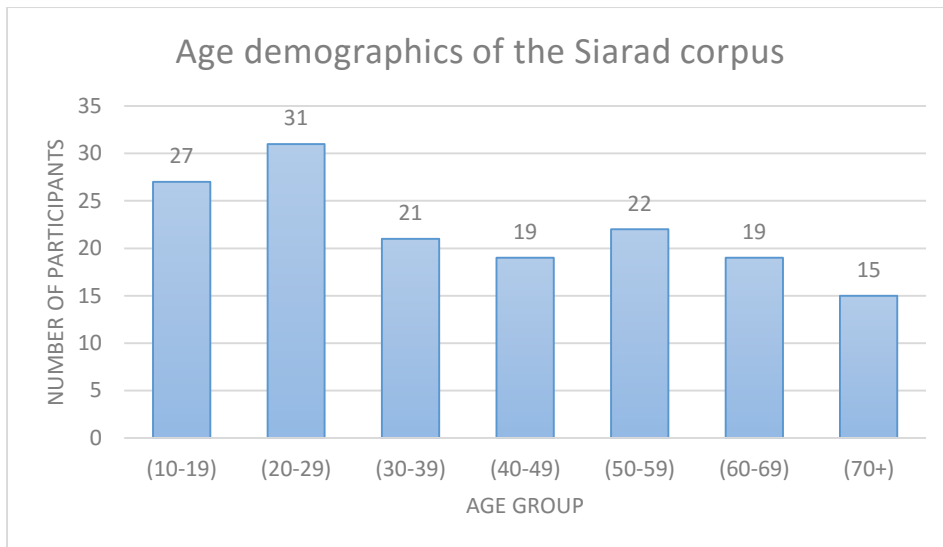


Figure 2 The age demographics of *Siarad* corpus participants

5.1.3.3 Level of Education

Information about participants' level of education was collected on a five point scale. A higher numerical value broadly represents a more advanced level of education. The options on the questionnaire were as follows:

- 1: None of the above
- 2: GCSE, O-Level/CSE, School Certificate, NVQ level 1 or 2 or equivalent
- 3: A/AS level, Higher School Certificate, GNVQ, BTEC National Diploma
- 4: Bachelor's Degree, Diploma of Higher/Further Education, PGCE, HND, NVQ level 4 or equivalent
- 5: Master's Degree, Doctorate, NVQ level five or equivalent

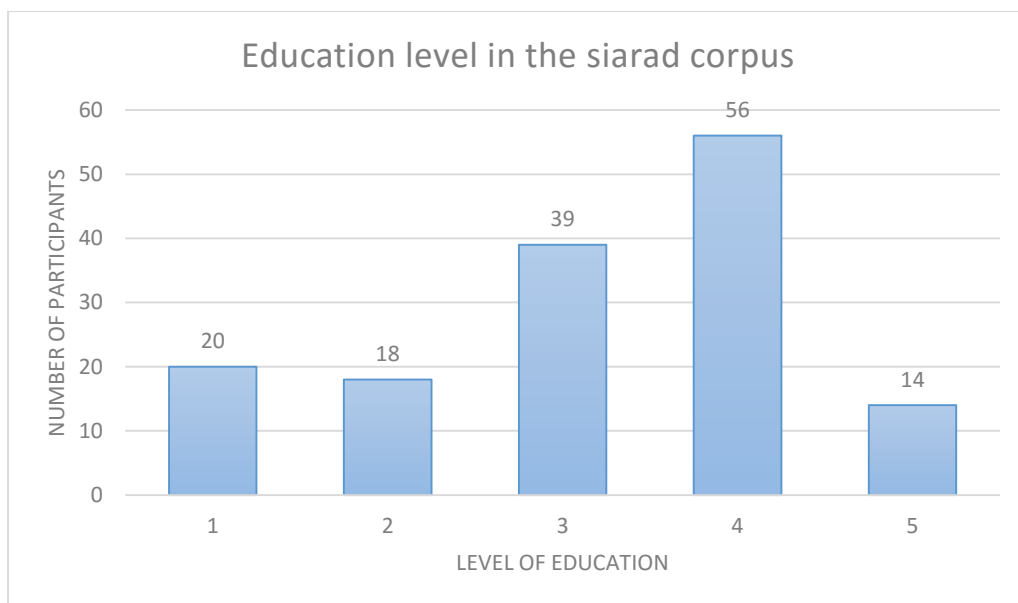


Figure 3 The number of *Siarad* participants by level of education

The demographics of the *Siarad* corpus clearly do skew towards participants who have received further education, which again could be partly due to the ‘friend of a friend’ approach and the traditional student bias of many University based research projects. Despite this, the researchers did manage to collect a broad range of participants, with Welsh speakers from every level of education represented.

5.2 Formality and the *Siarad* corpus

As has already been discussed, the notion of ‘natural speech’ is controversial among linguists. It can also be argued that true naturalistic speech is almost impossible to record without using unethical methods (e.g. recording without participants’ consent). In his interview methodology, Labov (1972) considered even the most informal part of the process atypical of participant’s natural speech, with genuine vernacular speech only glimpsed occasionally through unforeseen changes in the interview dynamic, such as phone calls or other intercessions by speakers familiar to the interviewee. In light of this phenomenon, known as the observers’ paradox, the *Siarad* corpus was deliberately collected with the intention of being as informal as possible,

particularly as the analysis of code-switching, which is associated with informal contexts, was the main objective of the researchers who collected it. One of the main factors conducive towards informality in the recordings is that the speakers were all recorded in conversation with people they knew well, and studies such as McNair & Knox (1994) have suggested that familiarity towards interlocutors is one of the main factors in modulating formality. Researchers also made sure that they left the area once recording was initiated, so that their presence would not be a factor in participant's behaviour. In theory, the recordings represent the natural behaviour of familiar dyads in natural conversation.

This is not to say that the corpus reflects completely natural speech. There are a few reasons why the language in the *Siarad* corpus cannot be considered completely naturalistic. Speakers were asked to speak together for an arbitrary period of time (thirty minutes). This may not be entirely conducive to natural conversation, since speakers might feel pressured or uncomfortable with the artificiality of the situation. This interpretation is borne out somewhat by meta-commentary by some speakers in the corpus data expressing nervousness, or an inability to think of further topics of conversation, along with some awkward silences. These kinds of comments only occur occasionally in the data, but it can be argued that this behaviour reflects a general tendency for participants to be more conscious of their speech in this dataset than they might be in their daily lives. If this is true then it can be conjectured that a global increase in attention to speech has probably filtered through to scores in the data, leading to more standard variants than is truly reflective of participants' completely unmonitored vernacular speech.

Other researchers have taken further measures to ensure the informality of the data, such as asking speakers to record themselves speaking to peers in their own time (e.g. Zentella 1990) although it might be argued that this method may have drawbacks of its own related to researcher control over the data. On the whole then, the corpus reflects an effective attempt to

systematically gather a very large corpus of informal speech in a relatively controlled way. Although I will have no illusions about the *Siarad* corpus representing the speech community's most unmonitored style, I feel justified in placing the corpus at the informal end of the radio corpora collected for this study, where participants generally monitor their speech more intensively.

5.3 Mutation variables

These variables comprise of seven different lexical items which, in their standard forms, trigger three different types of mutation, soft, aspirate and nasal. The variables are masculine possessive pronoun (ei.m), feminine possessive pronoun (ei.f), preposition (o), preposition (am), locative preposition (yn), conjunction (a) and possessive pronoun (fy). General information about these mutation triggers is provided in table 22.

Trigger	English Translation	Gloss	Type of mutation caused
(am)	about/for	for.PREP	soft
(ei.m)	his/its	his.ADJ.POSS.M.3S	soft
(ei.f)	her/its	her.ADJ.POSS.F.3S	aspirate
(o)	from	from.PREP	soft
(yn)	in	PRT/in.PREP	nasal
(a)	and	conjunction	aspirate
(fy)	my	my.ADJ.POSS.1S	nasal
(â/gyda)	with	with.PREP/as.CONJ	aspirate

Table 22 General information about the mutation variables selected for the study

These variables were selected from a large potential range of mutation triggers. As D. G. Jones (1988) mentions, there are around forty grammatical and lexical triggers governing mutation, although many of these are no longer used even in the most formal written registers. The variables selected were chosen primarily on the basis that they were the variables most commonly used in previous mutation studies (see chapter 3). I consider this an advantage because it should also allow for real time comparisons with past studies, which include Ball's (1984) Cwmtawe study and Jones' (1998) work in South and North East Wales.¹⁵ The variables are also intended to represent the three main types of mutation, with four causing soft mutation, two aspirate mutations and three nasal mutations. Additionally, they include a range in terms of trigger word type, with a mixture of prepositions, conjunctions and pronouns.

5.4 Finding mutation environments

The first step in analysing a corpus for mutation is finding and identifying environments where the desired variants can occur. For many variables this is a relatively simple task, since a lexical trigger usually precedes the initial consonant of the following word. Using the KWAL function of CLAN allows me to extract all instances of the desired mutation requirement. For the preposition (am) for example, the following command will extract all utterances containing the target variable from the transcript Davies1.cha:

```
KWAL +sam +f Davies1.cha
```

¹⁵ A disclaimer to this statement – both of the studies mentioned here were carried out in communities that differ in some significant respects from that represented in the *Siarad* corpus. These include differences in time (the nineties for Jones and the eighties for Ball), the status of Welsh in those communities (with lower vitality for Welsh in both studies compared to North Wales) dialect (previous studies have looked mainly at South Wales and the North East) and more. Any comparison would thus need to take into account these differences, and it should be seen as a broad exercise.

The +s indicates that the switch is activated to search for all instances of ‘am’ in a transcript. The +f indicates that an output file containing the data will be sent to another directory, where the data for all tokens of the variable (am) for each transcript can be kept together.

5.4.1 Dealing with dropped triggers

One problematic issue arising in identifying mutation environments is the fact that speakers in the *Siarad* corpus data have a strong tendency to drop the lexical trigger, particularly in the case of some variables. This is most common for the variables (ei.m), (ei.f) and (fy). This would seem to raise a methodological problem, particularly since at 40 hours the corpus is too large for all these variables to be tagged manually within the timeframe of this thesis.

As an example of the problem at hand, the variable (fy) seems to have multiple possible variants in the corpus, rendering it particularly difficult to capture all variations within a single search. The pre-nominal possessive pronoun (fy) can be dropped or retained by a speaker, for example ‘my cat’ could be realized as either:

fy nghath or *nghath*

Additionally, the first person possessive construction is also variably followed by an echoing pronoun *i* or *fi*, which is considered a recent informal variant by some grammarians (King 2003). This opens up more possible variants (I have represented dropped tokens of (fy) in the examples below by placing them between parentheses):

fy nghath / (fy) *nghath* / (fy) *nghath i* / *fy nghath i*

Finally, as with the other mutation variables, speakers can either mutate the initial consonant, or use the radical. Speakers often use different combinations of all these features, so that the variable (fy) might be considered in fact to have at least six variants rather than two, illustrated here in rough order of decreasing standardness:

fy nghath / (fy) nghath / fy nghath i / (fy) nghath i / fy cath / (fy) cath fi

These variants of the variable (fy) lack a single easily searchable constituent which can cover all tokens in a single search. Although I previously described an auto-glossed tier, which does indeed gloss the nasal mutation in words like ‘nghath’, this does not help in the issue of finding all other variants of the variable – for example tokens where the radical or a soft mutation is applied, and so on. A solution is apparent, however, in the fact that all utterances in the corpus are accompanied by a translation tier. Using the command +%teng, the translation tier is activated and makes it possible to extract all target environments containing English word *my* from the translation tier, thus allowing the process of finding and identifying tokens to be speeded up substantially. For (fy), the following CLAN command allows extraction of all main tiers which have ‘my’ in the translation tier:

```
KWAL +smy +t%eng +f robert9.cha
```

This works relatively well because the translation tier makes explicit any material which is dropped in speech but is understood by a listener, as demonstrated in the following extract from the transcript Fusser 8:

```
BRE: oe(dd)      thad  hi      (y)n      reolwr  bank@s:und yn Abersoch@s:und
```

```
%gls: be.3S.IMP  father PRON.3SF  PRT  manager bank      in Abersoch
```

```
%eng: her father was a bank manager in Abersoch
```

Extract 12 An utterance from the *Siarad* corpus where feminine pronoun (ei.f) has been dropped

The speaker above has dropped the aspirate mutation trigger (ei.f) before the word ‘tad’, but has still mutated the target word standardly as ‘thad’. The transcriber has understood the presence of the dropped trigger from the contextual information of the utterance (presumably

including the aspirate mutation of *tad*), and has included the English equivalent ‘her’ in the translation tier. When searching for the variables (ei.m) and (ei.f) the following command is used, which incorporates both the lexical trigger *ei* and possible semantic equivalents for it in English:

KWAL +shis +sher +sits +sei +t%eng +f davies10.cha

The search command used thus varies depending on the requirements of each variable. Table 23 shows which variables are dropped in this way in the *Siarad* data, as well as their variants and the search commands used to find them.

Mutation variable	Variants	Search commands used
(fy)	fy nghath / (fy) nghath / fy nghath i / (fy) nghath i / fy cath / (fy) cath fi	KWAL +smy +t%eng +f file.cha
(ei.m)	ei gath / (ei) gath fo - (ei) gath fe / (ei) gath o – ei gath e / cath fo - cath fe /	KWAL +shis +sits +sei +t%eng +f file.cha
(ei.f)	ei chath / (ei) chath hi / (ei) chath / cath hi	KWAL +sher +sits +sei +t%eng +f file.cha

Table 23 Constructions with mutation triggers which are variably dropped

Once the variables can be identified routinely in the corpus data, the next step is to classify the mutation data.

5.5 Classifying mutations

The process of classifying mutations involves identifying each token as belonging to one of four categories; standard, intermediate SM, radical or undetermined. For binary variables, which have two variants, I define each possible token as follows:

Standard: where the target word is mutated as expected according to Welsh standard grammar

Radical: where the target word is not mutated, but retains the radical form

Undetermined: where the mutation environment was ambiguous and could not be confidently classified as belonging to either of the above categories

For variables which have more than two variants, there is an additional category:

Intermediate SM: where soft mutation is applied to a target word which is expected to undergo aspirate or nasal mutation

Once all tokens of the target variable have been extracted with KWAL, they are organized into files and methodically counted and entered into an Excel spreadsheet in order of transcript number. As I placed each utterance in the spreadsheet, each token was examined to identify the type of variant.

5.5.1 Classifying binary variables

For example, in the following example from the transcript Robert 3, speaker LUN uses the standard soft mutation following the variable (o):

LUN: ah@s:und dw i (ddi)m yn rhoi llwyth o° **gardiau** tro (y)ma

%gls: IM be.1S.PRES PRON.1S NEG PRT give.NONFIN load of cards turn here

%eng: *ah, I'm not giving loads of cards this time*

Extract 13 An utterance showing a *Siarad* corpus speaker mutating standardly after (o)

Extract 13 is thus coded as a standard token for mutation for the variable (o). In the same transcript speaker LUN uses the radical form following the same variable trigger:

LUN: oh@s:und oedd o efo bunch@s:und o° **papurau** (y)n bob man neu (ry)wbeth
yeah@s:und

%gls: IM be.3S.IMP PRON.3SM with bunch of papers in every place or something yeah

%eng: *oh he had a bunch of papers everywhere or something, yeah?*

Extract 14 A *Siarad* corpus speaker using the non-standard radical after (o)

Since (o) is a soft mutation trigger, the initial consonant of the target word *papurau* would be mutated from [p] to [b] according to standard grammar. Since the radical form has been retained and no mutation occurred in this case, this token is coded as radical.

5.5.2 Classifying intermediate variants for index scored variables

For the aspirate and nasal mutation variables, I have included an additional variant *intermediate SM*, where speakers apply a soft mutation instead of the standard or radical variant. For instance, many researchers have noted that the variable (yn) seems to be moving from a binary radical/NM paradigm towards a three variant radical/NM/SM model (e.g. Awbery 1986). In order to provide a fully representative a model of this variable in his analysis, Ball (1988) used an index scoring system which provided a three level scoring hierarchy for (yn). Ball also applied an indexical scoring system to the variable (ei.f) in order to include the possibility of participants non-standardly applying SM where AM is traditionally expected. Although there are intermediate SM tokens for (a) and (â/gyda) in my data, they are infrequent enough that I decided against setting them up as index scored variables, which was also the approach taken by Ball (1984). Table 24 presents each variable along with all of its variants.

Trigger/Variable	Variants	Number of variants
(am)	SM, radical	2
(ei.m)	SM, radical	2
(ei.f)	AM, SM, radical	3
(o)	SM, radical	2
(yn)	NM, SM, radical	3
(a)	AM, radical	2
(â/gyda)	AM, radical	2
(fy)	NM, radical	2

Table 24 the number and types of mutation that each variable undergoes

For (yn) there are thus three possible ways that a token could be coded. The extract below provides an example of participant ANN using a standard variant of (yn):

ANN: oedden ni (y)n cychwyn yn° **Meddgelert@s:und** (.) a cerdded efo (.) yr afon ochr bella i (y)r afon

%gls: be.1PL.IMP PRON.1PL PRT start.NONFIN in Beddgelert and walk.NONFIN with DET river side farthest to DET river

%eng: *we started in Beddgelert and walked with the river, the other side of the river*

Extract 15 A *Siarad* participant using the standard variant of (yn)

ANN has applied a nasal mutation to target word Beddgelert, mutating the initial consonant [b] to the nasal [m]. As this is the standard mutation for this environment, the token is coded as standard. In the next extract ANN uses the non-standard radical variant:

*ANN: oedd y jug@s:und (y)na adre (y)n° tŷ mam@s:und ers blynyddoedd

%gls: be.3S.IMP DET jug there home in house mum since years

%eng: *that jug was at home at my mother's for years*

Extract 16 A *Siarad* participant using the non-standard variant of (yn)

Since ANN has retained the radical [t] in the target word *tŷ*, this token is coded as a radical token of (yn). Finally, below is an example of the intermediate SM mutation following (yn):

*ANN: ond yn° **Griccieth@s:und** ges i (y)r (.) potyn

%gls: but in Criccieth get.1S.PAST PRON.1S DET pot

%eng: *but I got the pot in Criccieth*

Extract 17 A *Siarad* participant using the intermediate variant of (yn)

Here ANN has applied a soft mutation to the place name Criccieth, mutating the initial consonant from [c] to [g] rather than using the standard nasal mutation [ŋ].

5.5.3 Classifying tokens for (ei.f) and (ei.m)

Identifying variants for these variables is slightly more complex than for other variables, since the presence of the trigger itself is not enough to disambiguate the mutation context. Both (ei.f) and (ei.m) are phonetically identical, and as I have already discussed, the trigger itself is in any case often dropped by speakers in informal speech. This means these variables must be disambiguated by finding the antecedent possessor that the pronoun refers to. If the possessor's gender is feminine, then the token is coded as a variant of (ei.f), but if the possessor is masculine it is coded as a variant of (ei.m).

TOS: wedyn er@s:und (.) John_Ifan@s:und a (e)i° **frawd** yn idiots@s:und

%gls: then IM John_Ifan and POSS.3SM brother PRT idiots

%eng: *then, er, John Ifan and his brother are idiots*

Extract 18 A *Siarad* participant using (ei.m) standardly

In extract 18 the possessor for *ei* is *John Ifan*, which is a masculine personal name, making this (ei.m). As TOS has applied the standard soft mutation to the target word *brawd* this is coded as a standard token of (ei.m). In the extract below BEN uses (ei.m) non-standardly.

*BEN: +, yn (.) dim ond lle (.) mae (.) llyn wedi cael ei° **llenwi** +/.

%gls: PRT NEG but where be.3S.PRES lake PRT.PAST get.NONFIN POSS.3S fill.NONFIN

%eng: *only where the lake has been filled*

Extract 19: *Siarad* participant using a non-standard variant of (ei.m)

The possessor for *ei* here is the word *llyn* (lake), which according to Welsh reference dictionaries is a masculine noun. Since BEN has retained the radical consonant of the target word, this is coded as a radical variant of (ei.m).

As an aspirate mutation variable with three possible variants, coding (ei.f) is slightly more complex than (ei.m), but relies on the same procedure of accurately identifying the possessor in each case.

AWE: a oedd un iâr yn (.) gwthio (e)i° **phen** trwy twll doedd (.) i ddeud hello@s:und (wr)tha
chdi

%gls: and be.3S.IMP one chicken PRT push.NONFIN POSS.3SF head through hole
be.3S.IMP to say.NONFIN hello to.2S PRON.2S

%eng: *and one chicken was pushing her head through the hole, wasn't she, to say hello to you*

Extract 20 *Siarad* participant using a standard variant of (ei.f)

In extract 20 *iar* (chicken) is the antecedent possessor. As *iar* is a feminine noun, I consider this a token of (ei.f). AWE has applied the standard aspirate mutation to the target word *pen*, and so the token is coded as a standard variant for (ei.f).

*GLA: um@s:und (.) mae [ʃ] (.) mae Jenny@s:und dim ond <yn ei> [ʃ] (.) um@s:und yn ei°

pedwardegau

%gls: IM be.3S.PRES Jenny NEG but in POSS.3S IM in POSS.3S forties

%eng: *um, Jenny's only in her...um, in her forties*

Extract 21 *Siarad* participant using a non-standard variant of (ei.f)

In extract 21, Jenny is the antecedent possessor. Jenny is a female personal name, making this a token of (ei.f). The target word *pedwardegau* (“forties”) has retained the radical consonant [p] here, making this a radical token of (ei.f).

The final variant for (ei.f) is intermediate SM, which is where soft mutation is applied to the target word, rather than the standard aspirate mutation:

*IOL: (be)cause@s [///] ac ydw i cofio efo &deiva dafad ar yn nwylo ac rwbio (y)r (..) saim ar y dafad a ei° **gladdu** o

%gls: because and be.1S.PRES PRON.1S remember.NONFIN with wart on POSS.1S hands and rub.NONFIN DET grease on DET wart and POSS.3SM

%eng: *because... and I remember having a wart on my hands, and rubbing the grease onto the wart, and burying it*

Extract 22 *Siarad* participant using a non-standard variant of (ei.f)

In extract 22 the possessor for *ei* is *dafad* (wart), which is a feminine noun. Since the speaker IOL has applied a soft mutation to the target word *claddu* rather than the standard aspirate mutation, this is coded as an intermediate SM token of (ei.f).

Possessors are not always clearly visible in the same utterance as the pronoun, and much of the challenge in coding (ei.m) and (ei.f) has been in attempting to identify possessors which are not explicitly stated. This issue is discussed further in section (5.8.6).

5.5.4 Classifying undetermined variables

The final classification, ‘undetermined’ refers to tokens which were originally coded as a mutation variants, but were ultimately removed from the study for accountability reasons. A comprehensive exclusion list for tokens which belong to this category can be found below in section (5.7).

5.6 Calculating scores for mutation variables

Part of the rationale for performing an analysis of the *Siarad* corpus is that it will allow for a comparison between naturalistic speech and the arguably performative, monitored, speech of the *Radio Cymru* corpus. This section will set out the method used to calculate mean scores for the use of each mutation variable. Scores will be presented in two ways – as percentage scores

for those variables which have only two variants, and as index scores for variables with more than two variants. An important problem confronting any analysis of mutation is that of the infrequency of certain variables. This is a greater issue in the *Siarad* corpus than in the Radio corpus, because each recording in the *Siarad* corpus is only 30 minutes in length. Even the most frequently occurring mutation trigger (o) is rare in the speech of some individuals, and some variables such as (â/gyda) are very infrequent, meaning that individual scores cannot be calculated. Tokens are thus added together to create age group scores, rather than averages of individual percentage and index scores for every speaker. The procedure for calculating group percentage and index scores is explained below.

5.6.1 Calculating percentage scores for binary variables

Binary variables are those that have only two possible variants, standard and radical. To begin, every instance of both variants is counted for each speaker. For instance, the *Siarad* participant LUN has seven standard tokens of the trigger variable (o), and one radical token of the same variable. LUN's standard tokens are added to those of the groups of which she is a member. For example, LUN is a member of the (10-19) age group, and her tokens are added towards the sum of this group's tokens, which add up to a 122 standard tokens and 56 radical tokens in total. The percentage score representing this age group's frequency of standard mutation after (o) is thus calculated as a simple proportion:

$$122 / (56 + 122) = 68.54\%$$

This score can then be considered representative of this age group's tendency to mutate standardly following (o). Scores for other social variables, such as age and gender, can be calculated in the same way.

5.6.2 Calculating index scores for non-binary variables

Calculating the index score for preposition (yn) follows the methodology used by Ball (1984:76). Each variant is given a value which corresponds to its level of standardness. For preposition (yn), the variants and their values are as follows:

- 3 Nasal mutation (standard)
- 2 Soft mutation (intermediate SM)
- 1 Radical (non-standard)

This system reflects the hierarchy of standardness that Ball applied to the variants – nasal mutation is considered the most standard variant, while using the soft mutation is slightly less standard. The least standard variant is using no mutation at all, the radical. The other variable using index scores, (ei.f) has a slightly adjusted hierarchy. In Ball’s study, AM is considered the most standard score with the radical coming second, while using the soft mutation is ranked least standard. This is more controversial than the hierarchy for (yn), since the decision to rank the use of soft mutation as least standard appears almost seems a punishment for a poor grasp of the gender system in Welsh. Conversely, I would argue that speakers who inadvertently apply the ‘wrong’ mutation (based on an error in identifying the gender of the possessor) are approximating the standard more closely than those who simply use the radical, simply because they arguably have greater awareness that a mutation rule exists at all for the relevant environment. Thus my revised hierarchy for (ei.f) places the SM variant in intermediate position, and the radical variant as the least standard.

- 3 Aspirate mutation (standard)
- 2 Soft mutation (intermediate SM)
- 1 Radical (non-standard)

I will now outline the procedure for coding tokens and calculating group scores in more detail. Each individual token used by a speaker is first identified and coded according to the categorization outlined above. For example, *Siarad* participant MAG mutates the target word *pa* (which) standardly as *mha* following (yn) in the utterance below:

MAG: ym° **mha** bryd ?

%gls: in which time

%eng: *when?*

Extract 23 Participant MAG mutating standardly following the nasal mutation trigger (yn)

This token is thus coded with a value of 1. In the utterance below, the same speaker mutates the target word *Caernarfon* using the soft mutation to produce *Gaernarfon* following (yn), which (following Ball) I have coded as the intermediate SM form:

MAG: diwrnod blaen yn° **Gaernarfon@s:und**

%gls: day front in Caernarfon

%eng: *the other day in Caernarfon.*

Extract 24 Participant MAG using the intermediate mutation following the nasal mutation trigger (yn)

This token is coded with a value of 2. In yet another extract, MAG retains the non-standard radical form in the target word *tŷ* (house) following (yn):

MAG: yn° tŷ Graham@s:und mae (y)n gynnes nice@s:und efo (y)r tân yn y living@s room@s:und

%gls: in house Graham be.3S.PRES PRT warm nice with DET fire in DET living room

%eng: *in Graham's house it's nice and warm with the fire in the living room.*

Extract 25 Participant MAG using non-standard mutation following the nasal mutation trigger (yn)

As the radical is placed at the bottom of the scoring hierarchy for (yn), this token is given a value of 1. As with the binary mutation scores discussed in section (5.6.1), index scores are calculated exclusively as group scores, because individual frequency for most mutation variables was too low to calculate individual scores. For this reason, MAG's tokens are contributed towards group scores. For example, MAG is a member of the 20-29 age group. MAG's tokens are added to the group's total which add up to 22 standard tokens, 12 intermediate tokens, and 42 non-standard tokens. To calculate group scores for each age group, the value of each variant is then multiplied by the number of times it is used in that group and divided by the total number of instances of the variable - (yn) in this case - by the group. Finally, the score is given a numerical value out of a range of 0 to 200 by subtracting 1 and multiplying by 100. For the (20-29) age group this is calculated in the following manner:

Step 1: Each token multiplied by its value: $22*3 = 66$ $12*2=24$ $42*1=42$

Step 2: The total number of tokens in the group is calculated: $22+12+42= 76$

Step 3: The combined sum of step one is then divided by the result of step two:

$$(66+24+42) / 76 = 1.736842105$$

Step 4: The result of step three is subtracted by 1 and then multiplied by a hundred to produce the index score:

$$(1.736842105 - 1) * 100 = 74$$

The final index score for the 20-29 age group's use of the variable (yn) is thus 74. This score can then be used to compare this age groups' tendency to use (yn) standardly with others. The next section will move on from calculating scores to discuss the process of identifying 'undetermined' tokens from the analysis.

5.7 Identifying undetermined tokens

Besides tokens which are identified as standard, non-standard or intermediate, certain tokens can be categorized as undetermined. Undetermined tokens are ones which at first glance might seem to involve mutation, but for various reasons are ambiguous in their actual status. The next section will provide a brief overview of situations where tokens are considered undetermined. In the process of identifying the status of each token, undetermined tokens are tagged with the code 'und' in the spreadsheet, with a brief explanation of the reason for the token's exclusion from further analysis.

5.7.1 Ambiguous Language classification

Stammers' (2010) thesis has provided evidence that Welsh speakers may be less likely to mutate lexical items which are perceived as switches rather than integrated borrowings. The likelihood of speakers applying mutation to words thus seems to be related, to some extent, to the words' integration into Welsh, which in Stammers' thesis, as in this one, was represented by their inclusion or absence in Welsh dictionaries. If speakers' likelihood of mutating borrowings and switches is indeed constrained by these kinds of internal factors, these items need to be controlled in the analysis. There are several examples of both English and ambiguous words being mutated standardly in the *Siarad* corpus, although the latter seems to be more common than the former. In the examples below, the plural noun *drinks*, which is tagged as an English word, and the plural noun *tiles*, which is tagged as an undetermined word (i.e. present

in both Welsh and English reference dictionaries), are both mutated standardly following the trigger variable (o):

NON: couple@s:und o° **ddrinks@s:en** ar_ôl practice@s:und wedyn wastad (gy)da ni

%gls: couple of drinks after practice then always with PRON.1PL

%eng: we always had a few drinks after practice then

Extract 26 Standard mutation applied to word tagged as English from Davies 1

AET: wchi ryw fath o° **diles@s:und**

%gls: know.2PL some kind of tiles

%eng: you know, some kind of tiles

Extract 27 Standard mutation applied to word tagged as undetermined from Fusser 17

As was discussed previously, the fact that many words tagged as ambiguous or English are sometimes mutated by speakers may indicate that some words are at variable stages of integration into Welsh among the Welsh speaking community. However, because different speakers may have differing perceptions of the relative integration of such words into Welsh, including them in the analysis would likely skew the results. In order to resolve this problem, all tokens including words which are tagged as either undetermined (@s:und) or English (@s:en) are removed from the study, so that the analysis can concentrate on those lexical items which may be most confidently considered ‘Welsh’.

5.7.2 Words that are never mutated

There is a category of Welsh words that are never mutated even when in environments which typically trigger mutation in other words. A list of these words is drawn from Ball and Müller (1992) and is displayed below. These words are noted and then excluded from any further analysis.

- tua (*about/towards*)
- mor (*so*)
- braf (*lovely*)
- Non-Welsh place names (e.g *Canada, Timbuktu*)
- Certain Welsh [g] initial borrowings (e.g. *gemau, graff*)
- Personal names (e.g *Ceridwen, Rhys*)

5.7.3 'Fossilized' formulaic constructions

A.E. Jones' (1988) observation that the more obscure aspirate mutation conjunction (a) is mainly heard in 'fossilised' constructions such as *ceffyl a° chert* has relevance for our analysis. According to Jones, speakers in this situation are not necessarily mutating the consonant according to standard grammar, but are simply repeating the formulaic or idiomatic phrase as a whole, which is itself a holdover from an earlier linguistic period. As these cases are not equivalent to the normal process of mutation, including them in the analysis could arguably adversely affect the resulting data by over-representing standard tokens of mutation. These fossilized constructions were added to the exclusion list as I collected the data. As can be seen in table 25 below, the main variable affected is (o), which appears in many formulaic constructions which have an adverbial quality, but the variables (am) (fy) and (yn) also appeared in similar forms.

Although many of these constructions were quite frequent, there were no examples in the data where the adjacent target word was un-mutated. For instance, no speaker ever retained the radical *te* ("tea") in the construction 'paned o° **te**' ("a cup of tea"), but always categorically mutated the target word, as in: 'paned o° **de**', and no speaker ever retained the radical *lleiaf* ("least") in the construction 'o° **lleiaf**' (*at least*), but always used 'o° **leiaf**', and so on. The main criterion for inclusion in this group was that the constructions exclusively involve categorical mutation – either always mutated or never mutated.

The majority of these constructions can be found in standard grammars of Welsh. A notable exception is *am bod*, which seems to be a frequently occurring dialect phrase meaning ‘because’, but which has not been described in standard grammars. When used in this sense, *am bod* is distinct from the more traditional sense of ‘will be’ where *bod* is an active verb, and (am) is present as an active mutation trigger. Here are two examples of ‘am bod’ being used in this way:

NEL: mae o nabod nhw right@s:und dda **am bod** mae (we)di bod yn gweithio efo nhw ar y prosiectau

%gls: be.3S.PRES PRON.3SM know.NONFIN PRON.3PL right good for be.NONFIN be.3S.PRES PRT.PAST be.NONFIN PRT work.NONFIN with PRON.3PL on DET projects

%eng: *he knows them quite well because he's been working with them on the projects*

Extract 28 ‘am bod’ functioning as a lexical item meaning ‘because’

NEL: yeah@s:und **am bod** o'n i isio gorffen cyfweiliadau fi erbyn mis yma

%gls: yeah for be.NONFIN be.1S.IMP PRON.1S want finish.NONFIN interviews PRON.1S by month here

%eng: *yeah, because I wanted to finish my interviews this month*

Extract 29 A second example of ‘am bod’ functioning as a lexical item meaning ‘because’

The fact that ‘bod’ immediately precedes the verbs ‘mae’ and ‘o’n’ in these examples seems to imply that it is not itself acting as a verb. This is arguably because ‘am bod’ in these cases has become a lexical item, which involves neither on-line mutation nor a working verb. This construction should be distinguished from ‘am° **fod**’ where the verb is active, and the intended meaning is ‘will be’ or ‘to be’, for example:

OSW: fyddi di (y)n hiraethu am° **fod** adre Cynddylan@s:und ?

%gls: be.2S.FUT PRON.2S PRT long.NONFIN for be.NONFIN home Cynddylan

%eng: *do you miss home, Cynddylan?*

Extract 30 ‘am fod’ when used to mean ‘to be’

The former type of *am bod* is typically not mutated, while the second is. All cases of *am bod* which are followed by a verb will thus be tagged as undetermined while all other cases will be retained in the analysis. Another fossilized lexical item worth discussing is *fath â*, which is a very frequent construction usually meaning either ‘like’ or ‘as’. Although standard grammars list ‘as’ as a legitimate sense for the trigger for aspirate mutation following *â*, there is evidence that *fath â* as a whole often acts as a lexical item, rather than a phrase with an active mutation trigger. Younger speakers in particular seem to use *fath â* in a sense which calques the widespread vernacular ‘BE + like’ quotative construction in English (Tagliamonte and Hudson 1999), as can be seen in extract 31:

TOS: a pan wnes i weld xxx rei (y)ma yn (.) chwech_deg_tri punt (.) xxx o'n i **fath â**
“wow@s:und” (.) methu coelio (y)r peth xxx

%gls: and when do.1S.PAST PRON.1S see.NONFIN xxx some here PRT sixty_three.M
pound xxx be.1S.IMP PRON.1S kind with wow fail.NONFIN believe.NONFIN DET thing
xxx

%eng: *and when I saw [...] these ones for sixty-three pounds [...] I was like "wow" couldn't believe it*

Extract 31 A *Siarad* participant using *fath â* as a calque of ‘BE + like’

Another piece of evidence which supports the lexical nature of *fath â* is the fact that it is often rendered in writing (although particularly in colloquial contexts) as a single lexical item: *fatha*.

Lastly, there are no cases in the *Siarad* or *Radio Cymru* corpora where *fath â* triggers an aspirate mutation in the adjacent word (although standard tokens following (*â/gyda*) were extremely infrequent in any case). Table 25 lists all of the constructions which were considered to be ‘fossilized’ in the sense described above. All tokens involving these constructions were removed from the analysis:

Fossilised construction	English Translation	Mutation variable affected
o° gwbl	At all	(o)
o° ran	In terms of	(o)
siŵr o° fod	probably	(o)
ADVB + o° dda	ADVB (e.g. terribly) good	(o)
paned o° de	a cup of tea	(o)
o° ddifri	serious	(o)
o° bwys	important	(o)
o° lawer	significantly	(o)
o° leiaf	at least	(o)
o° blaid	in favour	(o)
o° dro i dro	from time to time	(o)
byd o° les	very worthwhile	(o)
am° bod ¹⁶	because	(am)

¹⁶ Note that ‘am bod’ and ‘fath â°’ contrast with the other examples, in that they involve the mutation trigger becoming inactive, whereas the other cases seem to involve mutation being applied categorically.

am° dro	for a walk	(am)
brawd/chwaer yng° nghyfraith	brother/sister in law	(yn)
wrth fy° modd	overjoyed	(fy)
o (e)i° le	wrong	(ei.m)/(ei.f)
fath â°	like	(â/gyda)

Table 25 A list of formulaic constructions with ‘fossilized’ mutation triggers

5.7.4 ‘Relexified’ dialectal variants

A further source of confusion for the analysis of mutation involves words which have dialectal variants. In the corpus, many words can be found which have effectively shifted their initial consonant from unvoiced to a voiced form as the root form in the vernacular. P.W. Thomas (1984) referred to this process as *relaxification*, and provides as examples *bobol* (from standard *pobl*), *gegin* (from standard *cegin*), *ford* (from southern dialect *bord*), *fenyw* (from southern dialect *menyw*) and *bob* (from standard *pob*). In the *Siarad* corpus a similar but distinct example of this process can be found in a set of Northern dialect words. For example, the word *rhywbeth* (eng. *something*), has the initial consonant [f] in standard Welsh, but is typically pronounced with an initial [r] (which could be rendered orthographically as either *rywbeth*, or even more dialectically as *rw bath*) in the vernacular, even where there is no mutation trigger. Similarly the standard lexical form of the verb *gwneud* is most often rendered as *wneud* in the Northern dialect, and its inflected forms also usually follow this pattern. The shift from a root form like *pob* (eng. *every*), towards *bob* in colloquial Welsh can be particularly confounding for the analysis of nasal mutation trigger preposition (yn), because when speakers follow (yn) with *bob*, it is not clear whether they are applying a soft mutation intentionally or whether they

simply consider *bob* the root form. Furthermore, there is one example in the *Radio Cymru* corpus where a speaker seems to apply soft mutation to *bob*, producing the novel form ‘*fob*’:

GE1: er@s:und mae yna enghreifftiau di_ri o° **fob** math o fusnesau

%gls: IM be.3S.PRES there examples countless of every type of businesses

%eng: *there are many examples of all kinds of businesses*

Extract 32 a participant in the *Radio Cymru* corpus mutating *bob* as ‘*fob*’.

This provides further evidence that some speakers consider *bob*, rather than *pob*, to be the root form. However, since both *pob*, *mhob* and *phob* are also very frequent, and therefore very useful in the analysis of the nasal and aspirate mutations, I have decided to only remove *bob* from the analysis, and to retain the other variants.

The final list of relexified dialectal variants which were categorized as undetermined, and removed from the study, is presented in table 26:

Standard form	Dialect variant	English translation
rhywbeth	rhywbeth/rwbath	<i>something</i>
rhyw	ryw	<i>some</i>
pob	bob	<i>every</i>
potel	botel	<i>bottle</i>
tipyn	dipyn	<i>a little</i>
gwneud	wneud	<i>do</i>
gwnaeth	wnaeth	<i>did</i>

Table 26 A list of words which have relexified dialectal variants

5.7.5 Dropped items which block mutation

Another potentially confounding phenomenon related to this is the dropping of lexical items which are understood by the speaker, and which block mutation. A frequent example in the data is the dropping of the determiner *y*, *yr* and *'r* before nouns, which D. G. Jones (1988) describes as characteristic of Northern dialects. For example:

DAN: raid ti fynd â dy gar@s:und o° (**'r**) tŷ (a)cw

%gls: necessity PRON.2S go.NONFIN with POSS.2S car from (DET) house there today also

%eng: *you need to take your car from our house today, too*

Extract 33: An example of a dropped determiner from *Siarad* transcript Davies 6

Extract 33 could be inadvertently categorized as an example of a non-standard, radical token of (o), with the target word *tŷ* retaining its initial consonant [t]. However, a closer look reveals that the speaker has probably dropped either a determiner or a possessive pronoun following the trigger word. If the speaker has indeed dropped a determiner, there is no mutation environment in the utterance, while if a pronoun was dropped there could be a different mutation environment involved, for instance the AM trigger (ei.f). Ultimately, as it is impossible to tell from the context of the utterance what the dropped element is, the token must be removed from the count and tagged as undetermined.

5.7.6 Unknown antecedent possessors for (ei.m) and (ei.f)

When classifying these variables, identifying the gender of the possessor is a crucial step in recognizing whether the variable is (ei.m) or (ei.f), as both are phonetically identical. This problem is particularly apparent in the natural speech of the *Siarad* corpus, where participants often discuss possessors that are understood by the interlocutors but are only partially, or never, mentioned directly. This process of ellipsis is symptomatic of the informal register, and can be

contrasted with the elaborated registers (as opposed to economical registers – see section 2.6) in the more formal radio programmes which will be discussed in chapter 7, where possessors are often explicitly stated in the same utterance as the possessive pronoun. For example, in the extract below which is from the highly formal radio programme *Post Prynawn*, both the possessor *cystadleuaeth* and the pronoun *ei* are clearly stated in the same utterance:

*DEW: well@s:und mae un o atyniadau ymwelwyr Ynys_Môn wedi ei enwebu fel un o (y)r amgueddfeydd gorau mewn **cystadleuaeth** (.) sy (y)n cael ei° **chynnwl** gan bapur newydd y Guardian@s:und

%gls: IM be.3S.PRES one of attractions visitors Anglesey PRT.PAST PRON.3SM nominate.NONFIN like one of DET museums best in competition be.PRES.REL PRT get.NONFIN PRON.3SF maintain.NONFIN by paper new DET Guardian

%eng: *well, one of Anglesey's visitor attractions has been nominated in a competition that is being held by the Guardian newspaper*

Extract 34 An example of a possessor of (ei.f.) being stated explicitly in the highly formal programme *Post Prynawn*

Cystadleuaeth is feminine noun according to Welsh reference dictionaries, so this token is coded as a standard variant of (ei.f). By contrast, here is an utterance extracted from the *Siarad* corpus where the possessor is not made explicit:

*AET: <a mae (y)na> [//] (.) well@s:und mae (y)r trydydd yn canol +/-

%gls: and be.3S.PRES there IM be.3S.PRES DET third in middle

%eng: *and there's...well the third in the middle*

*BEN: yeah@s:und .

%gls: yeah

%eng: *yeah*

*AET: +, (we)di cael ei° **dorri** lawr am ryw reswm .

%gls: PRT.PAST get.NONFIN POSS.3SM¹⁷ cut.NONFIN down for some reason

%eng: *was cut down for some reason*

Extract 35 An example of a possessor of (ei.f.) not stated explicitly in the informal *Siarad* corpus

In extract 35, it is neither clear in the utterance itself nor in the preceding set of utterances what the possessor for the pronoun *ei* actually is. To find more information I had to search previous parts of the transcript to find some clue about the subject conversation. What I found can be seen in extract 36 below:

*AET: mae (y)na ynys bach a (y)r ddwy goeden (.) yn tyfu .

%gls: be.3S.PRES there island small and DET two.F tree PRT grow.NONFIN

%eng: *there's a small island and the two trees growing*

*BEN: +< mmm@s:und .

%gls: IM

%eng: *mm*

*AET: a mi oedd (y)na drydydd (.) yn canol .

%gls: and PRT be.3S.IMP there third in middle

¹⁷ This is the original gloss from the *Siarad* transcript of Fusser 17, which indicates that the transcriber has assumed that this is (ei.m) from the soft mutation applied to the target word. This illustrates the difficulty of applying this kind of analysis, where anaphoric reference needs to be taken into account.

%eng: *and there was a third one in the middle*

Extract 36 An extract showing the possessor being referred to in an earlier part of the transcript

Extract 36 indicates that the possessor is the noun *coeden* (tree), and reading the transcript more fully reveals that the participants were discussing trees being planted and cut down around Bangor. The token from extract 35 can thus be identified as an intermediate SM variant of the variable (ei.f), as *coeden* is a feminine noun, and the speaker has applied a soft mutation to the target word *torri*. Although in this case I was able to identify the possessor with some degree of confidence, there are many other examples in the *Siarad* corpus where the nature of the possessor could not be disambiguated satisfactorily. This difficulty seems to be an unavoidable aspect of performing this kind of analysis on naturalistic data, but the problem can be minimized by tagging any tokens where the possessor is not clear as undetermined.

5.7.7 Unclear Audio

Although the *Siarad* corpus researchers were highly skilled and effective transcribers, and the audio quality of the corpus is generally clear, there are some occasions where I have found myself either disagreeing with the transcribers' transcription of a word, or failing to distinguish the initial consonant with sufficient clarity to make a reliable judgment. Therefore, any tokens where I feel that the audio is not sufficiently clear to make a confident judgement have been removed from the study. Additionally, wherever I simply disagree with the original transcriber about the type of mutation that is taking place, I have coded the mutation to follow my own preference.

5.8.8 Affricate mutations

This is comparatively infrequent type of soft mutation which affects the affricate consonant [tʃ], voicing the consonant and producing [dʒ] (which is usually realized orthographically as a

‘j’). This normally affects established English loans, such as *tsips* (chips) or *tsicyd* (cheeky).

For example, in extract 37 the participant mutates the loanword *tsicyd*:

*GLE: am° **ji**clyd de am° **ji**clyd

%gls: for cheeky TAG for cheeky

%eng: *how cheeky, how cheeky*

Extract 37 An example of affricate mutation applied to the loanword *tsicyd*

There are comparatively few tokens of this mutation in the corpus, perhaps due to the fact that [tʃ] initial words are themselves rare and typically only affect borrowings from English. This mutation is considered to be a relatively novel development which is tied to informal registers in particular, and is not considered part of the traditional mutation system (Thorne 1993). Although this phenomenon of course seems worthy of a research project in its own right, any tokens with this kind of mutation are removed from the analysis.

5.8 Rbrul Analysis

This section describes the statistical analysis of factors which may affect the mutation process. The aim of this analysis is to establish which factors affect these mutation variables to the level of significance. Factors are divided into two types: social factors, which are intended to test whether the speaker’s background affects the variables; and language internal factors, which are intended to see whether the qualities of the words being mutated may in some way skew the data. The details for the social factors are taken from answers to questionnaires which participants in the original *Siarad* project completed before being recorded. For each token in the spreadsheet, the relevant social factors (such as age and gender) for the speaker are inputted on the same row. The factors were as follows.

5.8.1 Age

Age is a continuous variable, with the youngest participant being aged 10 and the oldest 89. I will be considering this variable to be the most important in the study, as it can provide information about the status of variables (such as whether they are stable or unstable) as well as whether variables seem more widely used by particular age groups, which may tie in to their relative statuses as sociolinguistic variables.

5.8.2 Gender

Gender is a categorical variable with two categories, male and female. As was previously discussed in the chapter on mutation, many sociolinguistic studies have found correlations between various kinds of sociolinguistic variables and gender (Tagliamonte 2012), while Ball (1984) claimed that men mutated more frequently than women due to higher exposure to formal Welsh. An analysis of this variable will allow comparisons to be made with such findings.

5.8.3 Level of Education

Level of education is an ordinal variable with a five point scale. Lower numbers denote a lower level of education than higher numbers. The scale is described in full in section (5.1.3.2). Based on the findings of Thomas and Gathercole (2007), which suggest that input strongly affects acquisition of mutation, my hypothesis in terms of this variable is that speakers with a higher level of education may be more likely to use mutation variables standardly, and that this tendency will be stronger for variables which are frequently used less standardly in vernacular speech, for example those which require formal education to acquire fully.

5.8.4 Initial consonant

This is a categorical variable which refers to the initial consonant of the word that is mutated. As each mutation type affects a different set of consonants, the categories for this variable depend on the trigger being analysed. See table 27 for the structure of the analysis:

Mutation trigger	Initial consonants affected
(am)	/k/ /t/ /p/ /g/ /b/ /d/ /l/ /r/
(ei.m)	/k/ /t/ /p/ /g/ /b/ /d/ /l/ /r/
(ei.f)	/k/ /t/ /p/
(o)	/k/ /t/ /p/ /g/ /b/ /d/ /l/ /r/
(yn)	/k/ /t/ /p/ /g/ /b/ /d/
(a)	/k/ /t/ /p/
(â/gyda)	/k/ /t/ /p/
(fy)	/k/ /t/ /p/ /g/ /b/ /d/

Table 27 The range of consonants mutated by each mutation trigger

This analysis will allow me to see whether certain consonants are more or less likely to be mutated standardly than others.

5.8.5 Place names

During the process of analysis I noted a pattern of place names seeming to resist mutation frequently following the triggers (o), (am) and (yn). There also seems to be an effect related to the relative frequency (or perhaps familiarity versus unfamiliarity) of place names, with e.g. *Cymru* (Wales) and *Bangor* more likely to be mutated standardly than e.g. *Trasbwl* and *Cwm*

Penllafar. This effect may have a particular effect on analyses of older versus younger and rural versus urban speakers, since older and rural speakers seem to use low frequency place names more often than their younger and more urban counterparts. For this reason, place name was added to the analysis as a categorical variable, with two possible categories: ‘yes’ and ‘no’. Additionally, in order to see whether the frequency of place names is a significant factor, I have included a second factor ‘frequency of place name’ for the variable (yn), which is the variable which seems most strongly affected by this pattern. The relative frequency of each place name following (yn) is calculated using the Excel function COUNTIF. These factors and their impact on the data will be discussed more fully in the next chapter.

5.9 Variables not considered

Certain variables were not considered in the analysis, for a variety of reasons. I had initially intended to use dialect as a factor, as it is a variable which is widely cited by earlier researchers. However, this variable had to be discarded due to difficulties concerning how to code for dialect. Using geographical information from participants’ questionnaire details was unhelpful, as many had complicated biographies and dialects which seemed not to ‘match’. For example, a teenage speaker who had spent the first ten years of their life in South Wales seemed to have acquired a Northern accent, and thus was difficult to categorise convincingly. An approach using features of ‘Southern’ speech to identify dialect could have been developed, such as whether speakers used *fe* or *fo* as the third person singular masculine object pronoun, but I decided that this would have been a misallocation of resources. Additionally, north Welsh dialect speakers substantially outnumber speakers of other dialects in the corpus, who are spread out across many different regions. Ultimately, the sample methodology and the make-up of the *Siarad* corpus make this type of answer unfeasible. A part-way solution for the *Radio Cymru* corpus analysis section will be to use programmes with exclusively Northern presenters, as well as with predominantly northern guests.

Class is another variable that will not be analysed directly. It has been described as one of the key sociolinguistic variables, but has not been widely studied in the Welsh language context. This variable will not be considered mainly because of the lack of data in the questionnaire pertaining to the subject. Although the occupation of each participant was noted on the questionnaire, which could perhaps have been used to try and categorize speakers into social class groups, I decided against including it as a factor, as this information alone does not constitute a sufficient basis for a thorough analysis of class. This issue is one which merits future research – I agree with G. Williams (1987) and Robert (2011) that class may be a significant factor in variation in Welsh, despite the arguments of Ball (1984) and others to the contrary - but the analysis in this case is constrained by the available data.

5.10 Code-switching

The code-switching analysis primarily relies on a methodology which counts proportions of English versus Welsh words for each speaker in the corpus (Deuchar 2006). This involves tagging words in order to classify them according to their language category. An important boon for this aspect of the study is the fact that this process has already been carried out fully for the *Siarad* corpus (Deuchar et al 2014). As with the mutation variables, the main aim of this part of the study is to produce percentage scores reflecting average usage of code-switching across age groups in an informal setting.

5.10.1 Coding and tagging code-switches

Since the explicit aim of the researchers working on the *Siarad* corpus was to collect data on code-switching, much of the work involved in categorizing switches has already been put in place. This analysis will largely rely on this fact. All words in the corpus have been coded for language as either Welsh, English or ambiguous. Words were considered Welsh if they appeared in any of the following three Welsh reference dictionaries: *Geiriadur Prifysgol*

Cymru (GPC Ar Lein 2015), *Geiriadur yr Academi* and *Cysgair*, but not in English dictionaries. Since the CHAT transcription system works on the basis that there is a ‘default’ language in each conversation, Welsh words in the *Siarad* corpus are not explicitly tagged by transcribers, are coded automatically in the underlying programme with the tag @cy. Words are tagged as English if they can be found in the Oxford English Dictionary, but not in any of the Welsh dictionaries mentioned above. English words are explicitly tagged by transcribers with the tag @s. A third category of word, undetermined words, are those which are present both in Welsh and English reference dictionaries. Also tagged as ambiguous are words which cannot be related to either language, such as brands like *Tesco*, or place names which are the same in both languages, like Bangor. These words are tagged with @s:und. Words tagged as undetermined are removed from the next stage of analysis to avoid influencing the data any further.

5.10.2 Calculating group scores for code-switching

In order to have a consistent index of code-switching use for comparison between the *Siarad* and *Radio Cymru* corpora, scores for all speakers are calculated as a percentage of English words that speakers insert or alternate with their Welsh. For example, in the transcript Davies 1, the speaker NON uses 1,875 words which are tagged as Welsh, 65 words which are tagged as English and 365 words which are tagged as undetermined. NON’s code-switching score is thus calculated as followed:

$$65/(1875+65) = 3.35\%$$

In order to calculate group scores, all percentage mean scores for individuals within a given group are averaged to create a score reflecting average usage for that group.

5.10.3 Statistical analysis of code-switching

Unlike the analysis of mutation described above, the code-switching patterns of the *Siarad* corpus have been described by previous researchers (e.g Lloyd 2008, Parafita et al 2011), and

their findings have been described in the code-switching section of this thesis. My statistical analysis of code-switching will focus on the same social variables as the mutation analysis, while removing the language internal variables (including initial consonant and place names) so that the influence of social factors on both kinds of variables may be compared.

5.11 Summary

This chapter has provided an overview of the methodology adapted to organize and code the *Siarad* corpus data, particularly in terms of mutation. Significant challenges were faced in trying to avoid confounding factors, and for this reason a comprehensive exclusion list was developed which allows problematic tokens to be removed from the data set. The next section will present the result of the analysis for each variable across a range of internal and external factors.

6. *Siarad* Corpus Analysis Results

6.0 Introduction

This section provides the results of the *Siarad* corpus analysis, and is split into two main parts. The first section provides the results of a statistical analysis using the free statistical software package RBRUL, which is a modified version of R specifically adapted to deal with sociolinguistic data (Johnson 2009). After briefly discussing the varying frequencies of tokens for each mutation variables (6.1) I will move on to describe the results of the analysis of internal variables; place name (6.2.1), frequency of place names (6.2.2) and initial consonant (6.2.3), and the way in which I adapted my analysis to take these findings into account. Following this I will present the results of social factors analysis for each mutation variable in turn, looking at four factors age (6.3.1), gender (6.3.2), level of education (6.3.3), and language attitudes (6.3.4). The final section presents a brief analysis of code-switching using the same social factors (6.4.1).

Following the statistical analysis using RBRUL, I will explore the distribution of each variable in apparent time, with figures for each variable showing their standard use by seven different age groups. This will provide insight into each variable's status in the speech community, including whether variables can be considered stable, undergoing community wide change, or subject to age grading. I will then summarize the results of the *Siarad* corpus analysis (6.5), before moving on to the next section.

6.1 Mutation Token Frequencies

This section displays the relative frequencies of tokens each variable in the *Siarad* corpus. Figure 4 below illustrates the relatively large range in frequencies of occurrence of the mutation variables in the dataset.

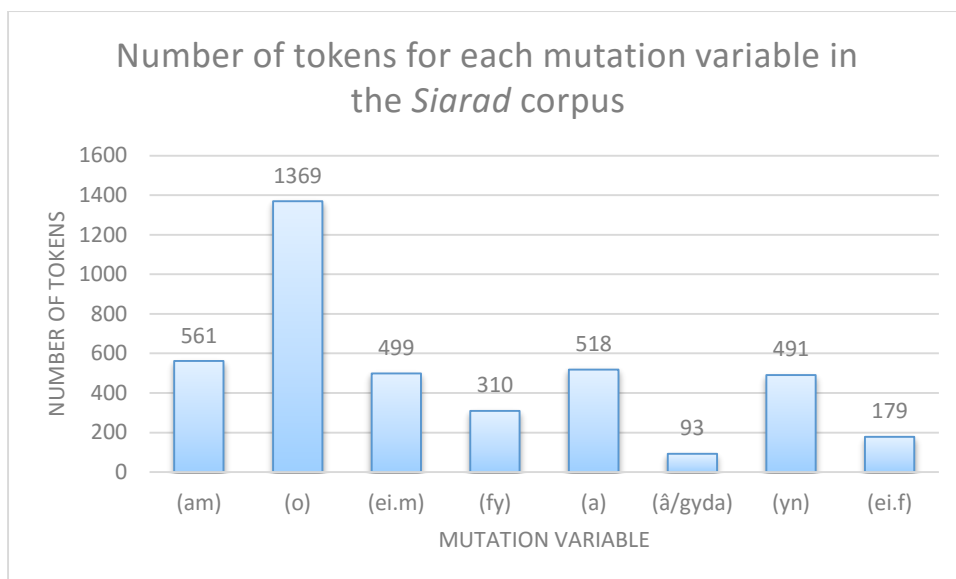


Figure 4 Frequency of tokens of mutation each variable in the *Siarad* corpus

The most frequently occurring variable is the preposition (o), which at a 1,369 tokens appears in the corpus more than twice as often as the second most frequent variable (am). The majority of the other variables form a group of intermediate frequency, including (am), (yn), (fy), (ei.m) and (a), which have an average of around 500 tokens each. The frequency of (ei.f) was somewhat lower than that of the other variables with only a 179 tokens, while (â/gyda) was notable for having slightly fewer than a hundred tokens, with 93 tokens.

6.2 Internal Factors

Before presenting the results for the language external social factors analysis, I will first present the findings of the internal factors analysis. This is because some of these factors had a significant effect, and may have skewed my initial analysis of the language external factors. These findings thus had to be taken into account when analysing the social factors. The internal factors are *place name* (6.2.1), *frequency of place name* (6.2.2) and *internal consonant* (6.2.3).

6.2.1 Place names

Three of the variables, (am), (o) and (yn), were all significantly affected by the place name factor. In other words, target lexical items that were place names were significantly less likely to undergo standard mutation than non-place names following (am), (o) and (yn). The frequency at which place names, as opposed to other kinds of lexical items, followed each of these variables differed to a great extent. This is important because the extent to which the factor affected each variable varied. Figure 5 shows that (yn) is by far the most likely to be followed by place names, with place names forming the lexical target for mutation 77.08% percent of the time. In contrast, (o) and (am) are followed by place names 10.58% and 4.64% of the time respectively.

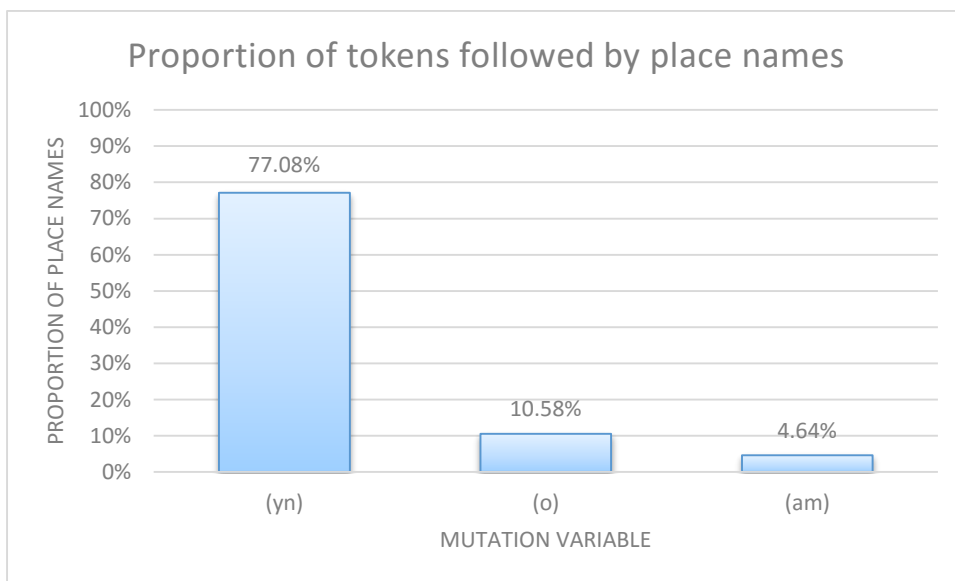


Figure 5 The proportion of tokens for each variable that were followed by a place name

6.2.1.1 Place names – (o)

For the variable (o), the independent variable ‘place names’ was found to have a highly significant effect on the standard usage of variable (o) to $3.41e-40^*$ in a one level analysis using RBRUL.

Factor	Logodds	Tokens	Standard/standard+nonstandard	Centered factor weight
Non-place name	1.526	1150	0.852	0.821
Place name	-1.526	136	0.265	0.179

Table 28 Results of analysis of factor place names for (o)

As can be seen in table 28, the factor *non-place name* has a very high centred factored weight of 0.821 compared to 0.179 for *place name*, indicating that non-place names are highly likely to undergo standard mutation following (o), while the opposite is true for place names. This confirms my prediction that place names would be significantly less likely to be mutated standardly than non-place names. Given the frequency of tokens involved, I believe that this factor may have had a slight skewing effect on the data. To illustrate this, figure 6 compares standard mutation percentages across three conditions: (o) including place names, (o) excluding place names, and (o) with only place names.

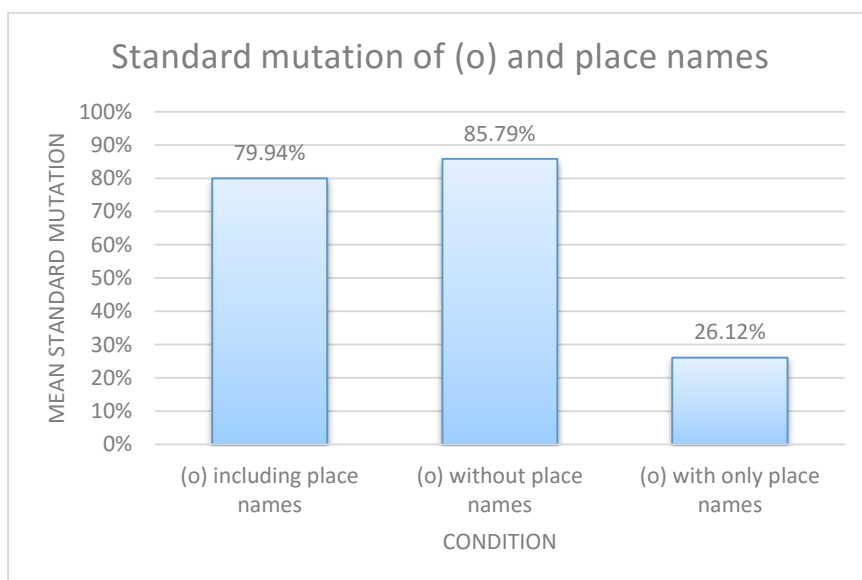


Figure 6 Standard mutation of (o) across three conditions, with, without or only place names

Place names are mutated only 26% of the time, while non-place names were mutated 85% of the time. Given that place names make up around 10% of target words for (o), this factor clearly has the potential to affect the data to a great extent.

6.2.1.2 Place names – (am)

As with (o), the independent variable *place names* was also significantly correlated with standard use of (am) at the level of significance of 4.38e-07*.

Factor	Logodds	Tokens	Standard/standard+nonstandard	Centered factor weight
Non-place name	1.173	534	0.747	0.764
Place name	-1.173	26	0.269	0.236

Table 29 Results of analysis of factor place names for (am)

This means that as with (o), place names are significantly less likely to be mutated following the mutation trigger (am), while the opposite is true for non-place names. Figure 7 illustrates the difference across three conditions for (am). The difference between the ‘with place names’ and the ‘without place names’ conditions is slightly smaller for (am) than it was for (o), although this might be due to the fact that place names constituted a larger proportion of tokens for (o) than (am).

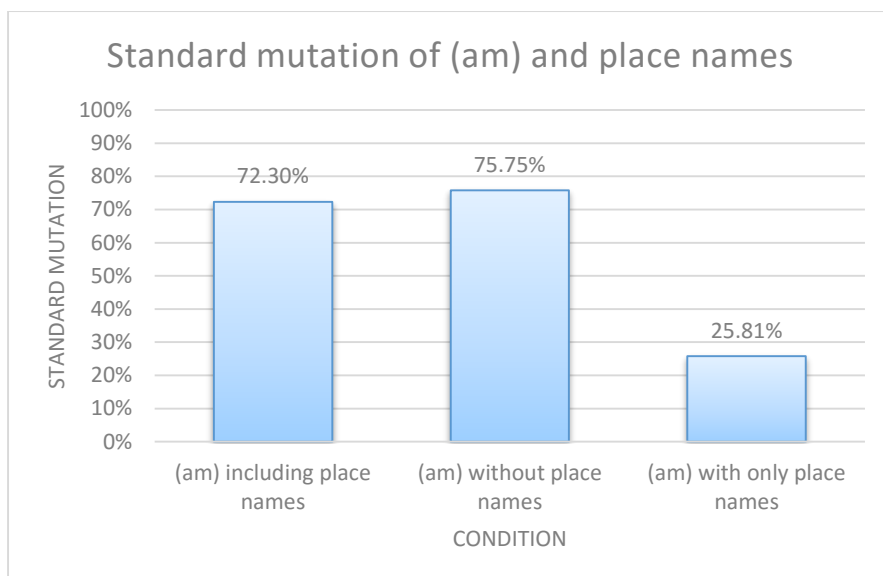


Figure 7 Standard mutation of (am) across three conditions, including, without and only place names

6.2.1.3 *Place names – (yn)*

The final variable in this section, (yn), was also significantly affected by the independent variable *place names* to the level $2.67e-07^*$. Table 30 provides more details about the results of the analysis.

Factor	Logodds	Tokens	Standard/standard+nonstandard	Centered factor weight
Non-place name	1.061	76	0.395	0.743
Place name	-1.061	279	0.222	0.257

Table 30 Results of analysis of factor place names for (yn)

As with (o) and (am), the *non-place name* factor favours standard tokens, while *place name* favours non-standard tokens. Figure 8 shows average index scores for each place name condition (including, excluding, and only place names) with (yn). Once again, the data follows the same pattern, with the ‘without place names’ condition showing a higher rate of standard mutation following (yn) than the ‘including place names’ condition.

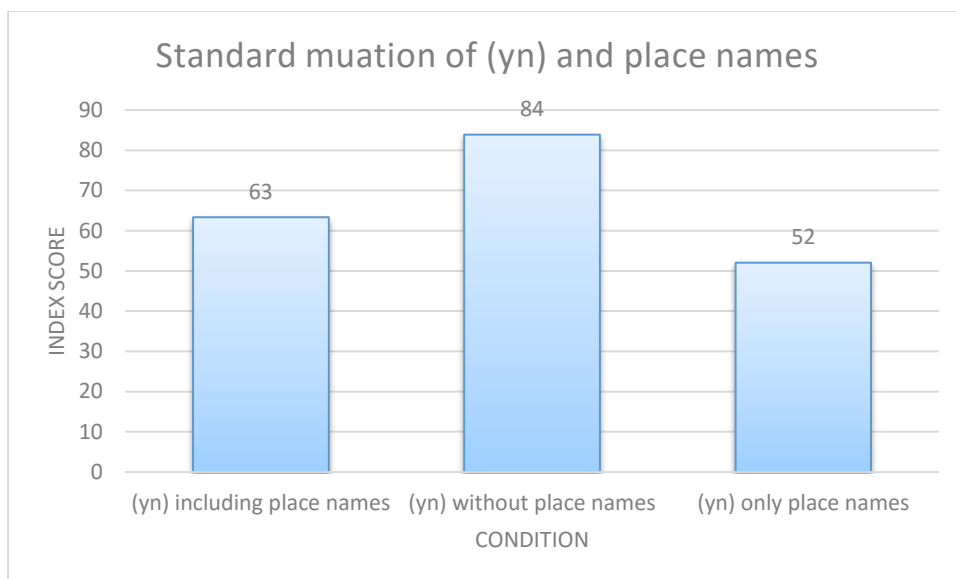


Figure 8 Standard mutation of (yn) across three conditions, with, without or only place names

6.2.1.4 Place names – Summary

Taken together, the results for all three variables seem to clearly suggest that target words which are place names are significantly less likely to undergo standard mutation than common nouns.

6.2.1.5 Frequency of place names

This analysis was applied exclusively to (yn) in order to test whether the frequency of place names could be related to the unique profile of (yn) when tested by age. The results show that the frequency of place name factor correlates with standard use of (yn) at a level of significance of $1.8e-07^*$, meaning that place names which occur more frequently after (yn) in the corpus tend to be mutated standardly more frequently.

continuous	logodds
+1	0.069

Table 31 Results for analysis of place name frequency factor for (yn)

By way of illustrating this point, figure 9 shows mean index scores for place names divided into five groups by frequency of tokens following (yn); (1-2), (3-5), (6-10) (11-20) and (20+):

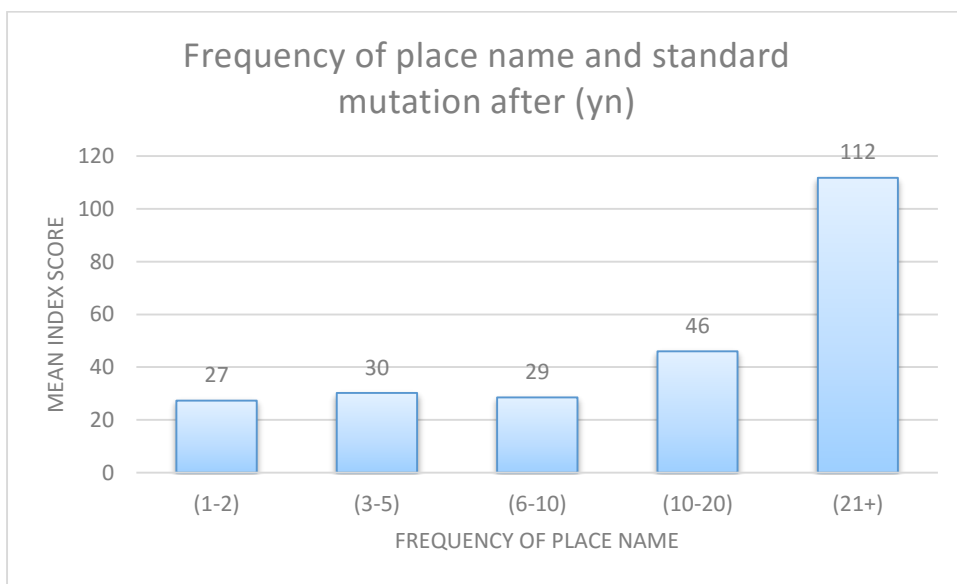


Figure 9 Standard mutation and frequency of place names in the *Siarad* corpus

Although rates of standard mutation following (yn) are relatively similar in the first three frequency groups, the mean index score is noticeably higher for the 21+ frequency group. Further, certain high frequency place names, such as *Cymru* (Wales) and *Caerdydd* (Cardiff) have very high mean index scores, at 125 points each, which are much higher than the average for (yn). I believe that this phenomenon may explain to some extent why (yn) differs so significantly from other variables, particularly in terms of the analysis of age.

There seems to be a pattern wherein age groups differ in the way they use place names, with certain groups generally using more frequent place names, while older groups tend to use less

frequent place names. To clarify whether this was happening, I conducted an analysis in RBRUL with place name frequency as the dependent variable and age group as the independent variable. The result was significant to the level of 0.00608* indicating that younger age groups were in fact generally more likely to use more frequent place names as I predicted, with the (30-39) age group with the highest mean at 21.

factor	coef	tokens	mean
(30-39)	7.166	33	21.000
(20-29)	1.226	50	15.060
(10-19)	1.203	27	15.037
(40-49)	-0.134	30	13.700
(70+)	-1.685	47	12.149
(60-69)	-3.651	71	10.183
(50-59)	-4.124	76	9.711

Table 32 Results of RBRUL analysis of mean frequency of place names by age group

What this seems to mean in practice is that younger participants (and particularly those in the 30-39 age group) seem to be using place names like *Cymru* and *Caerdydd* more frequently,

while older speakers are using less frequent place names, such as the names of local villages. This pattern has the potential to have a strong effect on the data, since it may confound attempts to measure standard mutation by age accurately if younger speakers have artificially inflated standard mutation means. This effect cannot be easily controlled for, but emphasizes the need for this variable to be calculated as two separate conditions – *only place names* and *no place names* – on some occasions. This will be discussed further in the section for correlating (yn) and age (6.3.1.4).

6.2.2 Initial Consonant

This section will provide the results for the analysis of the relationship between the independent variable *initial consonant* and standard mutation for each mutation variable. As place name was found to be a significant internal factor, I will analyse the affected variables (o) (am) and (yn) with place name tokens removed. The variable (â/gyda) will be excluded from this analysis as it has only two standard tokens. No specific predictions were made for the relationship between initial consonant and standard mutation, but any significant findings will be taken into account moving forward.

6.2.2.1 Initial consonant – (fy)

The RBRUL analysis of the variable (fy) and independent variable *initial consonant* found no significant relationship, with the level of significance at 0.874:

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
k	0.325	60	0.683	0.581
p	0.20	42	0.667	0.579
g	0.119	38	0.632	0.53
t	-0.015	87	0.690	0.496
b	-0.039	56	0.607	0.49
d	-0.710	14	0.714	0.33

Table 33 Result of analysis of variable (fy) and initial consonant

6.2.2.2 Initial consonant – (am)

The variable (am) was significantly correlated with the independent variable *initial consonant* at the level of 4e-04*.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
b	1.009	66	0.879	0.733
m	0.687	64	0.844	0.665
k	0.555	101	0.812	0.635
t	0.192	53	0.736	0.548
d	-0.037	73	0.712	0.491
g	-0.061	34	0.706	0.485
p	-0.094	103	0.680	0.476
ɾ	-0.597	22	0.591	0.355

ɫ	-1.653	18	0.389	0.161
---	--------	----	-------	-------

Table 34 Result of analysis of variable (am) and initial consonant

The consonants varied widely in their factor weights from 0.733 for [b] to 0.161 for [ɫ]. One notable pattern seems to be that marked consonants [r] and [ɫ] are much less likely to be mutated standardly. Also, consonants with fewer tokens are generally towards the bottom of the scale.

6.2.2.3 Initial consonant – (o)

The variable (o) was significantly correlated with the initial consonant variable, at the level of 0.000173*. Table 35 shows a relatively broad spread of factor weights for each consonant, with [p] showing the highest tendency towards standardness and [r] with the lowest.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
p	0.672	424	0.887	0.662
d	0.454	97	0.876	0.612
b	0.425	72	0.875	0.605
g	0.422	158	0.867	0.604
m	0.127	60	0.850	0.532
k	-0.103	143	0.839	0.474
ɫ	-0.423	93	0.763	0.396
t	-0.574	71	0.775	0.36
r	-0.999	32	0.688	0.269

Table 35 Result of analysis of variable (o) and initial consonant

As with the variable (am), the consonants [t] and [r] had particularly low rates of standard mutation. Also interesting is that the consonants with more frequent tokens seem to be generally more standard, while infrequent consonants are less standard.

6.2.2.4 Initial consonant – (yn)

The variable (yn) was correlated significantly with initial consonant to the level of $2.08e-11^*$.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
k	1.865	94	0.564	0.866
g	1.213	35	0.229	0.771
p	0.180	82	0.146	0.545
b	0.160	80	0.200	0.54
d	-0.221	26	0.077	0.445
t	-3.198	38	0.026	0.039

Table 36 Result of analysis of variable (yn) and initial consonant

The velar consonants [k] and [g] were mutated standardly at substantially higher rates than other consonants, with centred factor weights of 0.866 and 0.771 respectively. The consonants [p] [b] and [d] form an intermediate group, while [t] by contrast had an extremely low centred factor weight of 0.039. One explanation for this difference may relate to the correlation between place name frequency and standard mutation following (yn) – the very high factor weight for [k] in particular may reflect the fact that some of the most frequent place names, including *Cymru*, *Caerdydd* and *Caernarfon* begin with [k], while [t] occurred mostly in infrequent place names like *Tremadog* and *Trasbwill*.

6.2.2.5 Initial consonant – (a)

Initial consonant was not significantly correlated with the dependent variable (a), with the level of significance at 0.523.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
t	0.288	71	0.085	0.572
k	0.022	144	0.083	0.506
p	-0.311	281	0.046	0.423

Table 37 Result of analysis of initial consonant (a) and initial consonant

6.2.2.6 Initial consonant – (ei.m)

As the factor [r] had only standard tokens (a total of seven), I removed it from the Rbrul analysis for (ei.m) to avoid a potential skewing effect. The subsequent analysis found that (ei.m) has a significant relationship with initial consonant, at the level of significance of 0.00918*.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
g	1.275	99	0.960	0.782
t	0.552	65	0.908	0.635
b	0.333	57	0.912	0.583
k	0.164	72	0.917	0.541
ɫ	-0.178	28	0.893	0.456
m	-0.252	37	0.838	0.474

p	-0.745	61	0.803	0.322
d	-1.148	44	0.750	0.241

Table 38 Result of analysis of initial consonant (ei.m) and initial consonant

6.2.2.7 Initial consonant – (ei.f)

At the level of significance of 0.282, initial consonant was not a significant factor for the aspirate mutation variable (ei.f)¹⁸.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
p	0.572	41	0.732	0.639
t	-0.119	41	0.756	0.47
k	-0.453	68	0.662	0.389

Table 39 Result of analysis of initial consonant (ei.f) and initial consonant

6.2.2.8 Initial consonant – Summary

- There were significant differences for four mutation variables: (am), (o), (yn) and (ei.m)
- There were no significant difference for three variables: (fy), (a) and (ei.f)
- Despite these significant results, there seemed to be no clear pattern of variation that was consistent across all variables.

Although the analysis found several significant results in the analysis of initial consonants and the likelihood of standard mutation, there does not seem to be a clear pattern that can be

¹⁸ I removed the intermediate SM tokens from the RBRUL analysis of (ei.f), partly because the very low frequency of some consonants affected by such tokens (e.g. [g] in the utterance “wnes i (ei) weld hi”) caused the analysis to become skewed.

generalized. Because of this, in contrast to the removal of place names from subsequent analysis, I will retain all consonants for future analysis where possible.

6.3 Language external Variables

I will now turn my attention to the language external variables. These include age (6.3.1), gender (6.3.2), level of education (6.3.3) and attitudes (6.3.4).

6.3.1 Age

This section will look at each variable in turn and present the level of significance for each one.

Age was measured as a continuous variable.

6.3.1.1 Age - (fy)

The nasal mutation trigger (fy) has a strong positive correlation with age to the level of $1.97e-23^*$. The result of the RBRUL analysis is presented in table 40.

continuous	logodds
+1	0.096

Table 40 Results of RBRUL analysis of standard mutation after (fy) and age

6.3.1.2 Age – (am)

As place names were found to have a strong significant effect on the standard mutation of (am), I decided to remove all place names from the analysis of the variable. In the subsequent analysis, standard use of soft mutation trigger (am) is positively correlated with age to the level of significance 0.0125^* . Table 41 shows the details of the results of the analysis in RBRUL:

continuous	logodds
+1	0.015

Table 41 Results of RBRUL analysis of age and standard use of (am)

6.3.1.3 Age – (o)

Having removed place names, the mutation variable (o) correlates significantly with age at the level of 1e-06*.

continuous	logodds
+1	0.024

Table 42 Results for (o) without place names analysis with age

6.3.1.4 Age – (yn)

In order to try and correct for the potentially confounding effect of the place name and place name frequency on (yn), place names were removed from the analysis. This meant reducing the number of tokens substantially, from 424 to 91. Without place names included, the relationship between (yn) and age has a significant positive correlation at the level of 0.00904*.

Table 43 below shows the results of the one way RBRUL analysis of (yn) without place names:

continuous	logodds
+1	0.049

Table 43 Result of an one way analysis of age and (yn) without place names

6.3.1.5 Age – (a)

The aspirate mutation trigger (a) has a positive significant correlation with age at the level of significance of 0.000182*:

continuous	logodds
+1	0.044

Table 44 Results of a one way analysis of age and (a)

6.3.1.6 Age – (ei.m)

The variable (ei.m) was significantly affected by age to the level of 7.25e-06*, with older speakers in *Siarad* using more standard variants of the variable than younger speakers.

continuous	logodds
+1	0.041

Table 45 Results of RBRUL analysis of age and (ei.m)

6.3.1.7 Age – (ei.f)

The variable (ei.f) was found to have a significant positive correlation with age to the level of 6.1e-11*, with older speakers using more standard variants than younger speakers.

continuous	logodds
+1	0.1

Table 46 Results of RBRUL analysis of age and (ei.f)

6.3.1.8 Age – (â/gyda)

With just two standard tokens of this variable, a statistical analysis of this variable was not considered possible. The distribution of the tokens seems essentially random, with one standard token from the youngest (10-19) age group and the other from the (60-69) group.

6.3.1.9 Mutation variables and age – Summary

Here I briefly outline some of the main findings of the analysis of mutation and age:

- The mutation variables (am), (o), (fy), (ei.m), (ei.f) and (a) all showed statistically significant age variation.
- The variable (â/gyda) was not analysed statistically due to a lack of tokens, and of standard variants in particular. It was found to be a very rare mutation trigger.
- The variables (o), (am), and (yn) seem to have been skewed by the inclusion of place names -and the frequency of certain place names - in the analysis. Place names were thus removed from the analysis of age for these variables.

6.3.2 Gender

This section provides the results for the analysis of each mutation variable with the language external independent variable gender. No particular predication were made for this variable, although a study by Ball (1984) found that men were more likely to use standard mutation than women, while traditional variationist study of phonetic variants have consistently found that women are more likely to use prestigious variants, as well as incoming forms in some situations (Labov 2001.b).

6.3.2.1 Gender - (fy)

The variable (fy) correlates significantly with the independent variable gender to the level of significance of 0.0161*, with male participants using significantly more standard variants than female participants.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
M	0.457	123	0.78	0.612
F	-0.457	174	0.58	0.388

Table 47 Results of the RBRUL analysis of (fy) and gender

6.3.2.2 Gender - (am)

With the level of significance at 0.882, standard use of (am) does not correlate significantly with the gender of the speaker.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
F	0.017	225	0.747	0.504
M	-0.017	309	0.748	0.496

Table 48 Results of the RBRUL analysis of (am) and gender

6.3.2.3 Gender - (o)

The variable (o) did not correlate significantly with gender, with the level of significance at 0.798.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
F	0.024	588	0.849	0.506
M	-0.024	562	0.856	0.494

Table 49 Results of the RBRUL analysis of (o) and gender

6.3.2.4 Gender - (yn)

At the level of significance of 0.242 gender was not significantly correlated with the variable (yn).

Factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
F	0.201	154	0.266	0.55
M	-0.201	201	0.254	0.45

Table 50 Results of the RBRUL analysis of (yn) and gender

6.3.2.5 Gender - (a)

At the level of significance 0.638, the mutation variable (a) was not significantly correlated with the independent variable gender.

factor	logodds	tokens	standard/standard+nonstandard	centered weight	factor
F	0.107	208	0.067	0.527	
M	-0.107	288	0.059	0.473	

Table 51 Results of the RBRUL analysis of (a) and gender

6.3.2.6 Gender - (ei.m)

Gender was not found to be significantly correlated with the variable (ei.m), with a level of significance of 0.26.

factor	logodds	tokens	standard/standard+nonstandard	centered weight	factor
F	0.218	224	0.920	0.554	
M	-0.218	239	0.854	0.446	

Table 52 Results of the RBRUL analysis of (ei.m) and gender

6.3.2.7 Gender - (ei.f)

Gender was at the level of significance for the variable (ei.f) at 0.046*, with male participants using significantly more standard tokens than female participants.

factor	logodds	tokens	standard/standard+nonstandard	centered weight	factor
M	0.632	52	0.846	0.653	
F	-0.632	98	0.633	0.347	

Table 53 Results of the RBRUL analysis of (ei.f) and gender

6.3.2.8 Gender - Summary

There was a significant difference between the male and female factors in only two variables, (fy) and (ei.f), with a stronger significant result for (fy). Both analyses suggest that the male participants used more standard variants than the female participants.

6.3.3 Level of Education

This section presents the result of the analysis of level education for the mutation variables. As it has too few standard tokens, (â/gyda) is excluded from the analysis. The analysis was conducted as a one way multivariate analysis in RBRUL, with the independent variable level of education, and with the mutation triggers variables as the dependent variables. The analysis includes five factors on an ordinal scale from 1 to 5, with a greater value representing a higher level of formal education. Given that formal education may allow greater exposure to standard Welsh, my prediction for this variable is that speakers with a more advanced level of education will use more standard variants.

6.3.3.1 Level of education - (fy)

The RBRUL analysis of (fy) and level of education found a significant positive correlation at the level of 0.00953*, indicating that speakers with a higher level of education generally used more standard variants than speakers with less education:

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
4	1.003	93	0.817	0.732
1	0.072	39	0.513	0.518
5	-0.101	35	0.771	0.475
3	-0.379	97	0.515	0.406
2	-0.595	33	0.727	0.356

Table 54 Standard us of (fy) and level of education

6.3.3.2 Level of education - (am)

The variable (am) was not significantly correlated with the independent variable level of education, with a significance level of 0.796.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
4	0.195	242	0.769	0.549
3	0.109	120	0.733	0.527
1	0.045	51	0.725	0.511
2	-0.164	72	0.764	0.459
5	-0.185	49	0.673	0.454

Table 55 Standard us of (am) and level of education

6.3.3.3 Level of education - (o)

Level of education was correlated significantly with the mutation variable (o) at the level of $4.87e-05^*$, indicating that more educated participants tended to use more frequent standard tokens than less educated ones.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
5	1.000	112	0.929	0.731
1	0.403	125	0.896	0.466
4	-0.138	493	0.860	0.511
3	-0.621	255	0.808	0.35
2	-0.645	165	0.812	0.344

Table 56 Standard us of (o) and level of education

6.3.3.4 Level of education - (yn)

The mutation variable (yn) was correlated significantly to the level of $2.3e-06^*$ with the level of education variable. This indicates that speakers with a more advanced level of education tended to use standard tokens more frequently than speakers with less education.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
5	1.720	44	0.409	0.848
4	1.256	159	0.302	0.778
3	0.586	83	0.277	0.642
1	-1.107	34	0.059	0.248

2	-2.455	35	0.029	0.079
---	--------	----	-------	-------

Table 57 Standard us of (yn) and level of education

6.3.3.5 Level of education - (a)

Level of education correlates positively with standard use of the variable (a) at the level of significance of 0.000274*, which indicates that speakers who have received more advanced education use standard tokens more frequently.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
5	1.927	44	0.227	0.873
2	0.111	74	0.081	0.528
4	0.051	185	0.054	0.513
3	-0.507	125	0.032	0.376
1	-1.582	68	0.015	0.171

Table 58 Standard us of (a) and level of education

6.3.3.6 Level of education - (ei.m)

Level of education correlates significantly with the soft mutation variable (ei.m) to the level of 0.0132*, indicating that speakers who are more educated tend to use this variable more standardly:

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
5	0.772	57	0.895	0.684
4	0.750	194	0.923	0.679
1	0.097	52	0.885	0.524
3	-0.361	105	0.819	0.411
2	-1.258	55	0.873	0.221

Table 59 Standard us of (ei.m) and level of education

6.3.3.7 Level of education - (ei.f)

Level of education was significant for the variable (ei.f) with a level of significance of 0.0323*, indicating that speakers with more advanced education use more standard variants of this variable.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
5	1.708	19	0.789	0.847
3	0.989	39	0.667	0.729
4	0.646	59	0.729	0.656
1	-1.349	18	0.500	0.206
2	-1.994	15	0.867	0.12

Table 60 Standard us of (ei.f) and level of education

6.3.3.8 Level of education - Summary

Level of education was found to be significant for the majority of the mutation variables, with the sole exception of (am). The overall pattern in the data seems to imply that education has a strong influence over speaker's tendency to use standard variants in casual conversation, with more educated speakers tending to use more standard variants than less educated speakers.

6.3.4 Attitudes

This section uses participant responses (via the speaker questionnaires) to the statement 'people should not mix languages' as an indirect measure of the level of purist language ideology held by speakers. Speakers who held stronger beliefs that languages should never be mixed were predicted to use more standard tokens of mutation variables.

6.3.4.1 Attitudes (fy)

The variable (fy) was not found to be significantly correlated with the attitudes variable, with a level of significance of 0.252.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
4	0.757	65	0.662	0.681
2	0.221	65	0.723	0.555
5	-0.238	65	0.738	0.441
3	-0.260	75	0.613	0.435
1	-0.481	27	0.481	0.382

Table 61 Standard us of (fy) and attitudes

6.3.4.2 Attitudes (am)

The variable (am) was just above the level of significance at 0.0519, and thus not significant when correlated with attitudes.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
5	0.429	60	0.850	0.606
4	0.425	144	0.723	0.605
3	0.007	150	0.792	0.502
2	-0.279	128	0.703	0.431
1	-0.583	52	0.635	0.358

Table 62 Standard us of (am) and attitudes

6.3.4.3 Attitudes (o)

The variable (o) was significantly correlated with the attitudes variable to the level of $1.6e-06^*$, indicating that speakers with more purist views used more standard tokens.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
5	1.072	151	0.954	0.745
4	0.117	325	0.871	0.529
3	0.114	325	0.879	0.528
1	-0.519	153	0.791	0.373
2	-0.784	249	0.775	0.314

Table 63 Standard us of (o) and attitudes

6.3.4.4 Attitudes (yn)

The variable (yn) was not significantly correlated with the attitudes question, reaching the level of significance of 0.129.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
4	0.744	92	0.348	0.678
5	0.109	58	0.241	0.527
3	-0.194	79	0.139	0.452
2	-0.316	84	0.286	0.422
1	-0.344	42	0.262	0.415

Table 64 Standard us of (yn) and attitudes

6.3.4.5 Attitudes (a)

The least purist group in the analysis of (a) used no standard tokens, which meant that the factor had to be removed them from the analysis. The subsequent analysis found no significant correlation, with the level of significance at 0.882:

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
5	0.213	51	0.137	0.553
4	0.119	132	0.076	0.53
2	-0.056	154	0.058	0.486
3	-0.277	100	0.050	0.431

Table 65 Standard us of (a) and attitudes

6.3.4.6 Attitudes (ei.m)

The soft mutation trigger variable (ei.m) was correlated significantly with the attitude variable, to the level of $7.22e-05^*$.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
4	1.111	117	0.923	0.752
3	0.653	128	0.930	0.658
5	0.455	97	0.938	0.612
2	0.927	85	0.812	0.284
1	-1.293	36	0.639	0.215

Table 66 Standard us of (ei.m) and attitudes

6.3.4.7 Attitudes (ei.f)

The variable (ei.f) was not significantly correlated with the attitudes variable, with a level of significance of 0.218.

factor	logodds	tokens	standard/standard+nonstandard	centered factor weight
3	0.677	39	0.769	0.663
5	0.240	43	0.698	0.56
2	0.209	27	0.815	0.552
4	0.155	25	0.800	0.539
1	-1.280	16	0.250	0.218

Table 67 Standard us of (ei.f) and attitudes

6.3.4.8 Attitudes - Summary

An analysis of the attitudes variable did not reveal a consistent pattern, with only two of the mutation variables correlating significantly. The statistically significant variables were (ei.m) and (o).

6.4 Code-Switching Results

This section will present the results of the code-switching analysis, including an analysis of age, level of education and other factors. Work on code-switching has already been conducted by other researchers using the same data set, therefore these findings are presented here for reference and comparison with the mutation results. I will return to these findings in later chapters.

6.4.1 Code-switching - age

Code-switching had a significant positive correlation with age in the analysis to the level of $1.99e-07^*$, with younger speakers using proportionally more code-switching than older speakers.

continuous	logodds
+1	-0.001

Table 68 Rbrul results for analysis of code-switching and age

6.4.2 Code-switching by Gender

Gender was not correlated significantly with gender, with the level of significance at 0.598

factor	coef	tokens	mean
M	0.002	67	0.042
F	-0.002	77	0.038

Table 69 Results of RBRUL analysis of gender and code-switching

6.4.3 Code-switching – Level of education

Code-switching was not correlated significantly with level of education, with the level of significance at 0.664.

factor	coef	tokens	mean
1	0.009	20	0.050
3	0.002	38	0.045
2	-0.003	17	0.031
4	-0.003	55	0.037
5	-0.005	14	0.034

Table 70 Results of RBRUL analysis of level of education and code-switching

6.4.4 Code-switching – Attitudes

Code-switching was not correlated significantly with the attitudes variable, with significance at the level of 0.141.

factor	coef	tokens	mean
3	0.007	39	0.048
1	0.006	18	0.052
5	0.000	16	0.033
2	-0.001	33	0.044
4	-0.013	38	0.025

Table 71 Results of RBRUL analysis of attitudes and code-switching

6.5 Summary of RBRUL results

Table 72 presents the level of significance for each variable for each factor tested in the analysis:

Variable	Initial consonant	Age	Gender	Level of education	Attitudes
<u>(fy)</u>	0.874	1.97e-23*	0.0161*	0.00953*	0.252
<u>(am)</u>	4e-04*	0.0125*	0.882	0.796	0.0519
<u>(o)</u>	0.000173*	1e-06*	0.798	4.87e-05*	1.6e-06*
<u>(yn)</u>	2.08e-11*	0.00904*	0.242	2.3e-06*	0.129
<u>(a)</u>	0.523	0.000182*	0.638	0.000274*	0.882
<u>(ei.m)</u>	0.00918*	7.25e-06*	0.26	0.0132*	7.22e-05*
<u>(ei.f)</u>	0.282	6.1e-11*	0.046*	0.0323*	0.218
Code-switching	N/A	1.99e-07*	0.598	0.664	0.141

Table 72 Summary of the statistical analysis for all variables

Age was significant for all variables, with younger speakers using more non-standard forms than older speakers.

6.6 Apparent Time analysis

This section focuses on the distribution of each variable according to age group. This allows us to establish how frequently a variable is used standardly, what kinds of differences exist between age groups, and whether the variables resemble previously attested profiles of community change, stability or age grading (see chapter 2). This section uses mean percentages to represent age group scores for binary variables (i.e. those with two variants) and index scores for variables which have three variants.

6.6.1 Apparent Time – (fy)

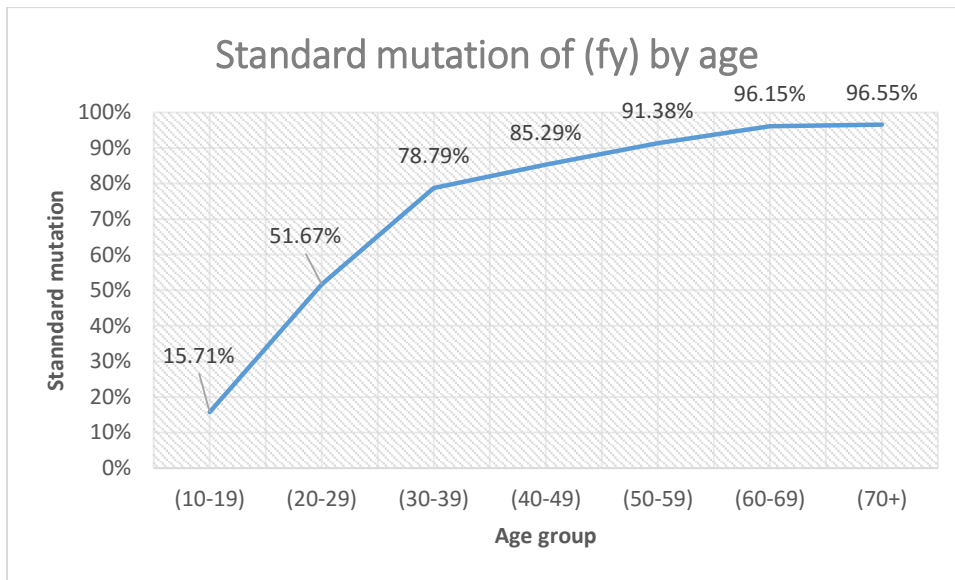


Figure 10 Mean use of standard mutation in the *Siarad* corpus following (fy) by age group

	(10-19)	(20-29)	(30-39)	(40-49)	(50-59)	(60-69)	(70+)
Radical tokens	59	29	7	5	5	1	1
NM tokens	11	31	26	29	53	25	28

Table 73 Number of SM tokens and radical tokens in *Siarad* corpus following (fy) by age group

As figure 10 illustrates, the level of standard mutation approaches a hundred percent among the older age groups, but decreases gradually starting with the (50-59) group, before dropping precipitously from 78.79% in the (30-39) group to 51.67% in the (20-29) year old group. This is followed by an even more extreme decrease in standard usage from 51.67% in the (20-29) group to 15.71% in the youngest (10-19) age group - a drop of 36%. The mutation variable (fy) exhibits the clearest and most dramatic contrast in terms of age of all variables which are analysed in this study.

6.6.2 Apparent Time - (am)

Chart 7 shows percentage means of standard means for (am) by age group, both with and without place names.

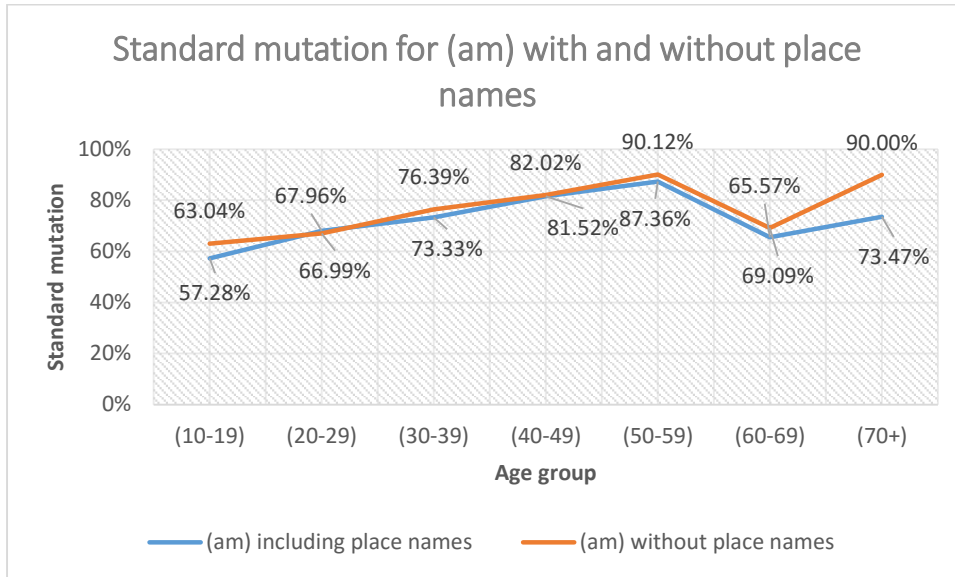


Figure 11 Standard mutation following (am)

	(10-19)	(20-29)	(30-39)	(40-49)	(50-59)	(60-69)	(70+)
Radical tokens	44	33	20	17	11	21	13
SM tokens	59	70	55	75	76	40	36

Table 74 Number of SM tokens and radical tokens in *Siarad* corpus following (am) by age group (including place names)

The data indicates a generally steady pattern of younger age groups using fewer standard variants than older age groups, although the (60-69) age group is something of an outlier to this pattern, with a mean percentage closer to the youngest age groups. The including and excluding place names conditions are generally similar for most age groups, but are highly divergent for

the oldest age group, which has a mean percentage of 90% in the without place names condition - a difference amounting to nearly 25%.

6.6.3 Apparent Time - (o)

As I have argued that (o) is significantly affected by place names as a factor, I decided to include (o) as two separate variables: ‘without place names’ and ‘only place names’. Figure 12 compares the standard mutation following (o) for ‘with place names’ and ‘without place names’.

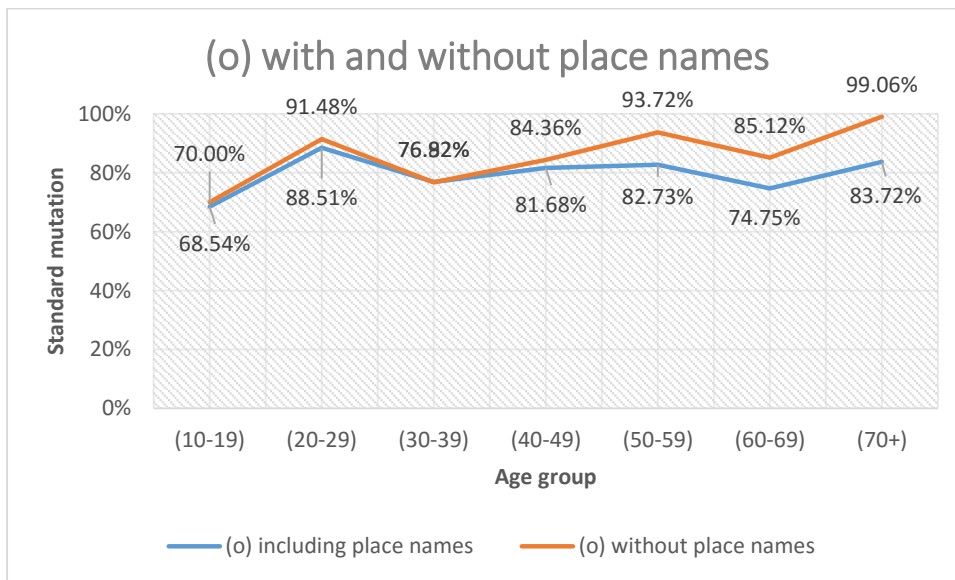


Figure 12 Standard mutation following (o), both with and without place names

	(10-19)	(20-29)	(30-39)	(40-49)	(50-59)	(60-69)	(70+)
Radical tokens	56	27	36	35	43	50	21
SM tokens	122	208	120	156	206	148	108

Table 75 Number of SM tokens and radical tokens in *Siarad* corpus following (o) by age group (including place names)

While both conditions are significant in terms of age, ‘without place name’ shows increased standard usage of the variable among the older age groups, while the scores for younger age groups are largely the same in both conditions. Taking the ‘without place name’ condition as a truer reflection of the actual distribution of the variables, a general, if inconsistent, decline can be seen from the oldest age group’s almost categorical mean use of 99.06% to the youngest group’s mean of 70.00%.

6.6.4 Apparent Time - (yn)

This variable presented the most significant challenges in terms of separating internal and external factors. The initial analysis of age group for (yn), which included place names along with non-place names, produced the following pattern:

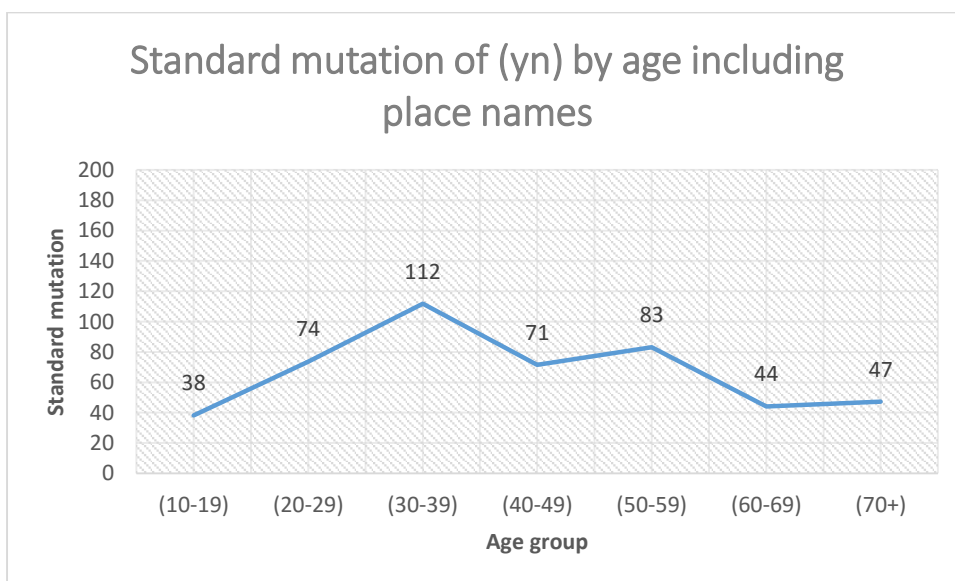


Figure 13 Standard mutation following (yn) for the ‘including place names’ condition by age

	(10-19)	(20-29)	(30-39)	(40-49)	(50-59)	(60-69)	(70+)
Radical tokens	33	42	19	32	28	71	36
SM tokens	11	12	14	8	6	17	9
NM tokens	5	22	26	16	19	14	8

Table 76 Number of SM, NM and radical tokens in *Siarad* corpus following (yn) by age group

According to these results, (yn) is relatively unique as a mutation variable in that it does not show a pattern of declining standard usage from older to younger age groups. Instead, it exhibits an inverted ν shape, with the (30-39) age group holding the highest score and both younger and older speakers using more non-standard variants. This reflects a similar pattern to that reported by Jones (1998) in the Rhosllanerchrugog community in North East Wales. One explanation for this pattern is that it involves a pattern of age grading (Sankoff and Laberge 1978), and this will be discussed below. Given the extent of the effect of place names on other variables however, and the fact that place names very frequently follow (yn), I decided to also examine (yn) with place names removed. As I discussed above, (yn) without place names had a positive significant correlation with age. Something that becomes immediately clear when looking at the apparent time profile is that the pattern is now closer to that of the other mutation variables analysed¹⁹.

¹⁹ The (70+) age group was removed from the apparent time analysis due to a low number of tokens.

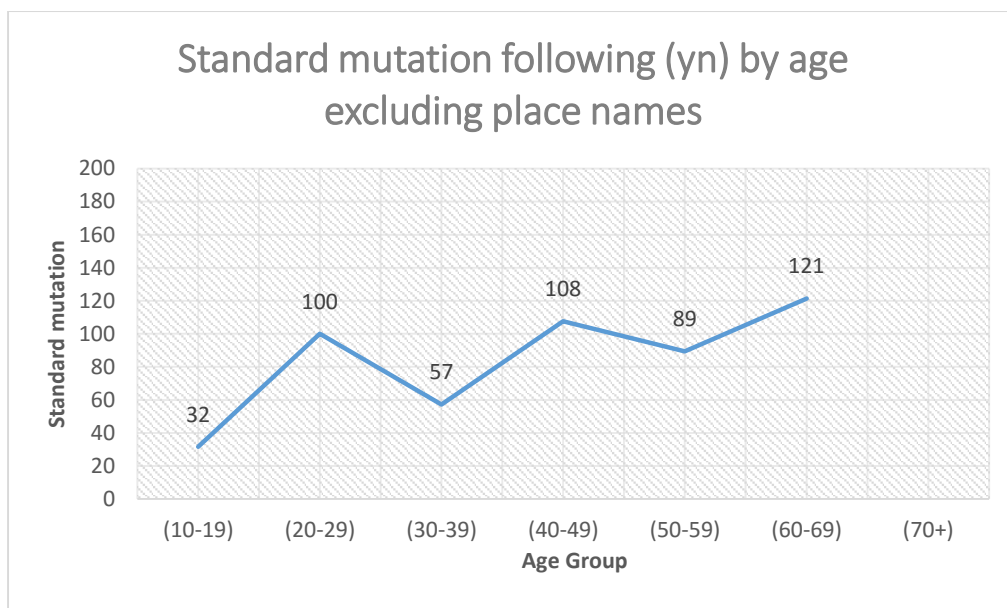


Figure 14 Standard mutation following (yn) for the ‘without place names’ condition by age

	(10-19)	(20-29)	(30-39)	(40-49)	(50-59)	(60-69)	(70+)
Radical tokens	13	6	9	6	8	4	1
SM tokens	3	2	4	2	2	1	0
NM tokens	2	4	3	6	5	8	4

Table 77 Number of SM, NM and radical tokens in *Siarad* corpus following (yn) by age group

Standard mutation of non-place names after (yn) is essentially similar to that of the other mutation variables, with older speakers using significantly more frequent standard mutations than younger speakers. The highest index scores are held by the two oldest age groups, while the lowest index score is held by the youngest age group, with 32 points. This suggests that the unique profile of (yn) may be associated in some way with the place names factor, although I will explore this further in the discussion.

6.6.4.1 Distribution of mutation type following (yn)

Figure 15 below shows the proportion of mutation types applied to target words following (yn).

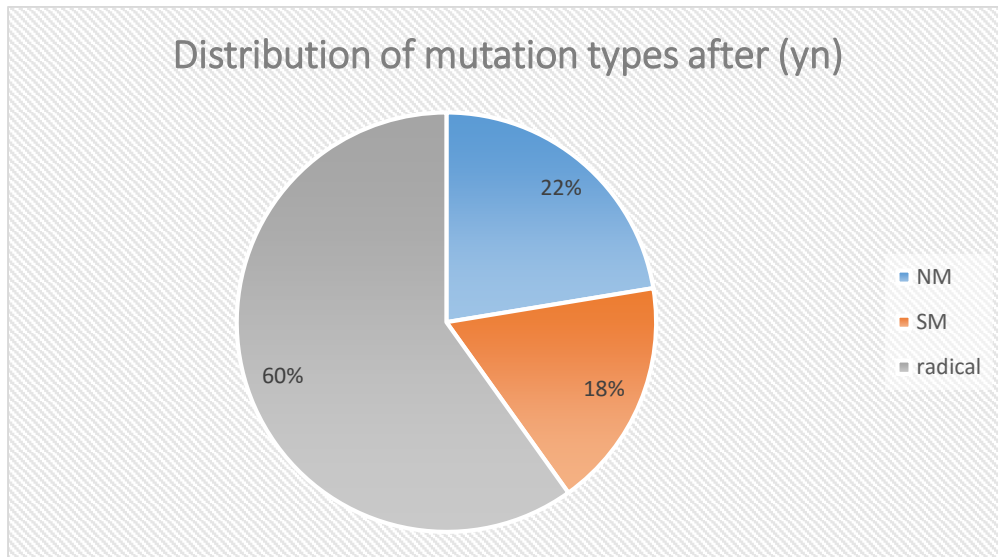


Figure 15 Distribution of mutation types following (yn) in the *Siarad* corpus

The chart shows that the radical is the most frequent variant used, while NM and SM were applied almost equally frequently.

6.6.5 Apparent Time – (a)

Like most of the other mutation variables in this study, (a) exhibits a profile of greater standard use by older than younger speakers, although all age groups have a lower mean standard usage than for the other mutation variables, and no group uses more than 22% standard tokens of (a).

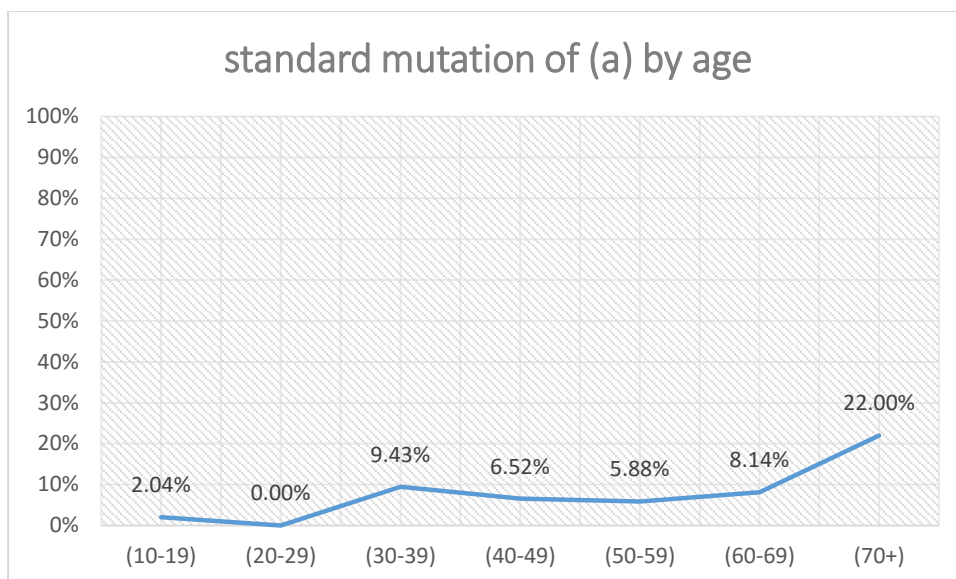


Figure 16 Standard mutation following (a) by age group

	(10-19)	(20-29)	(30-39)	(40-49)	(50-59)	(60-69)	(70+)
Radical tokens	96	100	48	43	80	79	39
AM tokens	2	0	5	3	5	7	11

Table 78 Number of SM tokens and radical tokens in *Siarad* corpus following (a) by age group

In terms of the range of scores, the (70+) group has a relatively high mean standard usage of (a), at 22% which is more than twice that of any other group. The four middle age groups (encompassing 30 to 69 year olds) seem to have comparable means of standard usage at between 5-10%. The youngest two groups show very low standard usage, reaching 0% for the 20-29 group and 2.04% for the youngest group. The profile seems to indicate a general low standard use of this variable, although with older speakers using more standard variants.

6.6.6 *Apparent Time – (ei.m)*

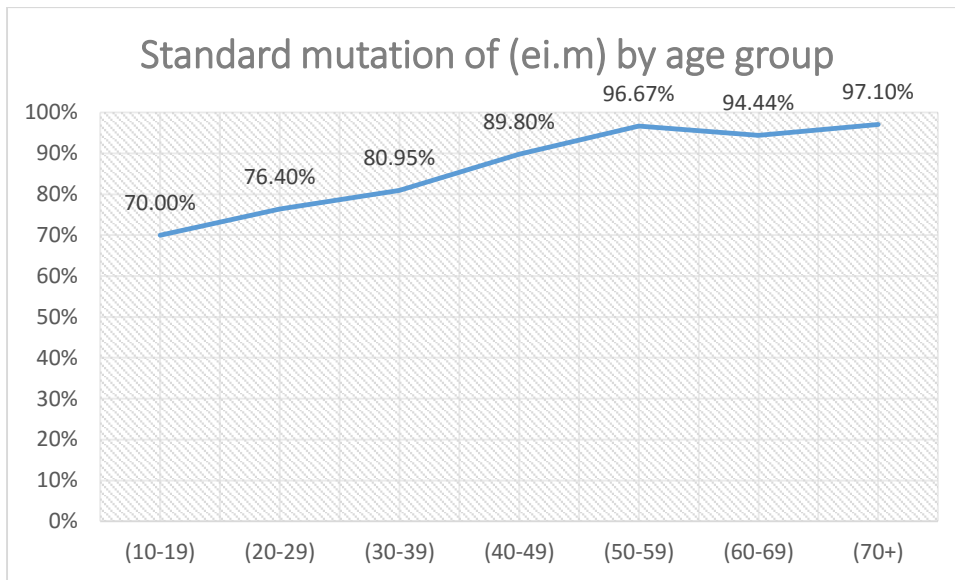


Figure 17 Standard mutation following (ei.m) by age group

	(10-19)	(20-29)	(30-39)	(40-49)	(50-59)	(60-69)	(70+)
Radical tokens	12	21	8	5	4	5	2
SM tokens	28	68	34	44	116	85	67

Table 79 Number of SM tokens and radical tokens in *Siarad* corpus following (ei.m) by age group

Figure 17 shows that standard levels of usage for (ei.m) have their highest point in the oldest age group at 97%, and are relatively consistent until the (50-59) age group after which they begin a gradual decline. The group with the lowest frequency of standard mutation is the youngest, with 70%. The data thus indicates a pattern of generally high usage among the oldest speaker groups approaching 100%, but which begins declining from the (40-49) group onwards.

6.6.7 Apparent Time - (ei.f)

As a variable with three variants in this analysis, standard mutation averages for (ei.f) were calculated using index scores.

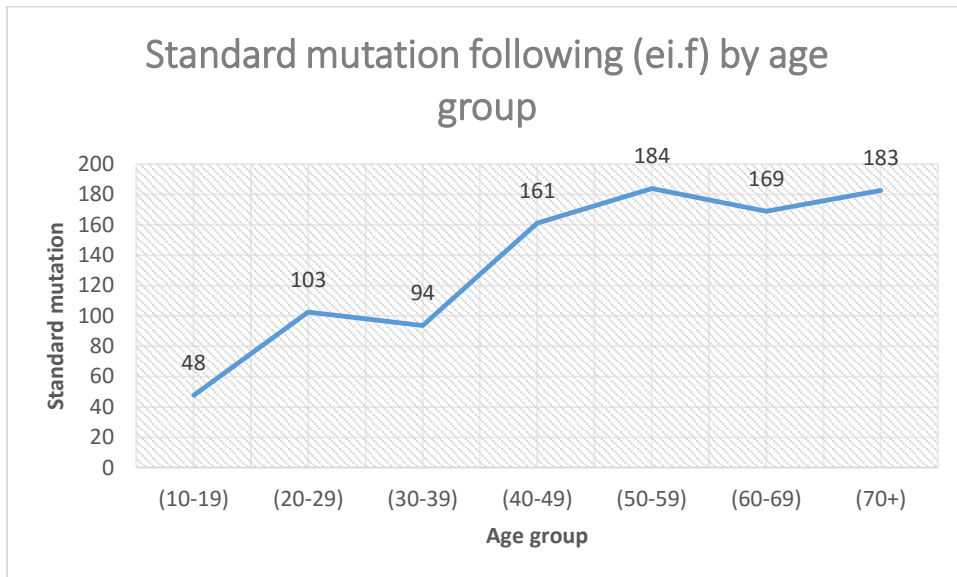


Figure 18 Standard mutation following (ei.f) by age group

	(10-19)	(20-29)	(30-39)	(40-49)	(50-59)	(60-69)	(70+)
Radical tokens	15	18	6	2	1	2	2
SM tokens	2	3	5	3	2	5	1
AM tokens	3	19	5	13	22	22	6

Table 80 Number of SM tokens and radical tokens in *Siarad* corpus following (ei.f) by age group

Figure 18 above illustrates a more or less consistent pattern of decreasing frequency of standard mutation following (ei.f) by age group. The scores of the four oldest age groups are relatively

consistent on a range of 161 to 184 index points, but standard usage declines radically between the (40-49) and (30-39) age groups, with a fall of 67 index points. There is then another large decrease between the (20-29) and (10-19) age groups, with a fall from 103 to 48 index points. The range between the highest index score, belonging to the (50-59) group, and the lowest which belongs to the youngest (10-19) age group, is 136 index points.

6.6.7.1 Distribution of mutation type following (ei.f)

Figure 19 shows the distribution of NM, SM and radical variants following (ei.f) in a pie chart.

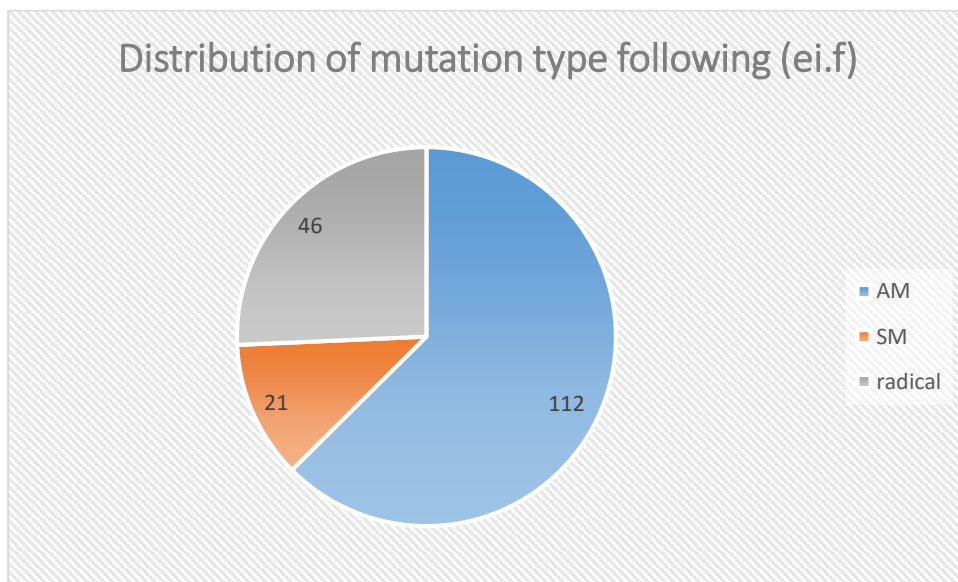


Figure 19 Distribution of mutation type following (ei.f) in the *Siarad* corpus

The standard AM mutation was most frequent, with the radical the second most frequently applied variant. SM was the least frequent variant.

6.6.8 Apparent Time – ($\hat{a}/gyda$)

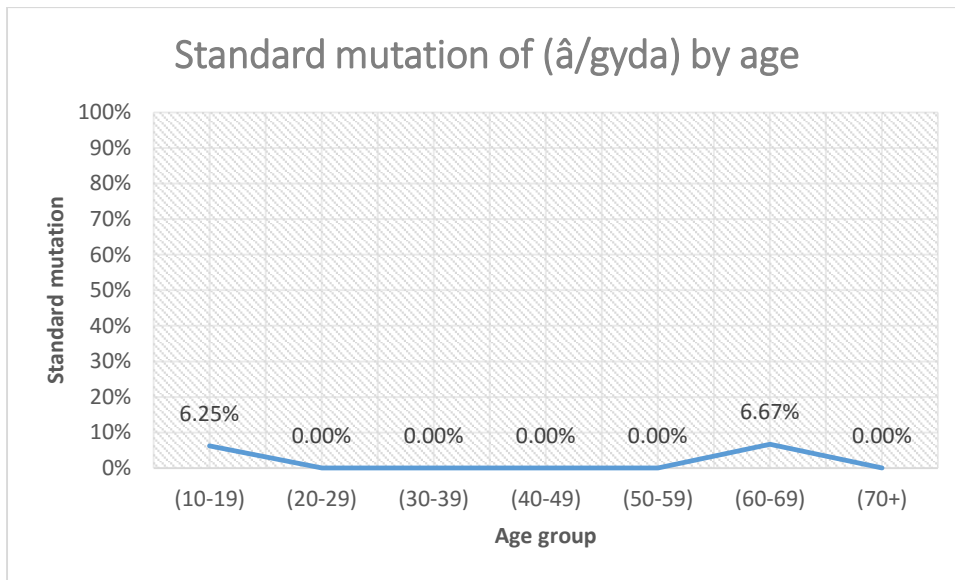


Figure 20 Standard mutation following ($\hat{a}/gyda$) by age group

	(10-19)	(20-29)	(30-39)	(40-49)	(50-59)	(60-69)	(70+)
Radical tokens	15	25	9	11	9	14	8
AM tokens	1	0	0	0	0	1	0

Table 81 Number of SM tokens and radical tokens in *Siarad* corpus following (a) by age group

Standard variants were very infrequent following ($\hat{a}/gyda$), with only two AM tokens. One standard token appeared in the speech of a speaker from the youngest age group, while the other was from the 60-69 year old age group. Given so few standard tokens, no firm conclusions on variation patterns can be drawn from this. It is clear however, that both standard realizations and the use of the mutation trigger itself are comparatively rare in the vernaculars of the *Siarad* corpus participants.

6.6.9 Apparent time – Code-Switching

This section presents the apparent time profile for code-switching, with mean code-switching percentages for each age groups.

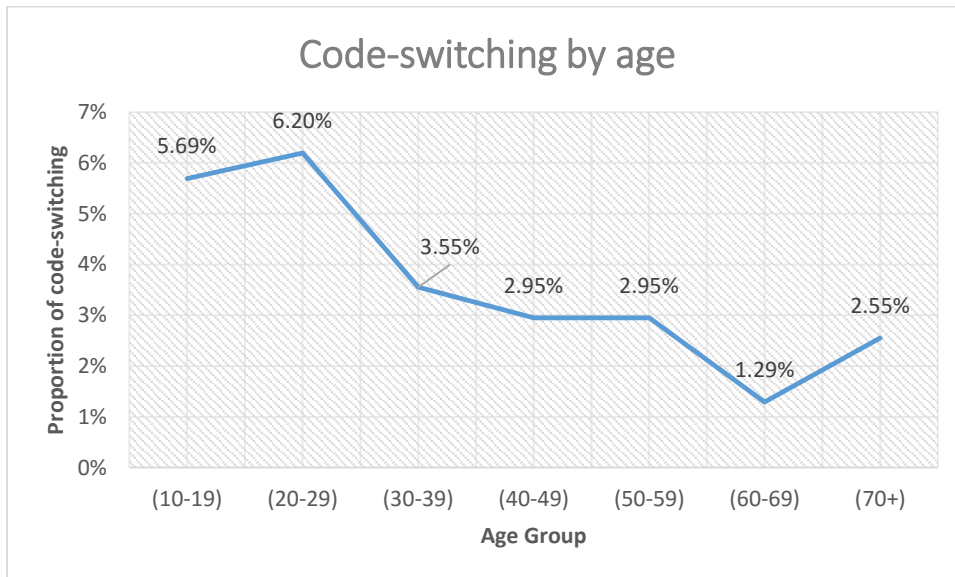


Figure 21 Average proportions of code-switching by age group

If the age variation data is interpreted via apparent time, where synchronic age variation is taken as an indication of diachronic change, figure 21 shows a general decline in the frequency of code-switching between younger and older participants. The second youngest group has the highest mean at 6.20%, which is slightly higher than that of the youngest group. The mean drops substantially between the (20-29) and the (30-39) year old age groups to 3.55%, and then seems to level off for the next two age groups, which both have a mean of 2.95%. There is another decline in frequency for the (60-69) age group, which has the lowest mean percentage in the Corpus at 1.29%. The mean for the oldest group is closer to the middle aged (40-49) and (50-59) age groups with a mean of 2.55%. The apparent time profile confirms the significant correlation between age and code-switching use, showing that younger speakers generally use more code-switching in informal speech than older ones.

6.7 Summary of Results

I summarize the main findings as follows:

- Age was highly significant across all variables, exhibiting a pattern in which older speakers seem to use standard variants at a higher rate than younger speakers, although some variables like (fy) and (yn) exhibit a greater difference between age groups than others like (o) and (am).
- There was variation in the highest standard score for the mutation variables, with (a) in particular holding a very low highest rate of standard mutation of only 22% by the oldest age group, while variables like (fy) and (ei.m) reached nearly 100% standard usage among the oldest age groups.
- Gender was significant for two mutation variables, (fy) and (ei.f), in both cases with men producing more standard tokens than women.
- Education was also significant for the majority of mutation variables, but was not significant for code-switching nor (am).
- The language attitudes analysis found two significant results, the mutation variables (o) and (ei.m).
- Tokens for (â/gyda) were very infrequent in the data, and there were only two standard tokens. This means that no real conclusions could be drawn from its distribution, other than it being very infrequent.

6.8 Discussion of Inter-speaker variation

This analysis was undertaken so that the stylistic analysis on the *Radio Cymru* corpus could proceed with a greater understanding of the social and internal factors operating upon the variables.

In terms of correlation with age, all mutation variables showed a statistically significant decrease in standard use across age groups. To some extent, this has been anticipated in work by other researchers, but it should also be remembered that this data is distinct from most other studies on mutation. Unlike Jones (1998) and Ball (1984) for example, this sample mainly reflects a linguistic community (Gwynedd, North Wales) where Welsh is language with high vitality, spoken by the majority of the population. Previous analyses have concluded that the aspirate and nasal mutations are declining while the soft mutations are basically stable (Ball 1984, Awbery 1986), but my analysis shows a slow decline even in the use of soft mutation, and this in a supposedly linguistically stable Welsh speaking heartland.

6.8.1 The aspirate mutations

The distribution of aspirate mutations (a) and (â/gyda) reflect the trend noted in previous literature (Sweet 1913; Roberts 1988) describing these variables as being infrequent in vernacular styles. The aspirate mutation following (â/gyda) is almost absent in the *Siarad* corpus, with only two tokens of the standard variant. One of these is used by a speaker in the youngest age group, while the other is used by an older speaker; but with so little data, no strong conclusions can be drawn about the relationship between (â/gyda) and age, or any other factor. What does seem clear is that aspirate mutation following (â/gyda) has all but disappeared from the vernacular of Welsh speakers across all age groups. This is relatively unsurprising, given that Fynes-Clinton (1913) remarked that the standard variant was rarely heard more than a century ago.

Aspirate mutation of (a) is more frequent than that of (â/gyda), with most age groups showing at least a few tokens of the standard variant. In the oldest age group (70+), the standard variant occurs 22% of the time. At the other end of the scale however, standard variants are used by only 2.08% the youngest group (10-19), while the (20-29) group used no standard AM variants.

One of the most important questions arising from the analysis of (a) and (â/gyda) is what the generally low frequency of both types of aspirate mutation in informal speech connotes for the natural acquisition of these variables. It seems unlikely that young Welsh speakers could acquire (a) without either being explicitly taught it, or having substantial opportunity to read and write through the medium of Welsh. Although a case could be made that younger speakers might be able to acquire the standard variant through frequent face to face contact with speakers from the very oldest age group, the frequency of use among that group may still be too low for this to occur. Even more so than (a), the aspirate mutation following (â/gyda) is rare enough in informal speech that children acquiring Welsh would almost never receive it in their home language input. This suggests that exposure to standard Welsh, either through reading of Welsh texts (which, according to A.R. Thomas (1982) and Rhys and Thomas (2013) is apparently rare amongst Welsh speakers), or through participation in a domain where formal Welsh is used (see section 1.1.4 of this thesis) is a prerequisite for the acquisition of AM mutations.

6.8.2 The distribution of (ei.m)

Use of the standard variant with (ei.m) is very frequent (above 90%) until the (40-49) age group, at which point standard usage begins a gradual decline, with the youngest age group applying the standard SM mutation only 70% of the time. The range in between the oldest group's standard use of (ei.m) and the youngest is 27%. This suggests that despite being a mutation which carries an important functional load (Ball and Müller 1992), standard usage of (ei.m) has still declined. Part of this can be attributed to broad use of the pattern that Jones (1998) noted in her study, where speakers tend to drop the initial mutation trigger *ei* and the SM or AM mutation to mark gender, instead using the post nominal or verbal pronoun *hi* or *fo* to signal the gender of the possessor: e.g. *cath fo* ("his cat") rather than *ei gath*. My analysis has found a similar pattern, which is increasingly widespread among younger speakers. Jones (1998) provided an effective explanation for this phenomenon as *compensation with loss*

(Trudgill 1983), where speakers are able to balance the loss of one feature with the introduction of another one. This suggests that the reason for the decline of this mutation is that (ei.m) no longer necessarily carries a functional load. Despite this, the use of (ei.m) across all age groups is still relatively high compared to NM and AM variables.

6.8.3 The distribution of (ei.f)

The aspirate mutation trigger (ei.f) shows a large decrease in mean standard use from 183 index points among the (70+) age group to 43 index points by the (10-19) age group, a range of 163 points. As with (ei.m) an argument can be made that this decline could be attributed to the loss of functional load of the variable due to the *compensation with loss* strategy, as described by Jones (1998). This means that younger speakers are able to avoid the difficult AM mutation by using the object pronoun *hi* to signal the gender of the possessor instead. The fact that (ei.f) has declined so much further than (ei.m) could perhaps be attributed to the general trend not to use aspirate mutations in the vernacular, although it should be noted that AM is still used with (ei.f) substantially more frequently than with the other two AM variables, (a) and (â/gyda).

Another factor affecting (ei.f) is the tendency for many speakers to apply SM to (ei.f) environments, previously reported by Thomas and Gathercole (2007), which implies that speakers often may not be aware of the historically appropriate gender for many possessors. This provides more evidence for the supposition that the gender system is being superseded by the compensation with loss strategy (Jones 1998). As in Thomas and Gathercole's (2007) work, I have found that speakers in my data seemed more likely to apply the appropriate gender based mutations in the case of human possessors, where the gender is arguably more immediately obvious (i.e. AM for female persons and SM for male persons), while in the case of non-human possessor, many speakers seemed to indiscriminately use the post-nominal pronoun *o* or *fo*, even where the possessors were feminine. As Thomas and Gathercole (2007) have argued, this

suggests that speakers' grasp of the Welsh gender system is declining particularly strongly in the case of inanimate possessors. This may be unsurprising if other cues to the gender of common nouns, such as appropriate AM mutation applied in (ei.f) contexts, are becoming rarer in the community.

6.8.4 (o) and (am)

In contrast with the NM and AM variables, the soft mutation variables seem relatively stable and standard forms frequent in the data. This reflects the findings of previous research which has presented SM as being a stable system, even expanding into other mutations' domains (Ball and Müller 1992), and to some extent this is attested here. The system's relative robustness could be attributed to the wide range of triggers environments which can cause SM, as well as the relatively high number of consonants which are able to undergo soft mutation (Thomas 2001). But while the fall in standard usage across these variables is much smaller and less rapid than that which has taken place with the nasal and aspirate mutations, a significant positive correlation was found between all of these variables and age. In other words, although standard forms are frequently used, there is still a tangible decline in process.

This reflects Jones' (1998) finding that the application of SM to the preposition triggers is undergoing gradual decline across all age groups, in contrast to other variables which show stability in older groups, but decline in school age children. Another noteworthy aspect of these variables in my data was that they seemed to be significantly affected by the place name factor, with place names less likely to undergo standard mutation than non-place names. This echoes the findings of Ball (1984). The place names factor seems to have had a greater effect on older than young speakers, with speakers' standard use of both variables substantially higher among older speakers when place names were removed from calculation, while younger speakers' means were generally less affected.

6.8.5 The distribution of (fy)

Like the aspirate mutations, the nasal mutation of (fy) is declining in use. This variable is among those which show the most rapid change, with speakers under 20 using the standard variant only 15.7% of the time on average, while the 70+ age group use the standard variant 96.55% of the time – a decrease in standard usage of 80.85%. The most significant shift seems to have taken place between the 10-19 and the 20-29 age groups, with a drop of 37.82%. This broadly corresponds with Jones' (1998) findings on the use of (fy) in North East and South East Wales, where she found that school age children were far less likely to use the variable than those who were in their twenties and over. Jones also argued that the surge in L2 speakers of Welsh may be influencing adoption of the form by L1 speakers, and this is one explanation for the rapidity and scale of the change in this age group.

The findings contrast with Jones' however, in that there is also a clear gradual decline in the use of the standard variant of (fy) among older age groups in the community. While the standard use of (fy) is stable in the mid-90% for the (70+) and (60-69) age groups, there is a gradual decline from the middle aged (50-59) age group onwards, followed by two much more striking falls in standard use between the (30-39) and (20-29) age groups, and then the (20-29) and (10-19) age groups. Thus, although it is possible that school related factors have played a factor in this decline - particularly given the magnitude of the fall in use between the (20-29) and (10-19) age groups – it may be that the process of change has been going on for some time, possibly starting with the (50-59) age group.

In most cases examined in the data, the use of the nonstandard radical variant with (fy) coincided with the use of the (noun) + fi construction, while the use of the nasal mutation coincided with either fy + (noun), yn + (noun) or with constructions which lacked any pre- or post- nominal pronoun (i.e. simply *ng hath* “my cat”). Thus, as with the dropping of pre-

nominal and pre-verbal pronouns with (ei.m) and (ei.f), young speakers' movement towards the novel (noun) + fi construction of (fy) seems to be connected with a tendency to use the radical rather than to apply the nasal mutation.

6.8.6 The distribution of (yn)

The locative preposition (yn) is perhaps the most unique in terms of its distribution. Rather than showing a pattern of steady decline in standard use, the standard form of (yn) is used most frequently by the 30-39 age group, with older and younger speakers showing comparably low rates. There are multiple possible explanation for this trend. Jones (1998) found a similar distribution of (yn) in her analysis of a village in the North Eastern village of Rhosllanerchrugog. Her explanation for the pattern was that the use of the radical following (yn) was actually the established vernacular form in the community. The older generation were thus more likely to use this traditional form, while middle aged speakers were more likely to have been influenced by Welsh-medium education and the standard Welsh norm of mutating after (yn). This is clearly a very plausible explanation, particularly as Rhosllanerchrugog is the nearest community to that represented in the *Siarad* corpus, both geographically and chronologically. Other studies like Hatton (1988) have also claimed that the nasal mutation following (yn) is not in fact the traditional variant in the vernacular, and that is a form that must be acquired in school. It could be argued then, that the *Siarad* corpus data reflects the same pattern, with older speakers using the radical as a dialectal variant, while younger speakers are showing the beginning of a trend of rule loss, as seems to be the case with other mutation variables.

An alternative, but related, interpretation of the data is that it is an example of age grading. One of the typical manifestations of age grading involves professional variants, which speakers begin using substantially more frequently in response to the pressures of work (Sankoff and

Laberge 1978). As they get older the same speakers tend to cease using the standard variables as these same pressures recede. Age grading also typically involves variables which are particularly salient for speakers, and we have at least some evidence for this from Ball (1985), who showed that (yn) was one of the variables of which speakers were most consciously aware. One problem with this theory is that the association of professional domains with Welsh is relatively recent, meaning that older speakers are not likely to have ever participated in the age grading process. Perhaps the most convincing explanation is a mix of both theories – middle aged speakers are influenced by Welsh medium schooling as well as professional pressures for those who may work through the medium of Welsh (for instance, in a local council or for the BBC) while older speakers came of age in the period before Welsh was used in such domains. There is another issue which complicates the picture further, however. One finding of the analysis of mutation seems to be the extent to which the use of mutation variables is governed by internal factors. This study replicated the finding of others (Ball 1984, P.W. Thomas 1984) who found that place names were particularly resistant to standard mutation. With (yn) however, which was predominantly followed by place names in the data set, speakers seem to follow a pattern of mutating place names according to their frequency. For example, *Cymru* (“Wales”), was the most frequently occurring place name, and was also mutated standardly far more frequently than less frequent place names, such as those only mentioned once or twice, like *Cibyn* or *Cerrigydrudion*.

Why only place names seem to have been affected by this phenomenon in my analysis (as far as I can currently see) remains unclear, but it may be related to the fact that the frequent use of certain place names following (yn) could become highly conventionalized collocations, and become easier for speakers to retrieve than low frequency place names. Another aspect that may have affected (yn) with regards to place names is that certain place names have historically attested forms which are lenited in comparison with the modern standard, such as *Caernarfon*,

which has the historical form *Y Gaernarfon* (GPC 2015). This may have artificially raised the index score for some place names, and Caernarfon did have a particularly high frequency of intermediate SM forms. However this does not explain the high incidence of standard mutation for other very frequent place names, such as *Cymru* (“Wales”), where I believe frequency, and some kind of process of entrenchment, is the most convincing explanation.

6.8.7 Level of education and gender

Although age was the factor which seemed to have the widest reaching effect upon the data, some of the other social factors were also significant. In terms of level of education, all mutation variables except for (am) were significantly correlated. This would seem to provide further evidence for the conception that the Welsh education system acts as a bulwark in favour of the reinforcement and dissemination of standard Welsh forms. Another possibility is that education here is reflecting class differences, as education is one of the factors typically used to determine participants’ socioeconomic class. However, it is not the only factor, and is not sufficient on its own to function as an index of class. Ultimately, further research should be carried out to clarify the matter.

In terms of gender, only two variables were significant, (fy) and (ei.f). In both cases men were more likely to use standard forms than women, which echoes Ball’s (1984) findings. Ball explained his data in terms of different degrees of exposure of male and female participants to prestigious forms in the community. Such an interpretation seems less convincing for my more modern data however, as it is unlikely that women now receive less exposure to standard Welsh than men. An alternative explanation might draw on the *gender paradox* (Labov 2001.b), and interpret the findings as reflecting female led change from below, with women more likely to adopt the non-standard innovating variants.

6.8.8 Inter-speaker variation - summary

The analysis of the *Siarad* corpus data was intended to aid in performing an accurate analysis of style in the *Radio Cymru* corpus data. What findings then, need to be carried over into the stylistic analysis?

- All standard forms of mutations are declining in the vernacular (although see below for (yn)), with younger speakers using less standard forms than older speakers. This suggests (Bell 1984) that informal situations will generally favour non-standard forms (i.e. non-mutation).
- Conversely, some mutations are declining at a faster rate than others, leading to a large gulf between generations for certain variables:
 - (fy) is strongly differentiated between generations, with a drop in the use of standard forms of 87% between the youngest and oldest generations. Similarly, (ei.f) has seen a large decline of 182 index points between the youngest and oldest age groups.
 - (o), (am) and (ei.m) are more stable, with decreases of 18%, 28% and 22% respectively.
 - The AM variables (â/gyda) and (a) both have the lowest standard use in the corpus. With only two standard tokens (and arguably one of those is a fossilized form) (â/gyda) is very infrequent in informal speech, while (a) exhibits a pattern which shows it declining from 22% in the oldest age group to 2.08% in the youngest group, and 0% in the second youngest group.
- I have argued that (yn) is affected by internal factors, namely *place name* and *frequency of place name*, although social factors like age grading may also be involved. With

place names included, (yn) exhibits an age graded type curve, while without place names it resembles the declining profile of the other mutation variables.

7. Radio Cymru Corpus Methodology

The aim of this chapter is to describe the collection and analysis of a small corpus (or rather corpora) of Welsh language radio programmes, which I will be referring to as the *Radio Cymru* corpus. The chapter will also introduce the programmes which were selected, explain the rationale behind their selection and briefly outline the main differences in format and structure between them. The programmes, which were selected to represent a range of formality, are *Post Prynawn* (7.1.1), *Dewi Llwyd Ar Fore Sul* (7.1.2), *Geth a Ger* (7.1.3) and *Tudur Owen* (7.1.4). Section (7.2) outlines the process of recording, converting and organizing the audio recordings for each programme, as well as describing the relative size of each corpus in terms of audio length and word count. The transcription process, which was deliberately undertaken in such a way to render it comparable with the *Siarad* corpus, is described in section (7.3). The process of calculating scores for mutation and code-switching scores are discussed in section (7.4). Section (7.5.1) describes the methodology used to rank each programme along a stylistic axis, indicating the relative degree of formality in each one using frequency of laughter and overlapping speech as quantitative measures, while section (7.5.3) discusses the ranking of programmes along the social axis, which is intended to represent the relative likelihood that participants in each programme use standard forms in their vernacular. Taking this approach will allow a clearer interpretation of the differences between variables, since variables that are not markers may be expected to correlate more strongly with the social axis, while variables which are markers should correspond more strongly with the style axis. The final section (7.6) will summarize the above, before moving on to the results chapter.

7.1 Selecting programmes

Programmes were selected bearing certain necessary criteria in mind. In order to control as much as possible for differences in gender and dialect, it was decided that the hosts of all programmes should be both male and from the same dialect area of North West Wales.

Crucially, programmes needed also to represent broad differences in level of formality. The four programmes were *Rhaglen Tudur Owen* (“Tudur Owen’s Programme”), *Rhaglen Geth a Ger* (“Geth and Ger’s Programme”), *Dewi Llwyd Ar Fore Sul* (“Dewi Llwyd on Sunday Morning”) and *Post Prynawn* (“Afternoon Post”). All four programmes differ in important ways, and I will be arguing that the continuum of formality that I have constructed for them is actually shaped by several factors, including the format, structure and ideology underpinning each one.

7.1.1 Post Prynawn

Post Prynawn is a daily, hour long programme mainly structured by news and interviews. The host is the television veteran Dewi Llwyd, who has previously worked on Welsh medium television for a lengthy period as a newsreader and narrator for news programmes. Typically, each programme involves alternation between brief sections where Dewi Llwyd address listeners directly, and short interviews with a mixture of established correspondents (i.e. other newsreaders and researchers working for the BBC), regular guests (including professionals who the BBC frequently consult for expertise on specific subjects, such as medicine or the print media) and members of the public, for example a local farmer for his opinion on a story about agriculture. Conversations on *Post Prynawn* never involve more than two speakers, and typically involve long stretches of well-formed utterances by both speakers with very few instances of interruptions or overlapping speech. The following example of an interview between Dewi Llwyd and a guest illustrates this pattern:

*DEW: ond beth ydy arwyddocâd y ffigyrau felly ?

%eng: *but what is the significance of the numbers then?*

*DEW: gan fod y wybodaeth yn cael ei chasglu siawns nad oes yna ryw fath o rybudd yma fod ym mhell dros hanner ein ysbytau ni mae graddfau marwolaethau yn waeth nac y dylai hi fod ?

%eng: *given that the information is being collected, there must be some kind of warning here given that over half of our hospitals have worse death rates than they should have?*

*DOC: well@s:und efallai fod hynny am efallai yng Nghymru cadw mwy o gleifion yn ysbytai a fod mwy o gleifion yn [/] yn gweld diwedd eu dydd mewn ysbyty

%eng: *well maybe that's because in Wales we keep more patients at the hospital, and more patients see the end of their days in a hospital*

*DOC: er@s:und mewn llefydd eraill efallai er@s:und fydden nhw yn mynd i gartrefi nyrsio neu yn sefyll adref .

%eng: *perhaps in other places they would go to nursing homes or they would stay at home*

Extract 38 Example of a typical exchange in *Post Prynawn*

The pitch of speakers is noticeably more consistent and monotonous than in natural conversation. Perhaps befitting the formal nature of the programme, laughter occurs only very occasionally, and usually when Llwyd is talking with established correspondents with whom he is familiar.

7.1.2 Dewi Llwyd Ar Fore Sul

Dewi Llwyd Ar Fore Sul is also an hour long programme that is broadcast each Sunday morning, again hosted by Dewi Llwyd. The programme has a more varied structure than *Post Prynghawn*, with much of the first section taken up by a relatively informal magazine style discussion of the news headlines. The news discussion is hosted by Llwyd, and involves two other interlocutors, who are typically eloquent members of Welsh public life. Although they aren't BBC staff, they do normally represent prominent Welsh medium institutions, and perhaps cannot be said to represent the general public. In one episode for example, the guests were the then head of The Welsh Language Society (a pressure group for Welsh language rights) and a Welsh-medium lecturer in Sociology at Bangor University. The next section involves a 'birthday interview' with a different guest each week who is celebrating their birthday. The guests are typically well known people within the Welsh speaking community, such as artists, actors and academics.

To give an impression of the typical guests, in the three episodes transcribed the guests were a sculptor turned art historian, the current archdruid of Wales and a respected (but now retired) professor of Welsh history. The birthday interview involves a relatively informal interview between Dewi Llwyd and the guest, with topics ranging from the guests' achievements to their family life, hobbies and other light topics. The final section of the programme involves a discussion between Dewi and an arts correspondent about various art exhibitions held around the country. Although this section is slightly more formal than others in this programme, there is still noticeably more levity than in the interviews that characterise *Post Prynghawn*:

- *MEN: um@s:und ond yeah@s:und mae (y)n stori erchyll â gweud y gwir .
- %eng: *but yeah it's an awful story to tell the truth*
- *MEN: a mae e (y)n atgoffa ni o straeon o um@s:und well@s:und dechrau yr ganrif ddiwethaf o bobl yn mynd ati i [/] i drin eu hunain .
- %eng: *and it reminds us of stories from well, the beginning of the century, of people going to treat themselves*
- *MEN: um@s:und +/.
- *DEW: yeah@s:und math o stori mae Geraint yn ei chofio dw i (y)n sure@s:und .
- %eng: *yeah, the kind of story that Geraint remembers I'm sure*
- *MEN: &=laugh oh@s:und ie .
- %eng: *oh yeah!*
- *GER: &=laugh well@s:und um@s:und odw Dewi odw .
- %eng: *well yes Dewi, I do!*
- *MEN: &=laugh .
- *DEW: +< &=laugh o'n i meddwl .
- %eng: *I thought so!*

Extract 39 A typical interaction from Dewi Llwyd AFS 1

Extract 39 above provides an example of a typical exchange from the newspaper discussion segment of Dewi Llwyd AFS 1, with a generally more involved conversation between the three participants than is typical of *Post Prynawn*, including more laughter, interruptions and overlapping speech. Also in contrast with *Post Prynawn*, which has no soundtrack, *Dewi*

Llwyd AFS has frequent musical interludes typically playing classical music and Welsh language folk songs.

7.1.3 Rhaglen Geth a Ger

Geth a Ger is an hour-long chat programme aimed at young people, with subversive humorous themes. The programme has two presenters: Gethin Evans and Geraint Jones, both young men in their early thirties and late twenties respectively. The programme mainly involves casual conversations between the presenters interspersed with contemporary music, including rock, hip hop and pop in both Welsh and English. In contrast to most of the other programmes in the study, *Geth a Ger* only ever has two contributors - the presenters Gethin Evans and Geraint Jones - taking part in the programmes (although the producer is often referred to as part of an ongoing in joke about his unseen presence). The low number of participants gives the programme the relaxed feel of two friends chatting at the pub, which contrasts with the more boisterous and anarchic feel of *Rhaglen Tudur Owen*, with its multiple contributors. There are some recurring sections in each episode where the tone arguably varies somewhat. One of these is *ffeithiau'r wythnos efo Ger* (facts of the week with Ger), which involves Geraint Jones providing 'facts' on themes that vary with every week. This section might be considered slightly more formal, with fewer overlaps and a more didactic tone than usual. The following extract is from a 'facts of the week with Ger' segment where the presenters are discussing conspiracy theories about the sinking of the Titanic, which gives a flavour of the type of discourse typical of the programme:

*GET: &=laugh i fod yn deg &=laugh y Titanic@s:und gath o yn anghywir de .

%eng: *to be fair, the Titanic got it wrong*

*GER: &=laugh .

*GER: +< oh@s:und Titanic gath o (y)n xxx .

%eng: oh the Titanic got it...

**GER: ia actually@s:en actually@s:en .*

%eng: yeah actually

**GET: +< ia i fod yn deg .*

%eng: yeah to be fair

**GER: oedd o un peth gath o mor bell oddi wrthi de .*

%eng: it was one thing that he got so wrong...

**GET: ia .*

%eng: yeah

**GER: wnaeth o ddeud yn ei lyfr bod y Titan@s:und (we)di hitio y rhew pedwar cant milltir oddi wrth Newfoundland@s:und tra oedd o yn wneud twenty_five@s:en knots@s:und*

%eng: he said in his book that the Titan had hit the ice four hundred miles from Newfoundland while it was doing twenty five knots

**GET: o faint ?*

%eng: how much?

**GER: ond (.) roedd hynna (y)n hollol anghywir .*

%eng: but that was completely wrong

**GET: &=laugh .*

Extract 40 A typical exchange taken from *Geth a Ger*

As can be seen in this extract, the discourse style is far from the well-formed utterances of Post Prynhawn, with half formed, unfinished constructions, interruptions and frequent laughter.

7.1.4 Rhaglen Tudur Owen

Rhaglen Tudur Owen is an institution in Wales. The presenter is a popular comedian who regularly appears on television and other media. Tudur is joined by a rotating cast of regular co-hosts, who usually include Dylan Meirion (a prominent Welsh language DJ and musician), Gareth Jones (the producer) and Manon Rogers. The programme typically involves a very dynamic feel, with a large amount of laughter and overlapping speech. This may be partly due to the fact there can be up to six speakers talking at any one time in the programme (taking into account the four regular presenters, as well as extra guests). Speakers often have to struggle to get a word in, and generally interrupt and talk across each other almost continuously. Additionally, and perhaps in contrast with *Rhaglen Geth a Ger*, *Tudur Owen* contains a larger range of emotional interactional markers than in any other programme. These include screams, shouts, silly voices, imitations and whispering. The programme also usually involves a segment on sporting news with celebrity sports correspondent Eleri Shon. Although ostensibly a straight interview between Tudur Owen and Eleri Sion about sporting news and events, this segment often breaks down in laughter and also involves interjections and interruptions by the other co-presenters. The extract below gives an example of a typical interaction:

*DYL: mae koala@s:und bears@s:en yn cario er@s:und chlamydia@s:und .

%eng: *koala bears carry chlamydia*

*ELE: +< xxx pam ?

%eng: [...] *why?*

*ELE: ah@s:und .

*TUD: ah@s:und ok@s:und .

*DYL: ffaith y diwrnod i chi .

%eng: *fact of the day for you*

*ELE: +< be xxx .

%eng: *what [...]*

*TUD: diolch yn fawr Dyl@s:und .

%eng: *thanks a lot Dyl*

*TUD: wnawn ni symud ymlaen dw i (y)n meddwl o hynna .

%eng: *we'll move on from that I think*

*ELE: +< o_k@s:und .

*GAR: yeah@s:und .

*DYL: &=laugh .

*ELE: paid â mynd yn agos iddyn nhw Dyl@s:und .

%eng: *don't go near them Dyl*

*GAR: &=laugh .

*TUD: &=laugh .

*DYL: &=laugh .

*GAR: um@s:und anyway@s:en +...

*TUD: +< da iawn does (y)na ddim koala@s:und bears@s:en yn Gaernarfon diolch
byth am hynny .

%eng: it's a good thing that there aren't any koala bears in Caernarfon - thank god for that

Extract 41 A humorous extract reflecting the typical style of *Tudur Owen*

As extract 41 suggests, the tone of the programme is often intentionally subversive, with humour that arguably pushes the boundaries of taboo subjects including sexual innuendo and controversial political jokes. Most relevant of all is the subversive humour aimed at linguistic purism among Welsh speakers, with frequent segments directly mocking standard language ideology. One example of this is the *Gloywi Iaith* segment, where the presenters mock the practice of taking ‘language improvement’ lessons by already fluent speakers with an ironic ‘lesson’ which presents ostensibly English words in the place of the expected advanced Welsh vocabulary.

7.1.5 Summary

The programmes described above were each selected to cover a broad variety of styles, and of levels of intensity and formality. Table 82 below provides an overview of some of the features that differentiate the programmes:

Programme	Host	Music Type	Maximum number of simultaneous speakers	Overall 'feel'
Post Prynawn	Dewi Llwyd	No Music	2	News programme
Dewi Llwyd Ar Fore Sul	Dewi Llwyd	Classic, folk, country (Welsh medium)	3	Light magazine programme
Rhaglen Geth a Ger	Gethin Evans	Rock, hip hop, pop (Welsh and English medium)	2	Pub chat
Rhaglen Tudur Owen	Tudur Owen	Rock, hip hop, pop (Welsh and English medium)	6	Comedy

Table 82 Summary of the four *Radio Cymru* programmes

The next section describes the way in which these four programmes were collected, organized and transcribed, before going on to detail the method used to categorize each programme in order of formality using quantitative measures.

7.2 Data collection

Data was collected using a software package capable of recording radio programmes directly from the internet. The material was recorded from the BBC's own IPlayer website, where all of *Radio Cymru*'s programmes are made available for a period of a week following their initial broadcast. Once downloaded, audio files were converted from MP3s into the WAV file format with the free audio editing software package *Audacity* in order to make them compatible with the CLAN transcription programme. I then edited each file to remove musical interludes, which were frequent in the more informal programmes, and any obviously scripted segments in the programmes which only involved a single speaker – for example news bulletins and reports.

My main aim in terms of corpus composition was to transcribe a representative cross section of each programme. I also wanted to ensure a sufficient number of tokens of each mutation variable for each programme. I decided upon a target of three distinct 'episodes' of each programme, which were each broadcast at around intervals of one month apart. This ensures that each episode may be considered relatively representative of the institutional style of each programme, and that the data is not overly skewed by a single unrepresentative 'off day', but it also provides coherent chunks of transcripts that can be submitted to a qualitative discourse analysis, in order to complement the quantitative work. Each individual episode varied in length from one hour to two hours each, but following the editing process the sum length of each programme averaged at around three hours. It should be noted that the length of each programme did not necessarily correspond with the final word count of each individual corpus. Although *Geth a Ger* was the shortest programme in terms of length post-editing, it still had a larger word count than *Post Prynawn*, which was almost an hour longer. This is largely due to paralinguistic differences between the programmes – in particular the higher speech rate and greater overlapping speech in *Geth a Ger* compared to the more measured format of *Post Prynawn*.

The variance in number of words does not seem to have affected the number of mutation tokens, as most of the variables were relatively well represented across all programmes. There was a problem with certain variables however, with (â/gyda) being particularly infrequent, reflecting its distribution in the *Siarad* corpus analysis. Interestingly, (â/gyda) was far more frequent in the two most formal programmes, with *Post Prynghawn* having many more tokens than *Tudur Owen*, despite the latter having a larger word count. This implies that simply recording and transcribing more material for the two more informal programmes may not have worked to increase the number of tokens of (â/gyda) significantly, since its presence may be intrinsically connected with variation in register. Another variable which was similarly affected was (fy), which was very frequent in *Dewi Llwyd AFS*, *Geth a Ger* and *Tudur Owen*, but almost absent in *Post Prynghawn*. These particular variables and their distribution will be discussed in more depth in the next chapter.

Corpus	Total number of words	Total length of recordings
Post Prynghawn	25866	02:59:59
Dewi Llwyd AFS	36167	03:37:15
Geth a Ger	28358	2:15:24
Tudur Owen	35092	02:57:37
TOTAL	125483	11:50:15

Table 83 The number of words and length of recordings (h/m/s) of each programme post-editing

The audio files and the relevant transcript for each program were labelled numerically according to the order in which the programme was recorded. So for example, for *Tudur Owen* the transcripts and audio files are labelled Tudur_Owen1, Tudur_Owen2 and Tudur_Owen3.

7.3 Transcription

One of the main aims in the collection and transcription of the *Radio Cymru* data was to create a corpus which could be directly compared with the *Siarad* Corpus. As such, the radio programmes were purposefully transcribed in a manner intended to correspond as much as possible with the way in which the *Siarad* corpus was transcribed. This means that I relied upon the *Siarad* corpus transcription manual (Deuchar 2014) for guidelines on how to transcribe and code the corpus, as well as calling upon my own experience working on the *Siarad* corpus' sister corpus, the *Patagonia* corpus, which was transcribed using the same techniques and guidelines.

7.3.1 Tier structure

As in the *Siarad* corpus, I transcribed a main tier as well as an English free translation tier, but I did not add a gloss tier. I added the translation tier in order to help identify dropped tokens (as discussed in section 5.4.1), as well as to add value to the corpus in the event that I choose to make it available for public access. I did not add a gloss tier, as I felt that it would not be necessary for the methodology used at this juncture, although I would not rule out making an automatic gloss tier available in the future, for use in further kinds of analysis of the same data set.

7.3.2 Mutation

The identification and coding of mutation in *Radio Cymru* corpus analysis uses essentially the same methodology as that used in the *Siarad* corpus analysis described in length in sections (5.3) to (5.8.8). Some of the lessons learned during the development of the *Siarad* Corpus mutation analysis, turned out to be relevant for the *Radio Cymru* corpus analysis. For example, the exception list from chapter 5 was also used here to tag and remove ambiguous tokens, such as 'fossilized constructions' and words tagged @s:und and @s:en. Another useful finding that

carries over from the *Siarad* analysis was the finding that place names behave very differently to non-place names in mutation environments. As I was aware of this, I was able to anticipate the skewing effect of this factor, and remove place names from the analysis of variables where this was feasible.

For (o), there were sufficient tokens to calculate scores in two conditions, ‘with place names’ and ‘without place names’. For (am) and (yn) however, there were not enough tokens to do this. (am) was simply calculated in the ‘without place names’ condition, with all place names excluded due to relatively low number of place names as target words for this variable. In the case of (yn) however, the opposite pattern was true, with almost all target words being place names. This is relatively problematic, particularly since I found a strong effect from the frequency of place names on the likelihood of standard variants being produced. As this effect is difficult to control, there was not much which could be done except to keep in mind the possible influence of internal factors on the data.

7.3.3 Transcribing Dropped Mutation Triggers

One innovation, relative to the way in which the *Siarad* corpus was transcribed, is the transcription of dropped variable triggers. As was described in section (5.4.1), finding tokens for certain mutation triggers was particularly difficult where speakers tended to drop those triggers, particularly in the case of the variables (fy), (ei.m) and (ei.f). While in the *Siarad* corpus analysis I had to rely on the translation tier to help me find dropped triggers, for the *Radio Cymru* corpus I decided to incorporate them in the transcription process so that these elements could be easily searchable after transcription was complete. Thus, in utterances where I inferred that a speaker had dropped a mutation trigger, I transcribed the dropped element between parentheses. This is demonstrated in in the example below:

*GER: o'n i (y)n meddwl mai ddy nes oedd yn (fy) cyfweld fi oedd (we)di (fy) galw fi yn hynna

%gls: be.1S.IMP PRON.1S PRT think.NONFIN PRT woman be.3S.IMP PRT (PRON.1S) interview.NONFIN PRON.1S be.3S.IMP PAST.PRT (PRON.1S) call.NONFIN PRON.1S PRT that

%eng: *I thought that it was the woman who interviewed me who had called me that*

Extract 42 An example of the dropped mutation trigger (fy) being transcribed between parentheses

In extract 42, I have inferred that the speaker has dropped two tokens of the variable (fy). I have transcribed the missing trigger (fy) in parentheses adjacently to the target words *cyfweld* and *galw*. Both these tokens can now be coded as radical variants of (fy). Using this method of transcribing dropped triggers as I encounter them, finding all tokens of variables like (ei.m), (ei.f) and (fy) becomes easier, and does not need to rely on the use of the translation tier.

7.3.4 Language tagging

As in the *Siarad* corpus analysis, the language tagging process relied upon the dictionary criterion. In the process of applying the criterion to my data, I made certain to draw upon the precedent set by the original *Siarad* corpus transcribers in classifying words. In any cases where I was unsure about the correct classification of a word for whatever reason, I used the classification of that word in the *Siarad* corpus as a final guideline. Using this method not only made this part of the process quicker and more efficient, but also ensured that the *Radio Cymru* corpus is more directly comparable to the *Siarad* corpus, as words are consistently classified in the same way.

7.4 Calculating scores

Once language tagging is complete and all tokens of mutation for each variable have been transcribed, the next step is to calculate scores that represent the mean use of the variables in each programme. Section (7.4.1) describes this process for code-switching, while (7.4.1) describes the same process for the mutation variables.

7.4.1 Calculating scores – Code-switching

In order to calculate a score representative of the use of code-switching in each programme, I used a method similar to that used in chapter five. The first step is to extract the frequency of both Welsh and English words from the programme being analysed. In the case of *Post Prynawn* for example, the number of Welsh words can be calculated as follows:

```
FREQ -s@s:en -s@s:und +f Post_Prynawn1.cha
```

The number of English words can then be extracted using the following command:

```
FREQ s+@s:en +f Post_Prynawn1.cha
```

Once the frequencies of English and Welsh words are extracted for every speaker in the episode, they are added together to represent the use of English and Welsh throughout the whole programme. A proportion of English is then calculated as an index of the use of code-switching in that programme. For example, in the case of *Post Prynawn* the total sum of English words is 43, and the total sum of Welsh words is 24630. The number of English words is calculated as a proportion of the Welsh words:

$$43 / (24630+43) = 0.14\%$$

The mean code-switching use of contributors to *Post Prynawn* then is 0.14%, and this percentage can be used to compare *Post Prynawn* with other programmes.

7.4.2 Calculating scores – Mutation

I decided against calculating individual scores for individual speakers in each radio programme largely because the majority of mutation variables had relatively few tokens in each programme, particularly in the case of certain variables like (â/gyda). Instead, tokens for each speaker will be added together to calculate a mean use representative of the average of the variable for that programme. For example, when calculating a mean score for (o) in *Tudur Owen*, the first step is to calculate the sum of both standard and radical tokens from all speakers in each programme. Once all tokens from each episode of *Tudur Owen* (i.e. Tudur_Owen1, Tudur_Owen2 and Tudur_Owen3) are added together, the sum is 112 standard tokens and 18 non-standard tokens. The mean standard use is then calculated as a percentage:

$$112 / (112+18) = 86.15\%$$

This mean percentage can then be used to compare the average use of (o) on *Tudur Owen* with that of the other programmes, as well as with the *Siarad* corpus.

7.5 Categorizing Programmes

Although I have already introduced some basic information about the four programmes selected, I believe that some further analysis is required before they can be confidently said to represent distinct levels of formality. Measuring the level of formality could feasibly be approached in many ways, with a variety of paralinguistic and situational features described in previous studies as being indicative of more or less formal situations, including pitch variation, laughter, attention to speech, speaker familiarity and so on. I have focused here on variables that are relatively easy to quantify, namely laughter, overlapping speech and the number of simultaneous speakers. My rationale behind choosing this, and the methodology I used, is described further in section (7.5.1). But as well as understanding how programmes vary stylistically, I also intend to analyse the programmes in terms of the nature of the participants

taking part. This is because, as I found in my analysis of the *Siarad* corpus, differences in demographic traits of a speaker can have a significant effect on the likelihood, and perhaps ability of that speaker to use standard forms. Thus, the programs will be categorized in two distinct ways, along a social axis and along a stylistic axis. The style axis will reflect the extent to which a programme exhibits traits previously associated with formality. The social axis will reflect the variation among participants in terms of two main factors: age and level of education. Categorizing programmes along both axes will require different methods, with more of a quantitative used with the formality axis, and a more qualitative approach used for the social axis.

7.5.1 The Style Axis - Defining formality

For Ball et al's (1988) radio study, formality was defined as the degree of scriptedness exhibited by a programme. This argument makes sense, since programmes which are more constrained by pre-prepared scripts are more likely to show influence of the literary language, and the speakers may be more likely to focus upon producing careful speech. The drawback of this approach is that involving too much scripted material in the analysis bypasses a necessary emphasis on spontaneous speech in favour of readings of written material. It could be argued that the predominant use of standard variables in reading of scripts by professional newsreaders is hardly remarkable. To avoid this problem, this study will instead attempt to assess level of formality based on various different criteria which should ideally allow a more refined method of categorizing each programme.

Scripting will be retained as one method of categorizing programmes. Although for the reasons given above, scripted material won't be included in the analysis, programmes which have more scripted material will be considered more formal than those that have less such content. A second diagnostic tool used will be channel cues, which include laughter, turn overlaps and

discourse markers such as self-correction and retracing. Laughter and turn overlaps will both be considered indicative of lowered formality since they may be associated with decreased attention to speech, and to situations where speakers are generally less constrained by formal norms and conventions. Speech repair, which includes self and other-correction, will be considered indicative of the opposite, as it can be considered a sign that speakers are self-consciously directing attention to speech, ostensibly in order to respond to ideological or social norms (Alfonzetti 1998).

Number of speakers will also be considered as a factor. Crucially, only the number of speakers taking part in a conversation simultaneously will be considered relevant – not the total number of speakers who have spoken at some point during the course of a programme. The relevance of this factor is that a smaller number of simultaneous interlocutors may be more likely (or have more opportunity) to be able to focus attention upon their speech than speakers in a larger group. Table 84 illustrates the quantitative variables used to categorize the programmes by level of formality.

Variable	Symbol
Laughter	&=laughter
Overlapping speech	+<
Number of simultaneous speakers	n/a

Table 84 Variables used to measure formality in the radio programmes

7.5.2 Analysing channel cues

This section deals with the methodology for the quantitative analysis of formality in the *Radio Cymru* corpus. My analysis of the relative formality of each programme centres on two dependent variables, laughter (6.5.1) and overlapping speech (6.5.1). For the purposes of the analysis I considered that a greater frequency of either variable found in a particular programme would be indexical of greater levels of formality.

7.5.2.1 Laughter

In the CHAT transcription system which was used in the transcription of the *Siarad* corpus, laughter is coded with the symbol *&=laugh*. This is a relatively broad marker, which generally covers any type of laughter from the participants. I considered one token of *&=laugh* to correspond with the period extending from the initiation of a ‘burst’ of laughter by a speaker, to its cessation. So if a participant continues laughing for twenty seconds without stopping, this is still transcribed with one token of *&=laugh*. However, if a participant stops laughing even briefly or begins speaking normally before laughing again, I transcribe a new token of *&=laugh* on a separate main utterance tier. Speakers sometimes laugh while speaking, which is coded by placing relevant part of the utterance brackets and adding the symbol [=! Laugh] to the end of the event. To keep the analysis as simple as possible, I did not include this kind of laughter in the analysis.

Once all tokens of laughter have been accurately transcribed in each of the programmes, the next step is to count the relative frequencies of each variable. Given that the final length of each programme varied somewhat once post-editing was complete, and that this analysis involved whole counts, I ensured that they were comparable by only performing frequency counts on the first half hour of each episode, which I edited into discreet CHAT files (e.g. *Post_Prynhawn1_halved.cha*). Frequency counts were performed using the CLAN command *FREQ*, which I previously used to calculate the numbers of English and Welsh words. To have

CLAN retrieve the number of tokens of &=laugh in the first half of the first recorded episode of Post Prynawn for instance, I used the following command:

```
FREQ +s&=laugh +f Post_Prynawn1_halved.cha
```

This creates a file containing an output of the calculation for laughter in Post_Prynawn1 in a special directory. I then perform the same procedure on the two other recordings Post_Prynawn2 and Post_Prynawn3. Once I have the number of tokens of laughter for each programme, I then average these scores to produce a mean frequency of laughter that is representative of *Post Prynawn*. The same procedure is then applied to the other programmes.

7.5.2.2 Overlapping speech

Overlapping speech in CLAN is coded using the *lazy overlap* symbol +< which covers any situation where speakers' utterances occur simultaneously. As with the coding of laughter, one of my main aims in the transcription of overlapping speech was to ensure that it was performed in a comparable manner to that of the *Siarad* corpus. The main principle I followed was to code an overlapping symbol if a participant at any point starts speaking during another participant's utterance, however brief this is. I also only inserted a lazy overlap where speech, or an attempt to speak (including phonetic symbols representing stuttering), was part of the overlapping utterance. This means that I did not code the lazy overlap symbol where there was overlapping laughter, or any other kind of paralinguistic feature such as whistling or coughing etc.

Once each token of +< was coded, the process of calculating frequency was similar to that used for laughter, including the use of only the first half of each episode for analysis. For example, when analysing Post_Prynawn1, the following CLAN command is used to extract all tokens of overlapping speech:

```
FREQ +s+< +f Post_Prynawn1_halved.cha
```

As with laughter, the same procedure is then applied to both other episodes of *Post Prynhawn*, and the number of tokens of overlapping speech in both are averaged to create a mean frequency for Post Prynhawn. The same methodology can then be applied to the three other programmes.

7.5.3 The Social Axis

The social axis is intended to reflect the fact that participants in the various radio programmes will naturally tend to vary in their use of standard Welsh in the vernacular. This is based largely on the findings of the *Siarad* corpus analysis of mutation and code-switching in chapter 6, which found that older, more educated speakers, as well as those holding purist beliefs about language, were generally more likely to use standard forms than younger ones. Unlike the analysis of stylistic factors, information about the social factors affecting the participants in the radio programmes cannot always be gleaned directly from the data. Although in some cases information could be requested directly from the hosts or other contributors to the programmes, it is unlikely that enough relevant information could be gathered to perform a full analysis of differences in age and education between them. This means that a slightly broader and more indirect qualitative approach is required in order to categorize the programmes in this way. I will now describe the two main factors in the social axis analysis.

7.5.3.1 Social Axis - Age

In the *Siarad* corpus analysis in chapter 6, age was the most important variable affecting standard use of all variables, with younger age correlating with a reduced likelihood of the speaker producing standard variants. Information about ages can be gathered both by simply asking participants where possible, and by making a judgement on the general ages of participants having listened to the audio recordings.

7.5.3.2 Social Axis – Education

Education was also found to be a powerful factor affecting speakers' likelihood of using standard Welsh forms of mutation, although this effect did not carry through to code-switching. Since finding information about all participants' level of education may not be practical, I will be taking level of education here to stand for a generalized concept of level of exposure to standard Welsh. In other words, the more participants are believed to have encountered standard varieties of Welsh during their lives, the higher their position on the social axis will be deemed to be.

7.6 Summary

This chapter has provided a description of the way in which the *Radio Cymru* data was both collected and transcribed. The next chapter will provide the results for both the categorization of the programmes by level of formality and prestige, and the results of the analysis of distribution of the mutation and code-switching variables across each programme.

8. Radio Cymru Corpus Analysis Results

8.0 Introduction

This chapter provides the results for the analysis of the *Radio Cymru* corpus. Section (8.1) presents of the categorization of the programmes by level formality, using the frequency of laughter (8.1.1) and overlaps (8.1.2) in each programme as a quantitative index of formality. Section (8.2) presents the results of the analysis of the social axis attributes of each programme, arranging each programme by level of standard Welsh likely to be used normally by the speakers. Section (8.3) provides the standard mean usage of each variable across all four programmes. Finally, section (8.4) places the results of the *Radio Cymru* corpus analysis in context, showing how the arguably performative speech of the presenters in each programme either resembles or is distinct from that used in the ‘natural’ speech *Siarad* corpus.

8.1 Categorizing programmes by level of formality

This section presents the result of the quantitative analysis of the level of formality for the four programmes which make up the *Radio Cymru* corpus. There are two variables in the quantitative analysis which I claim allow me to measure the relative level of formality in a radio programme; laughter (8.1.1) and overlapping speech (8.1.2).

8.1.1 Laughter

The results of this analysis reflect the mean frequency of laughter in the first half-hour of each episode of the programmes. A higher frequency of laughter was considered to be indexical of a lower level of formality in a programme.

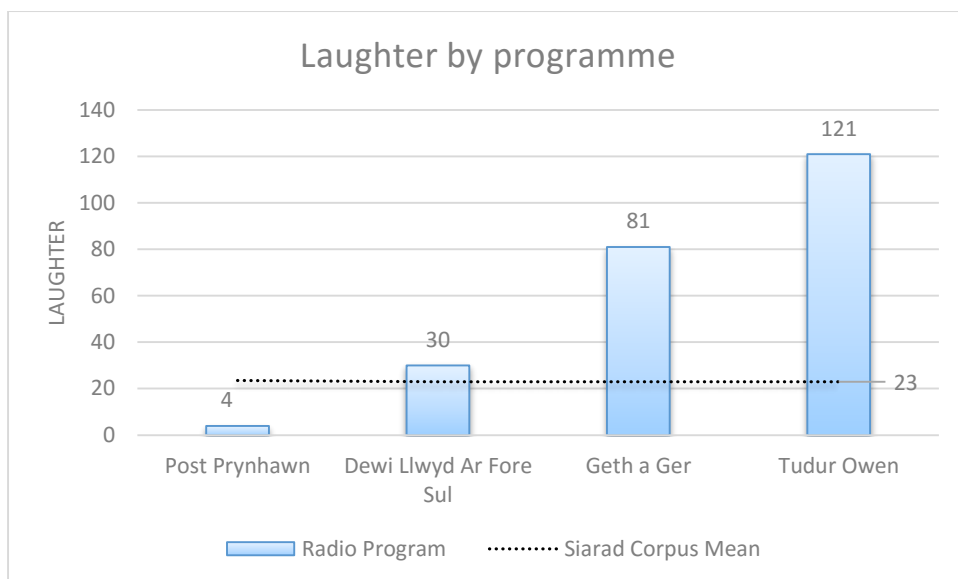


Figure 22 The average frequency of laughter in each programme

The programme with the lowest frequency of laughter was *Post Prynawn*, with a mean of 4 tokens of laughter across all three episodes. *Dewi Llwyd Ar Fore Sul* has the second lowest frequency, although with substantially more laughter than *Post Prynawn* with a mean of 30 tokens. *Geth a Ger* has yet more laughter with a mean of 81 tokens, while *Tudur Owen* has the most frequent laughter with 121 tokens. The mean frequency of laughter of *Siarad* corpus conversations was 23 tokens, which is slightly lower than that of *Dewi Llwyd Ar Fore Sul*.

8.1.2 Overlapping Speech

This section presents the results for the overlapping speech variable. As with the laughter variable, mean frequencies of overlapping speech were calculated from the first half-hour of each episode of all four programmes. More frequent tokens of this variable were considered indexical of a lower level of formality in a given programme.

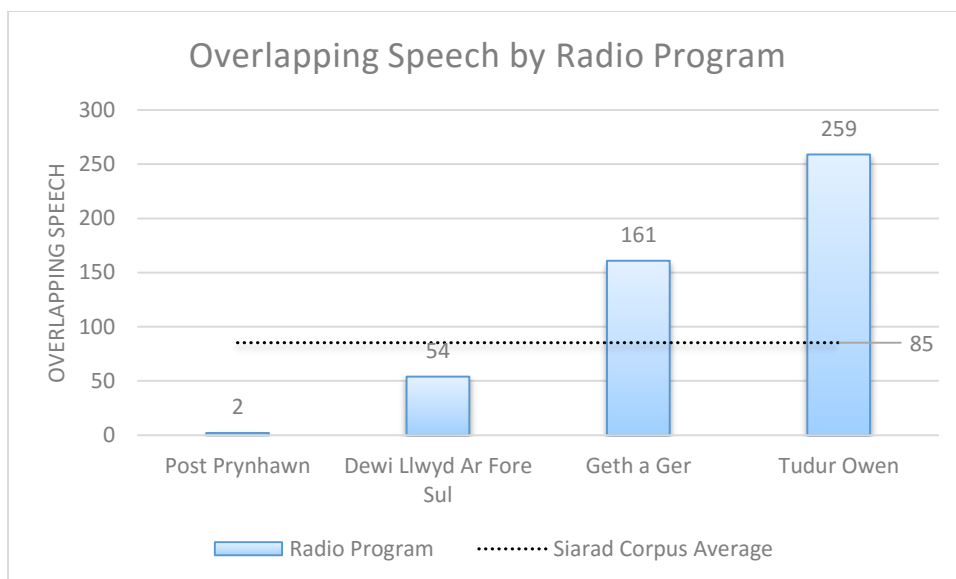


Figure 23 The mean frequency of lazy overlaps for each programme

The distribution of overlapping speech tokens mirrors relatively closely that of the distribution of laughter. *Post Prynawn* has the lowest frequency of overlapping speech, with a mean of only 2 overlap tokens per episode, while *Dewi Llwyd Ar Fore Sul* has the second lowest with a mean of 54 tokens. *Geth a Ger* has the second most frequent overlaps with a mean of 161 tokens, while *Tudur Owen* has the most frequent, with a mean of 259 overlap tokens. The *Siarad* corpus mean frequency for overlapping speech seems to fall between that of *Dewi Llwyd AFS* and *Geth a Ger*, although it was closer to the former.

8.1.3 Formality – Summary

Both formality variables follow a similar pattern, with *Post Prynawn* showing very few tokens for laughter and overlapping speech, and *Dewi Llwyd Ar Fore Sul*, *Geth a Ger* and *Tudur Owen* showing increasing mean frequencies, in that order. In both cases, the *Siarad* corpus mean was situated at a point which most approximated *Dewi Llwyd AFS*, with the *Tudur Owen* and *Geth a Ger* programmes exhibiting much higher levels of laughter and overlapping speech than occurs in typical *Siarad* corpus conversations. The next section will focus on using these

findings to decide on the placement of each programme along a stylistic axis, from the most formal to the most informal programme.

8.2 The style axis

The final structure of the style axis is displayed in figure 24, which illustrates how the corpora to the left are more formal than those to the right. My approach in categorizing the programmes involved not only considering the relative distributions of the quantitative variables laughter and overlapping speech, but also making observations about other factors that could be argued to affect formality in a programme, including the maximum numbers of simultaneous speakers at any one time, the presence and type of music played (or its absence), the subject matter discussed and notions of subversiveness and irony. The *Siarad* corpus is considered the most informal (situated at the far left of the axis) dataset analysed, despite exhibiting lower mean frequencies of laughter and overlapping and speech than three out of the four radio programmes. I am making the assumption here that the *Siarad* corpus contains fewer of the situational pressures which drive formal speech in the *Radio Cymru* corpus, including the institutional values of *Radio Cymru*, the expectations of a large audience and unknown audience who act as referees, and the frequently unfamiliar interlocutors. Most importantly, I argue that the *Siarad* corpus represents unmonitored speech which is representative of typical conversation, while the *Radio Cymru* corpus involves different types of tailored speech. In all programmes, the presenters are necessarily highly conscious of their own speech patterns to an extent that speakers in a naturalistic corpus are unlikely to be.

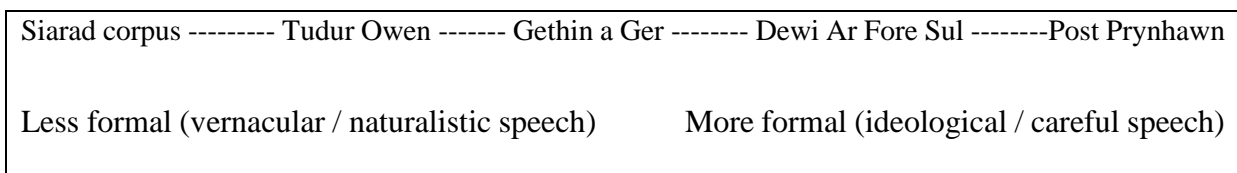


Figure 24 The distribution of each programme by level of formality

I will now discuss the position of each programme on the stylistic axis in turn, and explain my rationale behind arranging them as I have, with reference to the distribution of the formality variables from section (8.1) as well as some of the other factors mentioned above. The programmes will be discussed in order of decreasing formality; *Post Prynawn* (8.2.1), *Dewi Llwyd Ar Fore Sul* (8.2.2), *Geth a Ger* (8.2.3) and *Tudur Owen* (8.2.4).

8.2.1 *Post Prynawn* - Formality

At the rightward and most ‘formal’ end of the axis is the news programme *Post Prynawn*. The most obvious motivation for placing this programme here is the very low mean frequency of both laughter and overlapping speech in this programme. The scarcity of these channel cues implies a strong tendency for participants to speak carefully, monitor their speech intensively, and to avoid emotional engagement. The lack of overlapping speech is clearly related to the fact that participants in the programme generally speak one at a time in well-formed utterances, presumably allowing them substantial time to apply the maximum possible attention to speech. While this is the programme analysed that has the highest number of speaking participants per typical episode, there are never more than two interlocutors speaking at any one time, and interviews between participants are typically too short to build up any kind of rapport. The rarity of laughter meanwhile is arguably indicative of a tendency for participants to consider the context one where respectful, non-emotionally engaged speech is most appropriate. The formal ‘feel’ of programme is reinforced by the lack of a soundtrack - this is the only programme where there is no music played in between segments. In addition to this, the predominant topic of discussion is also the most ‘serious’ of the programmes in the thesis: news and current events. According to my predictions concerning formality and stylistic variation, this should mean that participants may be focused on avoiding the use of features influenced by English or otherwise non-standard speech. To summarize, the factors that make this the most formal programme can be summarized with the following points:

- It has the lowest frequency of laughter in any of the programmes
- It has the lowest frequency of overlapping speech in any of the programmes
- The most formal subject matter: *news*
- A general lack of familiarity between interlocutors
- Low maximum number of simultaneous speakers (two at most)
- No music

8.2.2 Dewi Llwyd Ar Fore Sul - Formality

Dewi Llwyd Ar Fore Sul is the second most formal programme on the style axis. Although it is similar to *Post Prynhawn* in that it shares a presenter and a news oriented theme, certain features make it arguably more informal by comparison. Although the programme does include discussion of quite ‘serious’ topics, including news and current events, a significant portion of the programme consists of an informal chat format where the presenter discusses newspapers with two interlocutors. This section in particular usually involves relatively frequent laughter and overlapping speech, although at lower intensity than *Geth a Ger* and *Tudur Owen*. The ‘birthday interview’ segment of the programme also typically involves a degree of emotional engagement between the presenter and interviewee, including irreverent topics, personal disclosures and humour. In terms of numbers of simultaneous speakers, *Dewi Llwyd AFS* beats not only *Post Prynhawn* but also *Geth a Ger*, with the chat section of the programme often involving dynamic and emotionally involved three way conversations between both guests and the presenter. Another factor that arguably favours an informal feel are the musical interludes which punctuate each section, particularly in contrast with *Post Prynhawn*, which is solely based around spoken word content.

There are also some factors which make this programme more formal in comparison with *Geth a Ger* and *Tudur Owen*. One of these is the lack of familiarity between speakers – all guests and the presenter *Dewi Llwyd AFS* are typically interacting as strangers who have not met

previously. This means that the context is essentially professional, in contrast to the more relaxed interactions between presenters on *Geth a Ger* and *Tudur Owen*, where the majority of those taking part are friends, and familiar with each other outside of the programme. One detail in particular which confirms this interpretation is the fact that the formal 2nd person pronoun *chi*, as opposed to the informal pronoun *ti* (or dialect form *chdi*), is used categorically by all participants in this programme, as it is in *Post Prynawn*. In terms of the musical interludes too, the type of music played on *Dewi Llwyd AFS* may be significant. Whereas *Tudur Owen* and *Geth a Ger* have musical playlists which tend towards youth-oriented and rebellious genres of hip-hop and rock music, *Dewi Llwyd AFS* primarily plays a mix of classical and folk music, which is arguably indicative of respectability and a more refined aesthetic sensibility, as well as being aimed at an older audience. Finally, as the quantitative analysis showed, laughter and overlapping speech are less frequent on *Dewi Llwyd AFS* than for both *Geth a Ger* and *Tudur Owen*.

To summarize, the factors that make *Dewi Llwyd AFS* the second most formal programme can be summed up in the following way:

- It has the second lowest frequency of laughter in any of the programmes (more laughter than *Post Prynawn*)
- It has the second lowest frequency of overlapping speech in any of the programmes (more overlaps than *Post Prynawn*)
- It includes more simultaneous interlocutors than *Post Prynawn* and *Geth a Ger* (a maximum of three)
- A general lack of familiarity between interlocutors
- Contains a mixture of serious and light hearted elements: featuring the news as well as a 'chat' format discussing show business and fashion
- Involves a 'highbrow' selection of music - i.e. a mixture of classical and folk songs

8.2.3 Geth a Ger - Formality

Geth a Ger is the second most informal radio programme. In terms of factors that support considering *Geth a Ger* as more informal than *Dewi Llwyd AFS*, perhaps one of the clearest is that the frequency of laughter and overlapping speech is relatively high in *Geth a Ger*. Despite there being only two presenters, the interaction between the presenters is dynamic and involved, featuring very frequent laughter and overlapping speech. This is perhaps driven by the fact that both presenters are familiar with each other outside of the programme, having been friends for years before starting the programme, and their close friendship is often a topic of discussion. Another factor favouring informality is the subversive humour which often characterizes the programme. This involves frequent ironic references towards Welsh cultural institutions such as the magazine *Golwg*, beloved long running television programme *Pobl Y Cwm* and Aberystwyth University's Welsh-medium hall of residence *Pantycelyn*. The topics of conversation that the hosts on this programme typically discuss could also be considered more informal, such as television programmes, films and various aspects of youth culture.

A few factors make me believe that this programme should be considered more formal in relation to *Tudur Owen*. One of them is that it has fewer simultaneous speakers than *Tudur Owen*, as well as *Dewi Llwyd Ar Fore Sul*. With only the two presenters in each programme, there is theoretically more of an opportunity to focus attention to speech as well as a less chaotic atmosphere in which to navigate, especially in contrast to *Tudur Owen* where participants often have to fight hard to be heard. Another striking feature of *Geth a Ger* is a recurrent pattern of reformulating and repairing speech, as well as metacommentary about the language being used, which will be explored in more depth in the next chapter. This behaviour seems indicative of frequent audio monitoring, as well as of self-consciousness about linguistic norms and speaking ability. Finally, one recurring segment; *gwers Ger* ("Ger's lesson") where one of the presenters gives a semi-serious talk on scholarly subjects has a didactic feel, which although sometimes

ironic, can be earnest enough that it might be expected to lead to a more formal style. This and other aspects of the programme will be discussed in greater detail in the discourse analysis in chapter 9.

- The second highest frequency of laughter in any of the programmes (after *Tudur Owen*)
- The second lowest frequency of overlapping speech in any of the programmes (after *Tudur Owen*)
- Light-hearted topics: films, music, and youth culture.
- Relatively didactic educational segment *Gwers Ger*
- Frequent audio monitoring, including speech reformulation and flagging
- Familiarity between the presenters
- The low number of simultaneous interlocutors (only two)

8.2.4 Tudur Owen - Formality

Tudur Owen is the second most informal corpus in this study (after *Siarad*), and is the most informal of the radio programmes. This is due to the very frequent occurrence of laughter in the programme and of overlapping speech, as well as frequent interactional markers indicating emotional arousal and lack of restraint including ‘silly voices’, impressions, shouting and whispering. It also has the largest number of simultaneous speakers at any one time in the corpora, with six speakers taking part in one exchange. Participants in this programme are also familiar with one another outside of the programme, and details about past and future social engagements involving the presenters are often discussed and mined for humour. Another factor contributing to my decision to categorize *Tudur Owen* in this way is harder to pin down – the degree of subversiveness. Many elements of the programme, in particular the humour, seems aimed at satirizing the ideology of *Radio Cymru* itself, and this takes place to a much

stronger and more pointed degree than in *Geth a Ger*. Much of this involves linguistic humour aimed at undermining puritanical language norms and conservative values.

- The highest frequency of laughter in any of the programmes
- The highest frequency of overlapping speech in any of the programmes
- Light-hearted topics: films, music, and youth culture
- Little evidence of audio-monitoring behaviour
- Familiarity between the presenters
- High number of simultaneous interlocutors (up to six)
- Frequently subversive and ironic content

8.2.5 Formality - Summary

Each programme has been placed along a continuum of formality, with *Post Prynawn* the most formal programme followed by *Dewi Llwyd AFS*, *Geth a Ger* and *Tudur Owen* in decreasing order of formality. This categorization mainly involved a quantitative analysis of two variables meant to measure formality, laughter and overlapping speech, but also incorporated other factors including the number of simultaneous interlocutors, type of musical content and level of subversive content in each programme. The next section will move on to consider social factors.

8.3 The social axis – categorizing the programs by social factors

The social axis should represent a continuum of programmes whose participants have more or less exposure to standard varieties of Welsh. The *Siarad* corpus analysis found that age was the most significant independent variable affecting both mutation and code-switching, and that level of education was also an important factor. My placement of programmes on the social axis will thus mainly depend on what I can glean about these factors – the relative ages of the participants and their level of education, as well their potential for exposure to standard Welsh.

Since I do not have access to in-depth information about all participants in these radio programmes (beyond what is revealed during the broadcasts themselves) this analysis will necessarily be of a broader and more qualitative nature than the *Siarad* analysis, and may involve some conjecture.

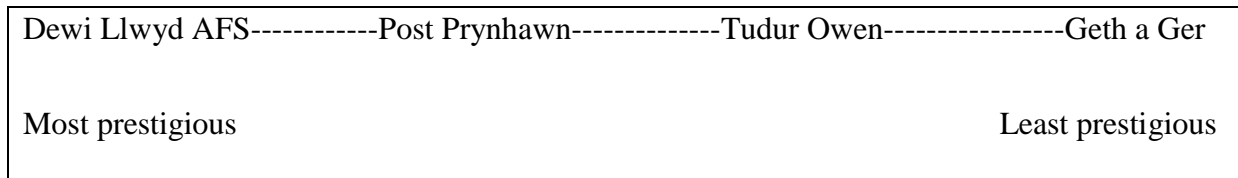


Figure 25 The Social Axis distribution of programmes according to prestige

The programmes will be in discussed in order of descending position on the social axis; *Dewi Llwyd Ar Fore Sul* (8.3.1) *Post Prynawn* (8.3.2), *Geth a Ger* (8.3.3) and *Tudur Owen* (8.3.4).

8.3.1 *Dewi Llwyd AFS – The social axis*

Dewi Llwyd AFS is placed at the most prestigious end of the social axis, having the highest concentration of participants bearing the characteristics of prestige speakers of Welsh. One important factor supporting this observation is that nearly all of the participants are middle aged or older. There is only one exception to this pattern, a young guest in her twenties who takes part in the newspaper segment in one episode, although this might be considered the exception that proves the rule. Another factor is the speakers’ experience and expected level of exposure to domains where formal Welsh may be used. Participants in the newspaper segment are selected on the basis of their level of insight into current affairs and their ability to speak knowledgably and eloquently in Welsh on serious matters. This means that guests often have prominent positions in prestigious institutions. For example, one episode recorded involves a PhD graduate and ex-chairwoman of the Welsh Language Society, a lecturer and academic at Bangor University and the head of a local community organization.

The interview segment (or ‘birthday’ segment) also deals almost exclusively with participants who are relatively well known for their work in Welsh medium institutions of high prestige. These have included Geraint Jenkins, a historian famous for revitalizing the study of history through the medium of Welsh, Peter Lord, who similarly advocated for the study of art history in Welsh, and Christine James, the then archdruid, who in addition to representing the apogee of Welsh speaking culture (the Eisteddfod) is also a prolific Welsh medium poet who lectures at Swansea University. It is reasonable to say that advanced degrees are relatively common among this cohort, which implies a certain level of exposure to higher registers of Welsh as well as experience in navigating prestigious Welsh medium domains. All three of the birthday guests are also published authors of multiple books, which also implies familiarity with standard written forms of the language.

- The majority of the participants are active in domains where standard forms of Welsh may be used
- The ages of the participants are generally middle age and over
- The majority of the participants are highly educated (including some current and ex-university lecturers)

8.3.2 Post Prynawn - The social axis

Post Prynawn arguably involves the most varied participants, since it is a programme in which any member of the public could potentially be involved by being invited as contributor. There are broadly three types of participants. First are the professional consultants and presenters of *Radio Cymru*, who frequently contribute to the programme and are well integrated into the institutional structure of the station. These participants are well versed in the values of the station, are usually highly educated and have a strong professional motivation to contribute in a way which conforms to the expectations of the institution. A second type of participant on *Post Prynawn* encompasses those who work for other prominent institutions, and who are

asked to comment on stories in specialized field where *Radio Cymru* correspondents lack the relevant expertise. Examples of these kinds of participants include a retired doctor who advises on various healthcare related issues (such as a breakout of measles), the head of a Christian charity discussing the impact of the Syrian war on refugees, and a classical music expert discussing an upcoming concert.

Despite being outside of *Radio Cymru*'s direct institutional structure, they might be expected to have both a higher average level of education and a more positive orientation towards Welsh than the ordinary speakers of the *Siarad* corpus, but the extent to which this is true varies depending on multiple further factors. The last category of participant are essentially ordinary members of the public who are invited to provide information about events they happen to have been caught up in. Examples are an English teacher who happened to be living in an American city where a widely reported kidnapping case took place, and a farmer from mid-Wales who was invited to discuss the governments' agricultural policy. These kinds of participants are presumably outside of the institutional structure of *Radio Cymru*, and may be comparable to the ordinary speakers of the *Siarad* corpus, with less exposure to standard Welsh than the guests typically invited to contribute on Dewi Llwyd AFS, for instance. Participants on *Post Prynghawn* could thus be generally expected to be relatively highly placed on the social axis, since they are generally highly educated and older, although they reflect a less consistently prominent and prestigious background than contributors in *Dewi Llwyd AFS*.

- A mixture of participant types, including phone-in members of the public and professional correspondents, many of whom will be familiar with standard Welsh to varying degrees
- The ages of the participants are generally middle age and over

8.3.3 Tudur Owen - The social axis

Although I have argued that *Tudur Owen* is the least formal programme in this study, on the social axis I have placed it in the second lowest position. The most important factor in making this decision was the ages of the presenters – the presenters and contributors on *Tudur Owen* are slightly older than those in *Geth a Ger*. The main presenter, Tudur Owen himself, has worked in the Welsh language media for a substantial period of time, and has recently written and published a novel written in Welsh. In terms of both exposure to Welsh and their mean ages then, I would argue that *Tudur Owen* should be placed at a higher position than *Geth a Ger*. In terms of its placement relative to the *Post Prynhawn* and *Dewi Llwyd*, I would argue that the presenters and participants have backgrounds tailored towards entertainment more than news reporting, and may be less used to speaking Welsh in the more formal contexts of these programmes.

- The participants are older than those of *Geth a Ger*
- The presenters have worked in the Welsh media for an extended period

8.3.4 Geth a Ger – The social axis

The defining feature of *Geth a Ger* in terms of the social axis is that both presenters are relatively young in comparison with the participants of the other programmes analysed, with GET in his early thirties and GER in his late twenties. Since age was a particularly significant variable in the *Siarad* corpus analysis, I would argue that this is a particularly important factor in shaping the presenters' vernacular speech patterns. Another important factor is that the presenters may have had less exposure to higher domains of Welsh, as well as less experience using higher registers, due to their relative inexperience in presenting compared with most other presenters on *Radio Cymru*.

- The presenters are the youngest in any of the four programmes
- They have less media experience than most other presenters

8.3.5 The social axis – Summary

Using age and exposure to standard Welsh of participants as an index of prestige, I have categorized the programmes by the likelihood of the participants' having been habituated to using standard Welsh. The placement of the programmes according to the social axis is distinct to the placement on the stylistic axis, with *Dewi Llwyd AFS* the highest placed, followed by *Post Prynawn*, *Tudur Owen* and *Geth a Ger*. The next section will use my analyses of the social and stylistic statuses of each programme to interpret the distribution of the linguistic variables.

8.4 Results for linguistic variables

This section provides the results for the distribution of each variable by radio programme. The prediction made for the *Radio Cymru* corpus was that the distribution of variants for marker variables would correlate with my categorization of each programme by level of formality, while indicators would reflect other factors, including contributors' ages and presumed level of exposure to standard Welsh.

8.4.1 Results – (am)

Figure 26 presents the distribution of standard mutation mean percentages for (am) in each programme:

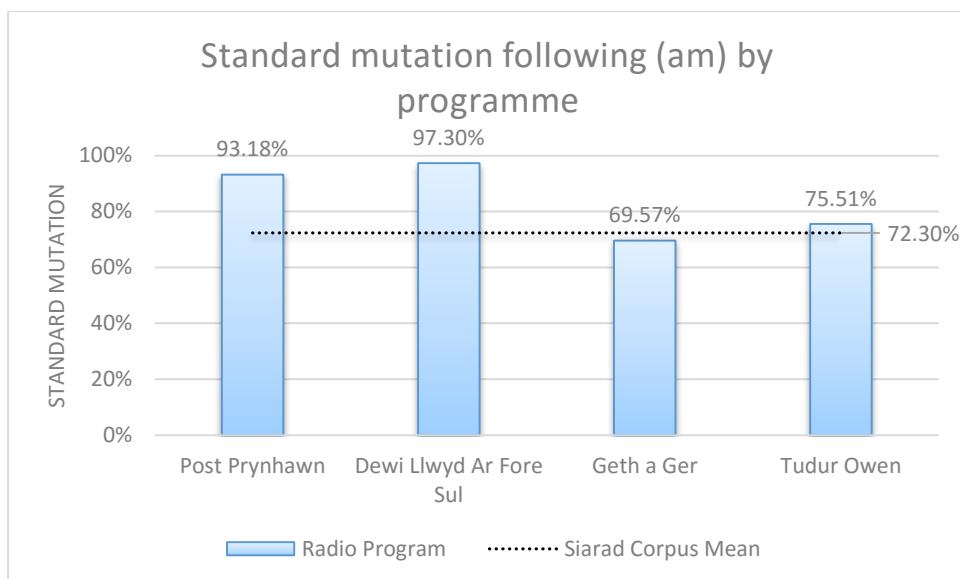


Figure 26 Distribution of standard tokens of (am) by programme

Radio Programme	Standard Tokens	Non-Standard Tokens
Post Prynawn	41	3
Dewi Llwyd AFS	72	2
Geth a Ger	32	14
Tudur Owen	37	12

Table 85 Number of standard and non-standard tokens of (am) in each programme

Dewi Llwyd Ar Fore has the highest standard mean use of (am) with 97.30%, while *Post Prynawn* follows this with 93.18%. The other two programmes form a distinct group, with *Tudur Owen*'s mean use of (am) slightly higher than *Geth a Ger*'s at 75.51% compared with 69.57%.

8.4.2 Results – (o)

Since this variable was strongly affected by place names, I have decided to calculate score for (o) in two conditions – one with place names removed, and one with only place names.

Figure 27 shows the results for (o) without place names:

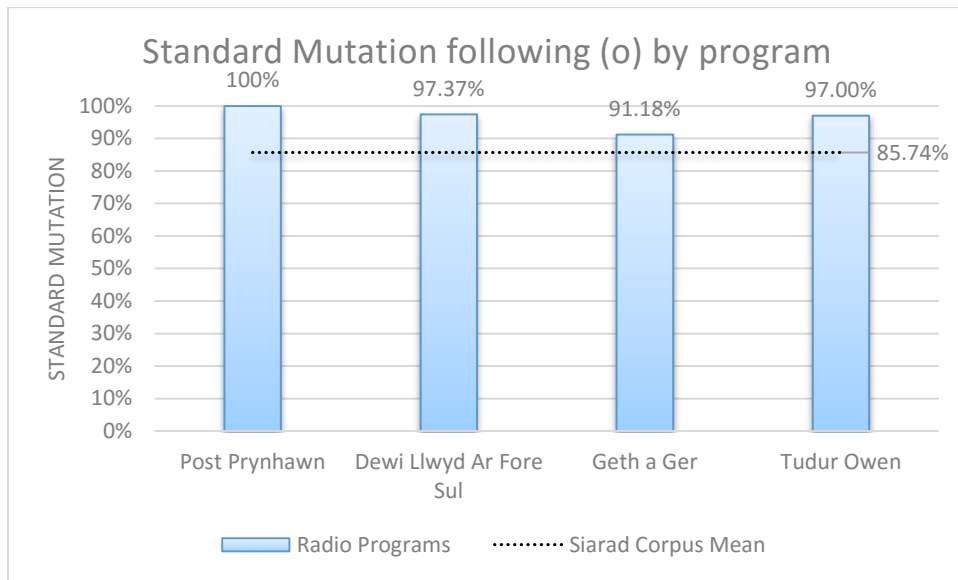


Figure 27 Standard mutation of (o) without place names by programme

Radio Programme	SM Tokens	Radical Tokens
Post Prynawn	171	4
Dewi Llwyd AFS	241	5
Geth a Ger	55	4
Tudur Owen	97	3

Table 86 Number of standard and non-standard tokens of (o) ‘without place names’ in each programme

All programmes have very high rates of standard mutation for (o) without place names. *Post Prynawn* has a categorical 100% mean use of standard mutation here, while *Dewi Llwyd AFS* have slightly lower means of 97.37% and 97.00% respectively. *Geth a Ger*’s mean score is the lowest of any programme at 91.81%, although this is still higher than the population mean of 85.74% represented by the *Siarad* corpus.

Figure 28 provides the mean percentage standard use of (o) with only place names:

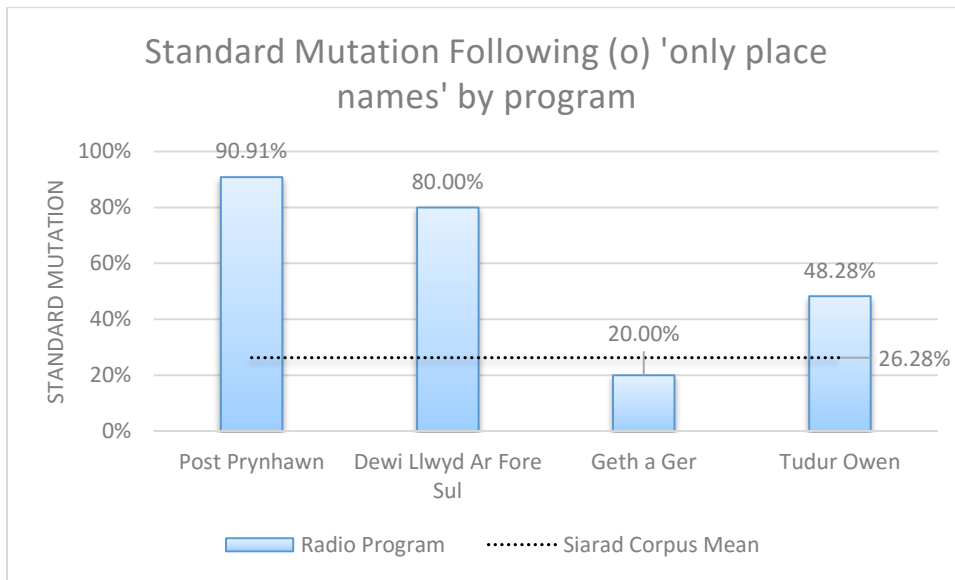


Figure 28 Standard mutation of (o) with only place names

Radio Programme	SM Tokens	Radical Tokens
Post Prynawn	10	1
Dewi Llwyd AFS	8	2
Geth a Ger	3	12
Tudur Owen	14	15

Table 87 Number of standard and non-standard tokens of (o) 'only place names' in each programme

With only place names analysed, the distribution of standard mutation following (o) is rather different, with *Geth a Ger* and *Tudur Owen* using far fewer standard forms with place names than *Post Prynawn* and *Dewi Llwyd Ar Fore Sul*. *Geth a Ger* has the lowest mean standard use with only 20% of place names mutated after (o), while *Tudur Owen* has a mean of 48.28%. The difference between both conditions for (o) seems to indicate that place names have affected

the two least formal programmes to a much greater extent than the two most formal programmes. Standard mutation of place names is higher on *Tudur Owen* than the population mean of 26.28%, while on *Geth a Ger* it was slightly lower.

8.4.3 Results – (yn)

The results for (yn), which are index scored, again show a division between the two groups of programmes, with *Post Prynawn* and *Dewi Llwyd Ar Fore Sul* both with very close mean standard usage scores at 189 and 190 respectively.

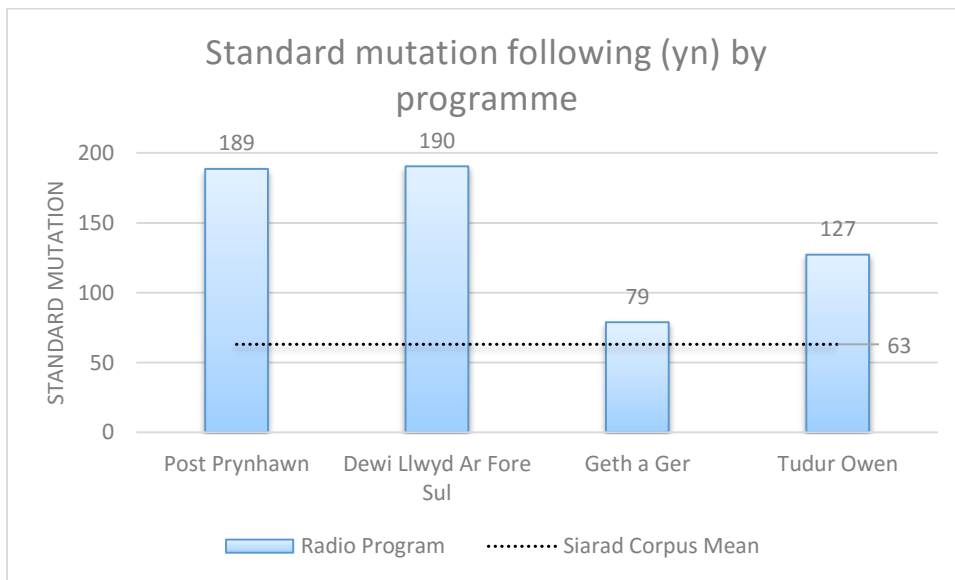


Figure 29 Standard mutation of (yn) by radio programme

Radio Programme	Nasal Mutation	Soft Mutation	Radical Tokens
Post Prynawn	66	2	3
Dewi Llwyd AFS	108	3	4
Geth a Ger	12	6	20
Tudur Owen	28	19	12

Table 88 Number of standard, intermediate and non-standard tokens of (yn) in each programme

Tudur Owen has a score of 127, which is almost double that of *Geth a Ger*, which has the lowest standard use of (yn) at 79 index points. All programmes, including *Geth a Ger*, have mean scores that are higher than the average population mean of 63 index points for this variable. In terms of the type of mutation applied, presenters on *Post Prynawn* and *Dewi Llwyd AFS* seem to favour the standard NM form, *Geth a Ger* mostly used the radical, while *Tudur Owen* mostly used the standard, but also frequently used tokens of SM.

8.4.4 Results – (fy)

Tokens of (fy) were very infrequent in the most formal programme *Post Prynawn*, with only four in total being used during the three hours recorded. For this reason I have decided not to calculate a mean percentage of (fy) for *Post Prynawn*.

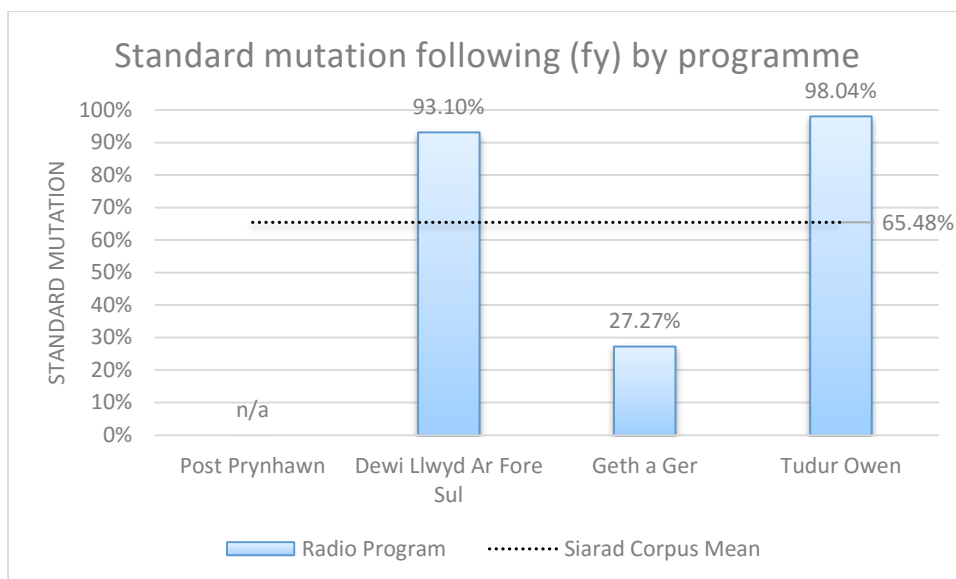


Figure 30 Standard mutation of (fy) by radio programme

Radio Programme	Standard Tokens	Non-Standard Tokens
Post Prynawn	4	0
Dewi Llwyd AFS	54	4
Geth a Ger	9	24
Tudur Owen	50	1

Table 89 Number of standard and non-standard tokens of (fy) in each programme

The mean standard percentage of the Siarad Corpus for (fy) is 65.48%, which falls roughly halfway between the high means of *Dewi Llwyd AFS* and *Tudur Owen* and the much lower mean use of *Geth a Ger*. Although it is the least formal programme, *Tudur Owen* has the highest mean use of standard (fy) at 98.04%, while *Dewi Llwyd AFS* has the second highest mean at 93.10%. *Geth a Ger*, which I have categorized as more formal than *Tudur Owen*, has a mean standard use far below that of the other programmes, as well as the mean use of Siarad speakers, with a mean of only 27.27%.

8.4.5 Results – (a)

The mean results for (a) can be split into two groups, with the two most formal programmes exhibiting substantially more frequent use of standard tokens than the two least formal groups. *Post Prynawn* has the highest mean standard use at 80.56%, while *Dewi Llwyd AFS* has a slightly lower standard mean of 71.62%.

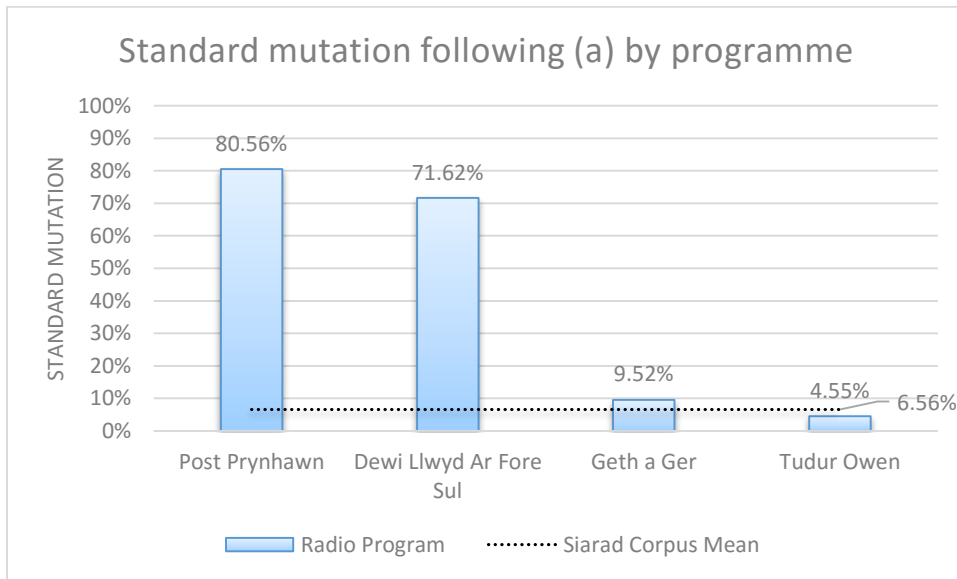


Figure 31 Standard mutation of (a) by radio programme

Radio Programme	Standard Tokens	Non-Standard Tokens
Post Prynawn	29	7
Dewi Llwyd AFS	53	21
Geth a Ger	2	19
Tudur Owen	2	42

Table 90 Number of standard and non-standard tokens of (a) in each programme

Both of the least formal programmes are far closer to the *Siarad* corpus mean of 6.56%. While *Geth and Ger*'s mean is slightly higher than the public mean with 9.52%, *Tudur Owen*'s mean is lower than this at 4.55%.

8.4.6 Results – (â/gyda)

The results for (â/gyda) again indicate that *Post Prynawn* and *Dewi Llwyd AFS* pattern similarly, with *Post Prynawn* having the highest standard use mean at 68.75% and *Dewi Llwyd AFS* following at 56%.

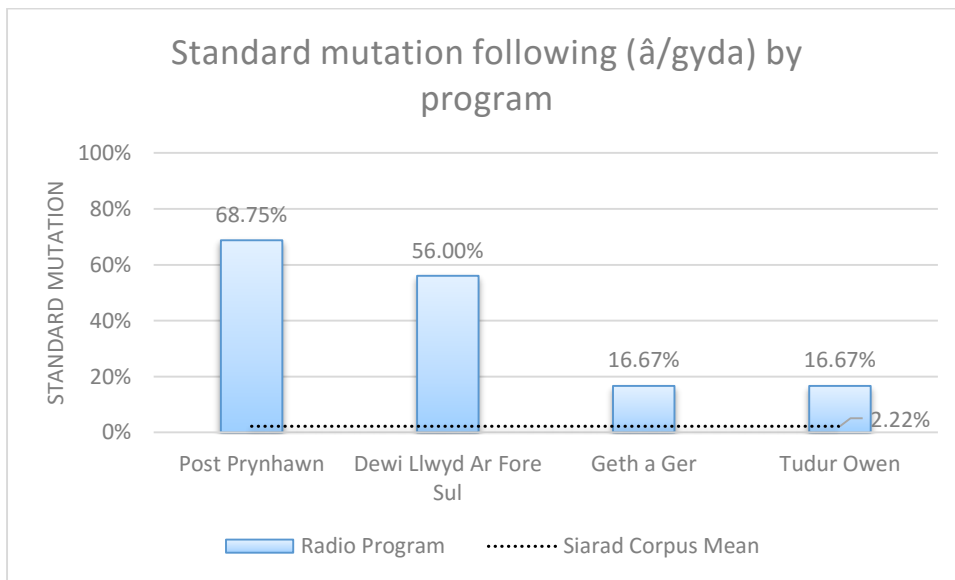


Figure 32 Standard mutation of (â/gyda) by radio programme

Radio Programme	Standard Tokens	Non-Standard Tokens
Post Prynawn	11	5
Dewi Llwyd AFS	14	11
Geth a Ger	1	5
Tudur Owen	1	5

Table 91 Number of standard and non-standard tokens of (â/gyda) in each programme

Geth a Ger and *Tudur Owen* both shared the third lowest use of this variable at 16.67%. The overall number of tokens in the two informal programmes was very low, which means that the results here should be taken with some caution. However, it should be noted that the single standard token used on *Tudur Owen* was spoken by a sports correspondent rather than a ‘core presenter’, while on *Geth a Ger* it was spoken by one of the two regular presenters.

8.4.7 Results – (ei.m)

The results of the radio analysis shows very high rates of standard use of (ei.m) across all programmes. The highest means are in the most and least formal programmes, *Post Prynawn* and *Tudur Owen*, which both had only standard tokens for this variable, and thus a mean of 100%.

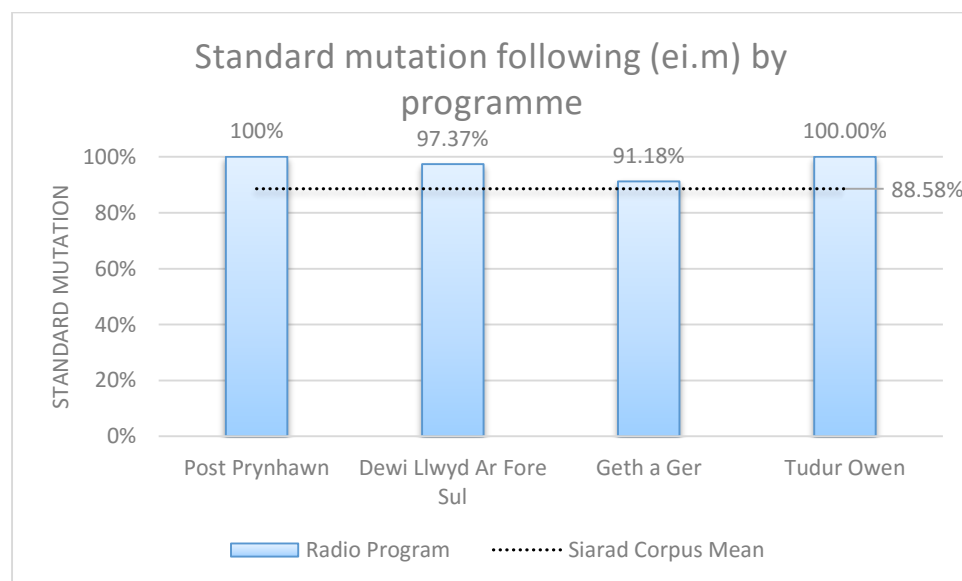


Figure 33 Standard mutation of (ei.m) by radio programme

Radio Programme	Standard Tokens	Non-Standard Tokens
Post Prynawn	63	0
Dewi Llwyd AFS	74	2
Geth a Ger	31	3
Tudur Owen	56	0

Table 92 Number of standard and non-standard tokens of (ei.m) in each programme

Dewi Llwyd Ar Fore Sul has a slightly lower mean at 97.37%, while *Geth a Ger* is lower again, with a mean nearly ten percent lower than the other programmes at 91.18%.

8.4.8 Results – (ei.f)

With only two tokens of (ei.f) in *Geth a Ger*, producing a representative score for this programme was not possible, so *Geth a Ger* is left out of this analysis.

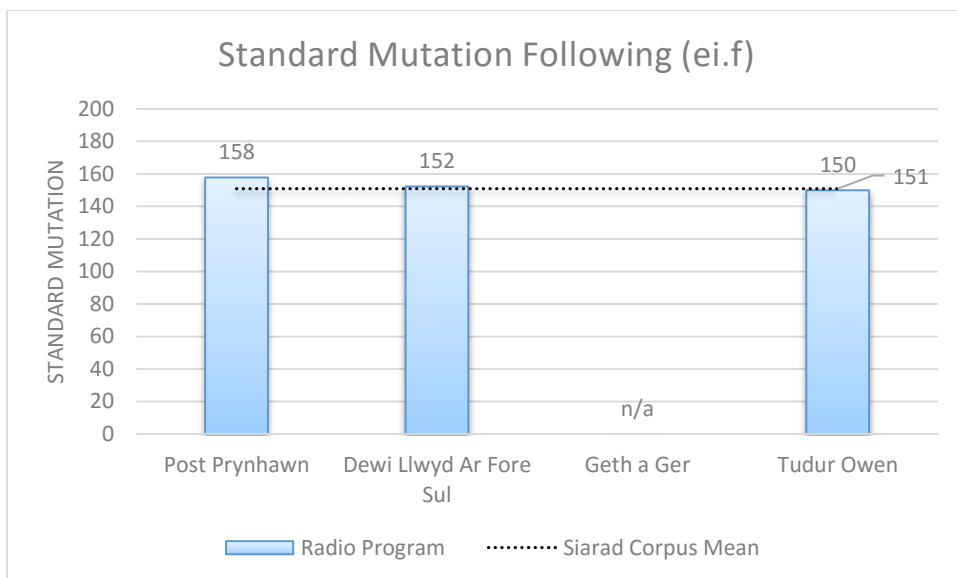


Figure 34 Standard mutation of (ei.f) by radio programme

Radio Programme	AM Tokens	SM Tokens	Radical Tokens
Post Prynawn	11	8	0
Dewi Llwyd AFS	12	8	1
Geth a Ger	0	2	0
Tudur Owen	7	7	0

Table 93 Number of standard, intermediate and non-standard tokens of (ei.f) in each programme

The other three programmes have scores that are relatively close to one another. *Post Prynawn* has the highest mean index score at 158, while *Dewi Llwyd AFS* and *Tudur Owen* are at 152 and 150 points respectively. A look at the distribution of aspirate, soft and radical variants shows that the soft mutation was frequent in all programmes, which suggests that even participants in the more formal programmes did not apply the standard AM mutation in many cases. There was only one example of the radical, which was used during *Dewi Llwyd AFS*.

8.4.9 Results - Code-Switching

This section presents the results for the code-switching analysis. Figure 35 presents the results of a simple mean calculation of the proportion of English words in each radio programme.

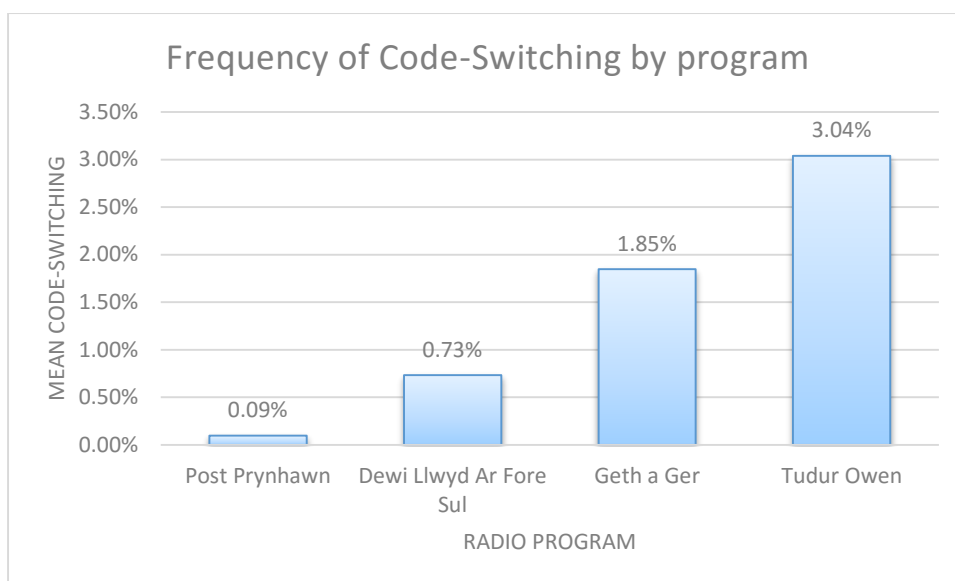


Figure 35 Proportion of code-switching by radio programme

Radio Programme	Welsh Words	English Words
Post Prynawn	24630	34
Dewi Llwyd AFS	33353	313
Geth a Ger	25034	510
Tudur Owen	30727	991

Table 94 The total numbers of Welsh and English words in each programme

The most formal programme, *Post Prynawn*, has the lowest mean frequency of code-switching, with only 0.09% switched words. *Dewi Llwyd AFS* has the second lowest mean at 0.73%, while *Geth a Ger* has a mean of 1.85%. The most informal programme, *Tudur Owen*, has the highest proportion of code-switching at 3.04%. Figure 36 shows the results calculated in the same way as the *Siarad* corpus mean analysis, as an average of each speakers' mean use of code-switching averaged across all three episodes of each programme. This allows the results to be compared with the *Siarad* corpus mean:

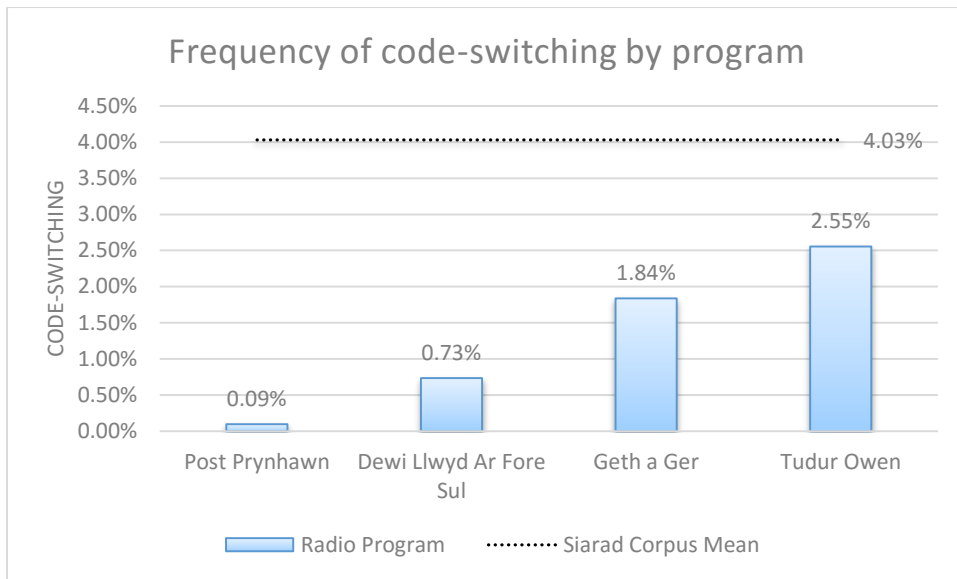


Figure 36 Average of means for all individual speakers averaged across all episodes

Presented in this way, the data still has a similar profile, with use of code-switching increasing gradually along the stylistic axis. However, the gap between *Post Prynawn* and *Dewi Llwyd AFS* is now larger, while the gap between *Geth a Ger* and *Tudur Owen* is smaller. The results also suggest all radio programmes used less code-switching on average than the speakers in the *Siarad* corpus. However, this impression may be partly given by the impact of various guest speakers on *Tudur Owen*, who used considerable less frequent code-switching than the three regular presenters.

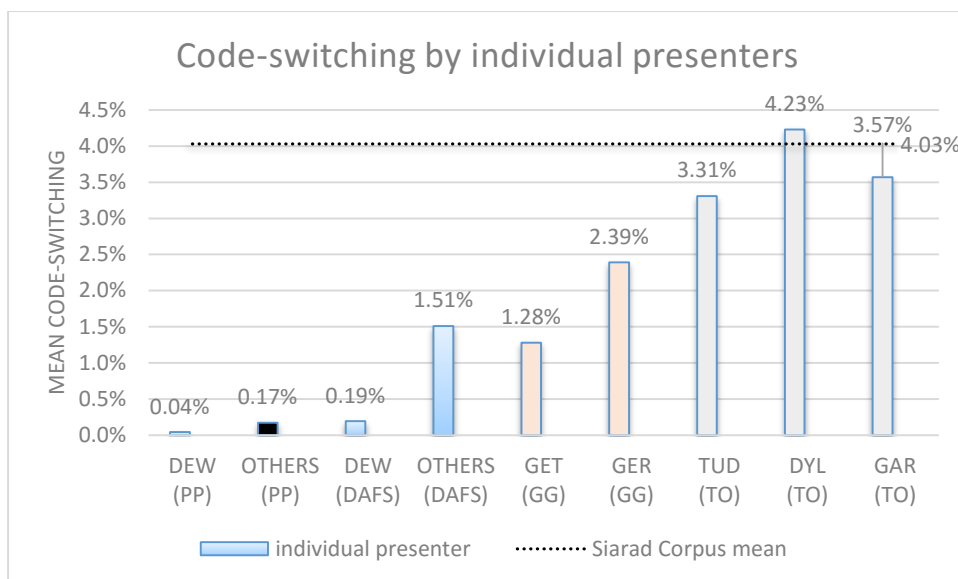


Figure 37 Code-switching means for each individual presenter

To compensate for this, figure 37 shows mean individual percentages of code-switching for each of the main presenters in each programme (i.e. those who were always present each episode, and excluding guests). It also shows the mean representing all contributors on *Post Prynhawn* and *Dewi Llwyd AFS* (i.e. all participants except for DEW). DEW's use of code-switching in *Post Prynhawn* is negligible at 0.04%, but rises slightly to 0.19% in *Dewi Llwyd AFS*. The mean code-switching of contributors to *Post Prynhawn* is slightly higher than DEW's mean in the same programme at 0.17%, while the contributors on *Dewi Llwyd AFS* have a substantially higher mean at 1.51%. In *Geth a Ger*, the host GET has a mean of 1.28% while his co-presenter GER has a much higher mean of 2.39%. The presenters of *Tudur Owen* all have the highest code-switching means in the *Radio Cymru* corpus, with 3.31%, 4.23% and 3.57% for TUD, DYL and GAR respectively. With regards to the *Siarad* corpus mean of 4.03%, it becomes clear that the three 'core' presenters of *Tudur Owen* are relatively close to the population average for code-switching use, with DYL in particular code-switching at a slightly higher rate than the community mean.

8.5 Results - Summary

The results of the analysis of the radio data seems to show that each variable is stratified stylistically in distinct ways.

Three variables arguably follow the stylistic stratification of the programmes; the aspirate mutations (a) and (â/gyda), and code-switching.

- The distribution of code-switching among the programmes was generally linear, with less formal programmes exhibiting proportionally more frequent code-switching than more formal programmes.
- Standard mutation of (â/gyda) was far more frequent in the two most formal programmes, with *Post Prynawn* exhibiting more standard forms than *Dewi Llwyd AFS*. Standard variants of the variable were used once on both informal programmes, although the frequency of tokens was very low.
- As with (â/gyda), the use of standard mutation following (a) was used much more frequently in the two most formal programmes *Post Prynawn* and *Dewi Llwyd AFS* than in the two least formal programmes *Geth a Ger* and *Tudur Owen*. Also, standard use of (a) seemed to be proportionally slightly more frequent in *Geth a Ger* than in *Tudur Owen*.

Four other variables were stratified in such a way that the social axis factors seemed to take precedent over stylistic ones:

- (o) and (ei.m) were mutated highly standardly across all programmes, although *Geth a Ger* (the lowest programme on the social axis) used slightly fewer standard forms than the other programmes.

- (am) was mutated very standardly on the two most formal programmes *Post Prynawn* and *Dewi Llwyd AFS*, and was mutated slightly less standardly by *Tudur Owen*, with *Geth a Ger* yet again possessing the lowest standard mutation mean.
- (fy) was used highly standardly in both the formal programme *Dewi Llwyd AFS* and the most informal programme, *Tudur Owen*, while the standard mutation mean in *Geth a Ger* was very low by contrast.

Two variables exhibit distributions which seem to contrast with the two patterns above.

- (yn) is used highly standardly in the two most formal programmes *Post Prynawn* and *Dewi Llwyd AFS*. The third highest standard mean was in *Tudur Owen*, where the tendency towards standard mutation is noticeably higher than the population average, while *Geth a Ger* had the mean lowest index score, which was still higher than the *Siarad* corpus mean.
- (ei.f) is relatively unique, as all programmes analysed (*Geth a Ger* was excluded) showed near equivalent mean index scores, including the mean scores for the *Siarad* Corpus.

The next chapter will take a closer look at the transcripts of the *Radio Cymru* corpus data, and will attempt to situate the use of the variables in discourse strategies used by the presenters themselves.

9. Qualitative Analysis - Discourse strategies for Code-Switching and Mutation

This chapter will focus on understanding stylistic variation from a different perspective, using the transcripts of the radio programmes to try and cast light on whether the varying frequencies of code-switching and mutation can be linked to particular discourse strategies in the data. I will argue that, particularly in the case of code-switching, which is strongly bound up with pragmatic and discourse related functions, the data cannot be appreciated fully without some textual analysis. I have identified five strategies under which code-switching and mutation related style shifting seems to occur in the radio data: retracing (9.1) flagging (9.2), increasing comprehension (9.3) direct speech (9.4) and humour (9.5). I argue that each strategy relates more or less to each programme depending on its level of formality, format and the values and aims – or the institutional style - underpinning each programme.

9.1 Retracing

In the previous chapter the possibility was discussed that the presenters of *Geth a Ger* in particular were intentionally suppressing their use of English in order to conform to a perceived appropriate level of formality for their programme. While transcribing the data for this programme, I noticed a particularly striking pattern where the hosts seemed to repair certain constructions – most notably those with English language material. Retracing here involves the production of an alternative form by a speaker intending to ‘correct’ an earlier utterance which is perceived to be inappropriate by the speaker. I consider this to be distinct from the strategy of ‘increasing comprehension’ discussed in section (9.3), where speakers produce an alternative form in order to clarify the meaning of an utterance, rather than to take its place. I will argue that the relative frequency of this strategy in each programme, as well as the nature of the linguistic material being retraced, provides us with insight into the process of style shifting observed during the analysis of the radio data. The majority of the instances of retracing

in the *Radio Cymru* corpus involve the reformulation of English language content to Welsh, although there is at least one case where the opposite may be taking place.

Programme	English to Welsh Retraces	Welsh to English Retraces
Post Prynawn	0	0
Dewi Llwyd AFS	2	0
Geth a Ger	12	0
Tudur Owen	3	1

Table 95 Comparison of the frequency and direction of the retracing of code-switching strategy across all four programmes

Table 95 provides totals of frequency, as well as direction, of retracing of English or Welsh constructions across all four programmes. One pattern that is particularly clear is that speakers almost always retrace material from English to Welsh, rather than from Welsh into English. Another interesting finding is that the programmes clearly differ in frequency of retracing. *Post Prynawn* has no retracing whatsoever, while *Geth a Ger* has a substantially higher frequency than any other programmes. I will now provide examples of the use of retracing in each programme.

9.1.1 Retracing - Geth a Ger

With twelve instances of retracing in the recorded material collected in this thesis, *Geth a Ger* has the most frequent use of this strategy in any of the radio programmes analysed. An example representative of this pattern is shown in extract 43. Here, GET has spent the previous turns introducing the next song on the playlist, and then directly addresses listeners with the English word *enjoy*, before playing the track:

*GET: enjoy@s:en .

(5 second pause with music)

*GET: mwynhewch dw i feddwl .

%eng: mwynhewch (enjoy) I mean

Extract 43 retracing from English to Welsh after a short pause in *Geth a Ger* (English → Welsh)

There is then a pause of approximately five seconds as the song begins to play, at which point GET interrupts the music to retrace the English imperative *enjoy* to its Welsh equivalent *mwynhewch*.

On a few occasions, the presenters go so far as to actually apologize for using English equivalents as they retrace to Welsh. In the following example, GER apologizes for using the English word *live*, before producing the Welsh equivalent *yn fyw* twice:

*GER: a oedden nhw yn wych live@s:en doedden ?

%eng: and they were great live weren't they ?

*GER: yn fyw sorry@s:und .

%eng: yn fyw (live), sorry

*GER: yn fyw .

%eng: yn fyw (live)

Extract 44: Retracing: live → yn fyw (from English to Welsh)

Similarly, in extract 7, during an exchange about television programmes, GER begins an utterance using the English word *series*, but abruptly stops and apologizes, before producing the Welsh equivalent *cyfres*:

*GER: am un series@s:en [///] sorry@s:und cyfres dw i yn trio ddeud .

%eng: *for one series, sorry I mean cyfres (series)*

Extract 45: Retracing: series → cyfres (from English to Welsh)

At times the struggle to produce Welsh equivalents for relatively low frequency English words causes relatively serious breaks in the flow of the programme. In the following extract, GER struggles to recall the Welsh equivalent of the English term *Civil Rights Movement*.

*GER: oedd y ddau ohonyn nhw yn (.) xxx yn frwd yn dilyn bod pawb yn +...

%eng: *both of them were keen that everybody in the...*

*GET: ++ +< civil@s:und rights@s:en movement@s:en .

*GER: yr civil@s:und rights@s:en .

%eng: *the civil rights...*

*GER: o'n i (y)n trio meddwl be oedd o yn Gymraeg &=laugh .

%eng: *I was trying to think what it was in Welsh!*

Extract 46: GER attempting to retrace a Welsh equivalent for *civil rights movement* in *Geth a Ger*

GER begins to trail off at the end of the first utterance when he cannot produce the Welsh equivalent, at which point GET seems to try and complete his utterance with the English equivalent *civil rights movement*. GER ultimately makes the metacomment “*I was trying to think what it was in Welsh*” before breaking down in laughter, and seems to give up on the

repairing process. The fact that the presenters are so intent on using Welsh terms that they would allow their discourse to break down in this way provides ample evidence that they are exerting considerable effort to avoid the use of English, even if doing so would come much more naturally to them.

Another interesting aspect of *Geth a Ger*'s retracing strategies seems to be the inclusion of mixed morpheme words in the category of undesirable material which needs to be retraced. In extract 5, GET retraces the mixed morphology word *smell_io* (eng. *to smell*), which has an English root and a Welsh verb ending, to a word with only Welsh morphemes, *arogli*. Integrated verb borrowings like *smell_io* are very common in colloquial Welsh, as described in Stammers (2010), and GET's retracing here seems to show that he believes such forms to be inappropriate for the context of *Geth a Ger*.

*GET: (y)dy headphones@s:en chdi fan (y)na Ger@s:und yn <smell_io@s:en+cy neis> [///] arogli neis?

Ger, do your headphones over there smell... arogli (smell) nice?

Extract 47: Retracing from a mixed morpheme (Welsh and English) word to a Welsh word in *Geth a Ger* (Welsh + English → Welsh)

In the next extract, GER flags the mixed morpheme word *inspirio* and humorously asks the public for help in supplying the 'Welsh' equivalent of the word. His co-presenter GET eventually provides the Welsh equivalent *ysbrydoli*, on his behalf.

*GER: a (ei) bod o (we)di bod yn arwr i xxx **inspire_io@s:en+cy** nhw .

%eng: *and that he had been a leader [...] inspiring them*

*GET: +< o_k@s:und .

*GER: **dw i ddim yn sure@s:und be ydy (y)r gair Cymraeg .**

%eng: *I'm not sure what the word is in Welsh*

*GER: text_iwch@s:en+cy ni chwech saith pum cant be (y)dy inspire_io@s:cy+en yn
Gymraeg .

%eng: *text us on 67500 if you know what “inspire-io” is in Welsh*

*GER: xxx (ei) bod o (we)di wneud ffrind da efo Kubhlai@s:und [/]
Kubhlai_Kahn@s:und o Mongolia@s:und wnaeth ddechrau y Wan@s:und
dynasty@s:en .

%eng: *[...] that he had been good friends with Kubhlai Kahn from Mongolia who
started the Wan dynasty*

*GET: +< **ysbrydoli** .

%eng: ysbrydoli (*inspire*)

Extract 48: GET retracing GER’s mixed morpheme word to a Welsh equivalent in Geth a Ger
(English --> Welsh)

9.1.2 Retracing – Tudur Owen

Rhaglen Tudur Owen is notable both in having far fewer instances of retracing than Geth a Ger, but also for at least one example where retracing is arguably taking place in the reverse direction, from Welsh to English. In terms of English to Welsh retraces, these are similar to the type seen in Geth a Ger, although they are never flagged with apologetic metacommentary.

*DYL: hyd yn oed gwell os mae (y)na deledu efo gêm peldroed Liverpool@s:und ac Arsenal@s:und **on@s:en** [///] **ymlaen** hefyd

%eng: *even better if there's a television with the Liverpool and Arsenal football game on... ymlaen (on) too*

Extract 49: Retracing in *Tudur Owen* on – ymlaen (English → Welsh)

In Extract 49, DYL retraces the English word *on* to the Welsh equivalent *ymlaen* in a smooth switch, without any pause or metacommentary. The initial use of the English word *laptop* in extract 12 also lacks a metacommentary, but is lightly flagged by the place holding interactional marker *er*, and the conjunction *a* (eng: and) which may indicate that TUD feels that he has made a mistake here.

*TUD: achos dw i (ddi)m yn gorfod cael **laptop@s:en** er@s:und a **gliniadur** i feddwl am ddeunydd ti (y)n gweld .

%eng: *because I don't need to have a laptop... and a gliniadur (laptop) to come up with material you see*

*TUD: er@s:und gyfeillion mae Dyl_Mei@s:und (we)di anghofio ei **liniadur** heddiw

%eng: *friends, Dyl Mei has forgotten his gliniadur (laptop) today*

Extract 50: Retracing in *Tudur Owen* laptop – gliniadur (English – Welsh)

TUD immediately switches to use the Welsh equivalent immediately afterwards, and maintains this pattern in the subsequent utterance, where he says *gliniadur* again.

TUD: ac oedd o (y)n [/] oedd o (y)n sôn am yr um@s:und [/] <yr timod **yr lladd** sydd yn> [///] yr **murder@s:en rate@s:en** sydd yn yr Unol_Daleithau .

%eng: *and he was... he was talking about the um, the you know, y lladd (the killing),
the murder rate in the United States*

Extract 51: Retracing in *Tudur Owen Lladd* – Murder rate (W-E)

In extract 51 TUD seems to retrace in the opposite direction to that normatively expected, correcting a broad Welsh equivalent *y lladd* (“the killing”) to an English language phrase ‘*murder rate*’. Again, as this is a smooth unflagged switch, it is difficult to tell what the intention of the speaker was at the time. However, given the higher frequency of code-switching observed in the previous chapter and the subversive nature of the programme, the possibility that speakers in *Tudur Owen* might sometimes show a preference for English equivalents over Welsh ones is intriguing.

9.1.3 Retracing - Dewi Llwyd AFS

The two instances of retracking for *Dewi Llwyd AFS* surprisingly (given his extremely low rate of code-switching) included one by Dewi Llwyd himself, where he repairs an English place name to a Welsh equivalent:

*DEW: ac wrth_gwrs oedd y fagwraeth yma yn digwydd <yn **Exeter@s:en**> [///] yng
 Nghaerwysg ac mewn byd tra wahanol i (y)r un dach chi yn byw ynddo fo ar
 hyn o bryd .

%eng: *and of course this upbringing was happening in Exeter... in Caerwysg (Exeter)
 and in a totally different world to the one you live in at the moment*

Extract 52: Retracing in *Dewi Llwyd AFS Exeter* – Caerwysg (E->W)

This could, arguably, be due to a lapse in concentration facilitated by the generally more relaxed level of formality seen in this programme, as compared with *Post Prynhawn*, where retracking did not take place in the recorded data. It does however, serve to indicate that the

same process of ideological pressure and audio monitoring to avoid English language material is operating upon DEW and other contributors on *Dewi Llwyd AFS*, although given their higher social prestige they may be generally more able to avoid producing English forms in the first place.

9.1.4 Retracing - Summary

In conclusion, although presenters on *Tudur Owen* sometimes seem to also retrace from Welsh to English, they do so on a far lesser scale than in *Rhaglen Geth a Ger*, and also with much less flagging. *Post Prynawn*, being more tightly structured, exhibits no retracing in the data set examined, while the more informal *Dewi Llwyd AFS* showed some retracing, which was exclusively from English to Welsh. In general the analysis of retracing seems to indicate two things – first, that *Radio Cymru* presenters generally seem to consider Welsh equivalents more appropriate for the context than English ones, and that retracing would seem to be a strategy to ‘correct’ for any use of inappropriate English forms. Secondly, the programmes vary in the frequency of the strategy, with *Geth a Ger* using it very frequently, and tending to flag retracing with apologetic metacommentary, while other programmes use it relatively infrequently.

9.2 Flagging

In this section I will be discussing the flagging of individual words that speakers in the radio data wish to draw attention to. The assumption will be that speakers draw attention to words considered somehow incongruous or remarkable within the context of the programme, thus indicating to some extent what the underlying values of the programme are. For example, if speakers on a programme regularly draw attention to English words in a negative way, we might assume that the underlying values are connected with linguistic purism. If they draw attention to Welsh words however, particularly in a mocking way, this might be considered indicative of a more subversive anti-purist ideology.

Programme	Welsh words flagged	English words flagged
Post Prynawn	0	1
Dewi Llwyd AFS	0	4
Geth a Ger	6	3
Tudur Owen	4	0

Table 96 Frequency and language of flagging across all programmes

In terms of the distribution and direction of flagging, which is presented in table 96, there is a clear divide between the two most formal programmes and the two most informal programmes. The flagging that occurs in *Post Prynawn* and *Dewi Llwyd AFS* is exclusively of English terms, or words that are perceived as being too English, which seems to include mixed words with both Welsh and English morphemes. The flagging in *Geth a Ger* on the other hand is particularly interesting in that it seems to be divided between flagging both English words and a particular kind of low frequency Welsh words. Presenters on *Tudur Owen* exclusively flagged Welsh words - never English ones - and specifically drew attention towards words which I argue are low frequency or archaic.

9.2.1 Flagging - Post Prynawn

Post Prynawn had one instance of flagging in the three hours transcribed, which is shown in extract 53 below.

*LLY: ond bod yna le i human@s:en error@s:en oedden nhw yn galw fo yn y system@s:und felly ynde .

%eng: but that there is room for human error - as they called it - in the system

Extract 53 Flagging English term 'human error' in *Post Prynawn*

The speaker LLY flags the English term *human error* with the metacomment *as they called it*, which arguably may serve to distance the speaker from the English words used, by ascribing it to another speaker, the abstract ‘they’. It is interesting to note that as English is generally absent in *Post Prynghawn*, any use of English is of course likely to be highly salient to both speakers and listeners. LLY may be claiming the right to use English here in the interest of objectivity and accuracy in reporting on other speakers’ words.

9.2.2 Dewi Llwyd AFS

As in *Post Prynghawn*, the flagging in *Dewi Llwyd AFS* is exclusively of English terms. In extract 54, PRY uses the English term *zero tolerance approach*, which is heavily flagged by the word ‘well’ and the metacomment ‘*as they say over and over in the article*’.

PRY: ac [/] ac be maen nhw yn ddeud ydy fod angen mynd i yr afael â (y)r broblem
yn ddifrifol iawn a cael well@s:und **zero@s:en tolerance@s:en
approach@s:en maen nhw yn (ei) ddeud drosodd a drosodd** yn yr erthygl

%eng: *and what they're saying is that the problem needs to be tackled seriously and
that a zero tolerance approach is needed, as they say over and over in the article*

Extract 54 Flagging the English term ‘zero tolerance’ in Dewi Llwyd AFS

In attributing the English term to another source, *they*, it could be argued that PRY is trying to distance himself from the use of the words, and to justify his use of them. A similar example is shown in extract 55, where the contributor CAT is about to quote an article headline from an English newspaper. Before making the switch to English to read the headline, she flags her intentions beforehand, using a metacomment to say that it is *werth ei ddarllen yn Saesneg* (“worth reading in English”):

*CAT: ond um@s:und yn yr erthygl olygyddol ar dudalen dau ddeg pump hynny yw mae yna un frawddeg dw i (y)n meddwl sydd yn ddiddorol iawn a **sydd yn werth ei ddarllen yn Saesneg** yndyfe .

%eng: *but in the editorial page on page twenty five there's one sentence which is particularly interesting and which is worth reading in English*

*DEW: +< <oes dw i (we)di> [=! laugh] gweld honna fy hun ia .

%eng: *yes, I've seen that one myself!*

*CAT: +" encouraging@s:en our@s:en neighbours@s:en to@s:en learn@s:en our@s:en language@s:en is@s:en not@s:en a@s:en punishment@s:en but@s:en a@s:en positive@s:und bene(fit)@s:en [/] benefit@s:en .

Extract 55 A participant flagging an intersentential shift to English in Dewi Llwyd AFS

As with extract 54, it seems that CAT feels the need to signpost and apologize for her use of the English language. As well as flagging purely English words, contributors also flag words which have mixed morphemes of both Welsh and English origin. In the following extract from *Dewi Llwyd AFS*, the birthday guest PET flags the mixed morpheme word *gorcompensateo* (eng: “to overcompensate”), which has two Welsh morphemes, *gor-* and *-io*, and one English one, *compensate*, with the metacomment *os oes yna fath gair yn Gymraeg* (eng. “if there is such a word in Welsh”), which is also followed by laughter:

*PET: yeah@s:und well@s:und w i (y)n sure@s:und (f)y mod i wedi gor_compensate_o@s:cy+en+cy **os oes <yna fath gair yn [/] yn>** [=! laugh] **Gymraeg** .

%eng: *yeah well I'm sure that I've gor-compensate-o (overcompensated), if there is such a word in Welsh!*

Extract 56: Flagging of mixed morpheme word *gorcompensateo* in *Dewi Llwyd AFS*

This would seem to provide further evidence that Welsh speakers consider such constructions English words, even though non-Welsh speakers would obviously not perceive them as such. In summary, the institutional style of *Dewi Llwyd AFS*, like *Post Prynawn*, seems to reject use of English words, given that speakers tend to flag and apologize for using them.

9.2.3 Geth a Ger

One of the most interesting aspects of *Geth a Ger* is that presenters seem to flag both English and Welsh words. Where Welsh words are flagged, there a tendency for them to be a specific class of words, and particularly ones that are low frequency or archaic, and flagging is often carried out in a humorous or mocking way. In the extract below, GET uses the low frequency word *herio*, and then repeats it after a short pause.

*GET: yn dilyn ni heno am hanner (we)di saith er@s:und mae Cymru yn herio yr Alban

%eng: *following us at half past seven, Wales challenges Scotland*

(3 second pause)

*GET: **herio .**

%eng: challenges

Extract 57: Flagging of low frequency Welsh word *herio* in *Geth a Ger*

In another extract, while talking about going to see a performance by the magician Derren Brown, GET uses the low frequency Welsh word *consuriwr*. The word is both said with a particular emphasis and met with laughter by his co-presenter GER.

*GET: dw i (we)di prynnu pump ticed i mynd i weld sioe (.) um@s:und (.) yr
consuriwr Derren_Brown@s:und .

%eng: *I've bought five tickets to go and see the... um... consuriwr (magician) Derren
Brown's show*

*GER: &=laugh .

Extract 58 Flagging low frequency Welsh word *consuriwr* in Geth a Ger

9.2.4 Flagging - Tudur Owen

The flagging in *Tudur Owen* seems to differ again from that observed in all other programmes, in that English words are never flagged. Instead, the presenters tend to flag low frequency, antiquated, technical or dialectal Welsh words, usually in a humorous or mocking way. In extract 19 below, co-presenter GAR uses the low frequency technical term *morwrol* (eng: “nautical) in a normal unflagged delivery while describing a building:

*TUD: yr un mawr gwyn (y)na .

%eng: *the big white one*

*GAR: yr un morwrol ydy hwnnw dw i meddwl ynde .

%eng: *that's the nautical one I think, isn't it*

*TUD: +< morwrol &=silly_voice .

%eng: *nautical!*

Extract 59: Flagging low frequency Welsh word ‘morwrol’ in *Tudur Owen*

His use of this word is immediately seized upon by the host TUD, who repeats *morwrol* in a silly voice, presumably in order to mock the use of a relatively technical and obscure word. In

the next extract, in the middle of a relatively serious section on a charity event he attended, he flags the low frequency word *bwrlwm* (hubbub/commotion) by shouting it:

*TUD: be oedd yn ddiddorol wrth_gwrs mae nhw (y)n roid yng nghanol <yr yr> [/] y
bwrlwm [=! shouts] oedd (y)na er@s:und videos@s:und bach o lle mae (y)r
arian yn mynd .

%eng: *what was interesting of course is that in the middle of the commotion there were
little videos showing where the money was going*

Extract 60: Flagging low frequency Welsh word ‘*bwrlwm*’ in *Tudur Owen*

In extract 69, TUD flags a standard (and perhaps rather antiquated) variant of *oherwydd* (eng: because) *o’r herwydd* which he inadvertently uses, again using his usual tactic of repeating the construction in a high pitched comical voice:

*TUD: er@s:und <o (y)r &h> [/] <o (y)r herwydd> [&=silly_voice] (.) oedd o (y)n
gyfnod oeddwn i yn fy ardegau (.) hmm@s:und eitha hwyr

%eng: *o’r herwydd (because)... o’r herwydd! it was a time when I was in my late
teens...*

Extract 61 Flagging low frequency Welsh construction *o’r herwydd* in *Tudur Owen*

The ridiculing of Welsh words isn’t necessarily restricted low frequency, elevated terms. In the following extract TUD mocks his co-presenter DYL for using the Northern Welsh dialect word *cuddiad* (standard form: *cuddio* “to hide”) when referring to the pop song *Chwarae Cuddio*:

*DYL: Chwarae_Cuddiad@s:und .

*GAR: &=laugh .

*TUD: +< Chwarae@s:und [/] Chwarae_Cuddio@s:und .

*TUD: **dim Chwarae_Cuddiad@s:und .**

%eng: *not "chwarae cuddiad"*

*TUD: Chwarae_Cuddio@s:und .

*DYL: be (y)dy gwahaniaeth ?

%eng: *what's the difference?*

*TUD: well@s:und mae (y)na d ychwanegol heb fod angen o i ddechrau efo (h)i .

%eng: *well there's an extra d which isn't needed to start with*

*DYL: +< silent@s:en d .

*GAR: &=laugh .

*TUD: &=laugh .

Extract 62 Flagging Welsh dialect word *cuddiad* in *Tudur Owen*

9.2.5 Flagging – Summary

All four programmes vary in terms of whether they tend to flag either Welsh or English words. As with retracing, *Post Prynawn* seems to be formal enough that flagging could seem unprofessional and disruptive, while speakers in *Dewi Llwyd AFS* predominantly flag only English words, probably because they are seen as being incongruous with the programme's purist institutional values. Low frequency words such as *allyriadau carbon* are used often on *Post Prynawn* and *Dewi Llwyd AFS* without being flagged, again presumably because they are considered appropriate and congruous with the value system of both programmes. Flagging patterns in *Geth a Ger* are more complex, perhaps because the presenters seem torn between purist and subversive values. In almost equal measure the speakers both ridicule low frequency Welsh words and apologize for and retrace English words that they use. This arguably justifies

the categorization of *Geth a Ger* as being more formal than *Tudur Owen*, where by contrast the speakers never flag English words, instead tending to flag low frequency standard and dialectal Welsh constructions.

In order to justify my labelling of the Welsh words that were flagged as ‘low-frequency’, I have used the CEG corpus (N. C. Ellis et al 2001), a million word database of Welsh words which allows for frequency counts, as well as the *Siarad* corpus to analyse each word by frequency.

The words flagged have the following frequencies in both corpora:

Flagged word	Word category	CEG Frequency (per million words)	<i>Siarad</i> frequency (per 400,000 words)
consuriwr	antiquated	2	0
morwrol	technical	11	1
bwrlwm	antiquated	17	0
herio	antiquated	43	0
gliniadur ²⁰	technical	0	0
cuddiad	dialectal	1	5

Table 97 Flagged words and relative frequencies in the CEG corpus

As can be seen in table 97, the words have a generally low frequency, although *herio* and *bwrlwm* are rather more frequent than I expected in CEG, while they are absent in the *Siarad* corpus. All of the words are in the lowest frequency band used by Gathercole and Thomas

²⁰ It should be noted that the term *gliniadur* (“laptop”) may have gained wider currency after the CEG corpus was published, given the relatively recent ubiquity of laptop computers. Because of this, its absence in that corpus is unsurprising.

(2009) of below 25 occurrences per million words, with the exception of *herio*, which fits in the penultimate 25-49 frequency band. The most frequent item in the *Siarad* corpus is the dialect word *cuddiad*, which DYL was mocked for using. The discrepancy may be partially due to CEG being largely a corpus of written texts, while *Siarad* is entirely based on vernacular speech, and is more recent. Although the *Siarad* corpus is smaller, I would be inclined to consider its word frequencies as more indicative of the true frequency of these words in colloquial speech.

9.3 Reinforcing comprehension

Another strategy, most typical of the more formal programmes, is the statement of an English equivalent immediately following the Welsh one, in order to clarify its meaning. This is distinct from retracing in the sense that the speakers are not trying to erase or ‘take back’ their previous code choice, but instead are deliberately placing an English equivalent following a Welsh word in order to increase their interlocutors’ and audience’s comprehension of it. The most reasonable explanation for this is related to the speakers’ consideration of their audience. If presenters in the more formal programmes understand the general norm that the typical Welsh speaking audience may not have a full grasp of low frequency equivalents, but feel themselves constrained to avoid English equivalents as much as possible, this strategy allows a way to fulfil both obligations. In a similar manner to the way in which speakers seem to feel that direct speech exempts the speakers from the usual social cost of code-switching, reinforcing low frequency Welsh words with English equivalents may be seen as permissible because speakers are using code-switching in a potentially constructive, helpful way.

Programme	Welsh words reinforced	English words reinforced
Post Prynawn	1	0
Dewi Llwyd AFS	3	1
Geth a Ger	1	0
Tudur Owen	0	0

Table 98 Frequency and language of words reinforced by radio programme

9.3.1 Reinforcing comprehension - Post Prynawn

In extract 63, the host of *Post Prynawn*, DEW, introduces a story about an accident involving a zip wire at a park. The Welsh equivalent for zip wire “*wifren wib*” is unlikely to be widely known by Welsh speakers, as it is in neither the main Welsh dictionaries, nor the *Siarad* corpus. DEW seems to acknowledge this by using the English equivalent first before adding the words “*neu wifren wib*” (“or wifren wib”) to clarify:

*DEW: a (y)r brif stori heddiw crwner yn galw am lunio rheolau newydd ar gyfer atyniadau zip_wire@s:en **neu wifren wib** wedi marwolaeth bachgen ym mharc Gelli_Gyffwrdd ger y Felinheli@0 ddwy flynedd yn_ôl .

%eng: *and the main story today: a coroner calls for the creation of new rules for zip wire - or wifren wib (zip wire) - attractions following the death of a young boy in Gelli Gyffwrdd park near Felinheli two years ago*

Extract 63 Reinforcing comprehension of low frequency Welsh word in *Post Prynawn*

This is particularly significant due to the extremely rare use of English terms by DEW in *Post Prynawn*, which makes this concession highly marked. Interestingly, DEW uses the Welsh equivalent without clarification earlier in the same programme, while another correspondent

who is interviewed for the same programme uses the English term exclusively, without flagging. This perhaps suggests that presenters in very formal contexts can adopt a variety of strategies for dealing with low frequency Welsh equivalents, as was suggested by Ball et al (1988).

9.3.2 Reinforcing comprehension - Dewi Llwyd AFS

Dewi Llwyd AFS has the most frequent use of this strategy, with speakers primarily using it when introducing a low frequency Welsh term. For example, in extract 64, a guest on the newspaper discussion segment follows his initial use of the term *ailwylltio* (referring to the practice of repopulating an area with previously extinct species) with its English equivalent term *re-wilding*:

PRY: <a mae o> [/] well@s:und mae o yn deud na yr ateb ydy ailwylltio y
rewilding@s:en (y)ma

%eng: *well, he says that the answer is ailwylltio (re-wilding), this re-wilding”*

Extract 64 reinforcing comprehension of low frequency Welsh word in *Dewi Llwyd AFS*

As the term *ailwylltio* is highly technical, with a frequency of zero tokens in the CEG corpus, it can be inferred that the implied purpose of this switch is to provide the audience with the English translation so that no one is left unclear as to the meaning of the concept. In at least one other case, the increasing comprehension strategy moves the other way, from English to Welsh:

MIK: a sylw i (y)r ddau quaterback@s:und **sef y maswr** sydd yn llywio popeth

%eng: *And some attention has been given to the two quarterbacks, the fly-halves who guide everything*

Extract 65 Reinforcing comprehension of American English word in *Post Prynawn*

In the quote above the speaker, a sports correspondent for *Dewi Llwyd AFS*, translates the English term quarterback into Welsh. This is interesting partly because the translation is a cultural one. Quarterback has no real Welsh language equivalent, perhaps because American Football isn't widely played in Wales. The speaker's strategy here however, is to adapt the word to the Welsh cultural context, using *maswr* (fly-half), which is a similar position in rugby. Rugby, being widely played in Wales, is used as an analogue which allows listeners to understand this relatively foreign concept. Although this switch involves the opposite direction in comparison with the strategy of Welsh to English switching, which we have already argued is intended to increase audience comprehension, here we can see that English to Welsh switching can perform the same function in some situations.

9.3.3 Geth a Ger – Reinforcing comprehension

Although the use of the reinforcing strategy is most prevalent in *Dewi Llwyd AFS*, there is one example from *Geth a Ger*, which is presented in extract 66 below. Here, during GER's *Gwers Ger* (*Ger's lesson*) segment, where the co-host provides a semi-serious lecture on various subjects, he introduces the highly technical term *goblygiadau cyfreithiol*, which he immediately follows with the English equivalent *legal consequences*:

*GER: aeth hi (y)n enwog ar y pryd de yn papurau newydd ac ar newyddion achos y goblygiadau cyfreithiol .

%eng: *it became famous at the time in the newspapers and on the news because of the goblygiadau cyfreithiol (legal consequences)*

*GER: legal@s:en consequences@s:en .

*GET: &=laugh .

*GER: **diolch i dad@s:und am honna .**

%eng: thanks to dad for that one

*GER: lle y legal@s:en consequences@s:en oedd wnaeth y crew@s:und i_gyd dynnu enwau allan o het i weld pwy oedden nhw yn mynd i fyta

%eng: where the legal consequences were that the crew all pulled names out of a hat to see who would be eaten

Extract 66 Welsh term *goblygiadau cyfreithiol* reinforced with English equivalent in *Geth a Ger*

This is met with laughter by host GET, before GER goes on to quip *diolch i dad am honna* (“thanks to dad for that one”). As with other examples of reinforcing, GER has produced the English equivalent of a low frequency Welsh term, apparently to clarify its meaning. But the implication of the ‘thanks to dad’ comment, and GET’s laughter in reaction to it, is that he could not have produced the term himself, and had to ask an older speaker to provide it before the programme was recorded.

9.3.4 Reinforcing – Summary

The increasing comprehension strategy was used most frequently in Dewi Llwyd AFS, and seems to be associated with more formal programmes in which there is a tension between the need to avoid English terms, and the problem that many Welsh speakers may not fully understand all Welsh equivalents for technical or low frequency terms. It may be that *Post Prynawn* does not use this strategy more frequently because it is too formal for this kind of digression, or that even the small amount of English involved may be considered inappropriate for the institutional style of the programme. *Geth a Ger* had only a single use of the strategy, which was notable for being centered on a low frequency Welsh term *goblygiadau cyfreithiol* which was explicitly given to the host by an older family member. This highlights the fact that the contributors of *Geth a Ger* themselves may not use the increasing comprehension strategy

more frequently because they don't generally use words that need to be explained to the audience in this way. *Tudur Owen* saw no usage of this strategy, which is unsurprising due to the highly subversive style of the programme, which would make this kind of didactic strategy inappropriate.

9.4 Direct speech

One of the most important pragmatic functions of code-switching in this dataset is direct-speech, the quotation of words attributed to one speaker by another. One reason why this is so relevant is the sheer frequency of direct speech in the *Radio Cymru* corpus. Direct speech occurs as a core element part of the format in the programmes which form the dataset. In *Post Prynawn* and *Dewi Llwyd AFS* for example, it usually takes the form of quotes and newspaper headlines. In *Geth a Ger* and *Tudur Owen*, where the format more closely resembles casual conversation, this usually involves quoting jokes or lines out of films and other media, often with a humorous intent. The style dimension of direct speech also raises particular question in the context of this study, since its use could be considered to be constrained by the original language in which an utterance was made. If speakers are attempting to accurately convey the words of someone who used English words, then it might be expected that they will shift to English in order to provide a truer account of what was said (see the discussion of Gal 1979 in section 4.3.1). Because of this, we might expect to see all speakers shifting to English when relaying things that were originally said in English.

Looking at the data however a more complex picture emerges. The impact of the need to communicate objectively and accurately is apparent in that direct speech is by far the most important motivator towards code-switching in the two most formal programmes. A large proportion of the switched material in *Post Prynawn* in *Dewi Llwyd AFS* are thus 'caused' by direct speech. The constraint against using Welsh equivalents while performing direct speech

is not categorical however, since speakers are able to translate or rephrase other people's speech - particularly if they are motivated enough to do so. In terms of the radio data, this is particularly clear in the most formal programmes, where the overwhelming norm is to avoid the use of English words. In *Post Prynawn*, when Dewi Llwyd quotes other speakers he generally translates their words to Welsh.

However, one of Dewi Llwyd's few uses of English in *Post Prynawn* is a quotation of the American Democratic Party's famous electoral campaign slogan "*it's the economy stupid!*" from the early nineties:

*DEW: + "it's the economy und .

*DEW: fel ddeudodd er und Bill_Clinton und ers lawer dydd ynde

%eng: "it's the economy", as Bill Clinton used to say

Extract 68 Direct speech of an originally English quote, from *Post Prynawn*

Here the quote is followed by an allusion to the source (supposedly Bill Clinton), which could be construed as a kind of flagging strategy to distance the speaker from the words. It may also be significant that the wordplay element of the quotation would not function as well in this context were it to be translated, so that the use of any Welsh equivalent is blocked. Effectively, it might be argued that if DEW wants to make this comment he does not have the option of using a Welsh equivalent. In *Dewi Llwyd AFS* direct speech frequently takes the form of quotations of newspaper headlines, particularly in the early newspaper discussion segment of the programme. During this segment, Dewi normally discusses events from the news with frequent allusions to specific newspaper headlines. However, in contrast to the tendency in *Post Prynawn* to translate direct speech to Welsh, in *Dewi Llwyd AFS* the pattern is more mixed, with more examples of repeating the original text as in the following extract:

*DEW: hynny yw yng ngeiriau yn y Saesneg felly bod ei gyfnod llwyddianus o wrth y llyw +"/.

%eng: *that is, in English, that his successful period at the wheel:*

*DEW: +" shows@s:en the@s:en benefits@s:en of@s:en playing@s:en the@s:en long@s:en game@s:en ynde .

Extract 69: Direct speech of an originally English source in Dewi Llwyd AFS, Flagged

In the example above, as in the majority of DEW's uses of English, the switch is heavily flagged. Conversely, there are also examples where DEW translates headlines into Welsh, as in the following extract:

*DEW: a (y)r penawd trawiadol yn y Mail@s:und +"/.

%eng: *and the striking headline in the Mail:*

*DEW: +" siaradwch Saesneg neu mi fyddwch chi yn colli eich budd_daliadau

%eng: *"speak English or you'll lose your benefits"*

Extract 70 Direct speech in Dewi Llwyd AFS, translating an originally English quote to Welsh

Although I have discussed the notion that direct speech is constrained by ideology from one direction and pragmatic considerations from the other, this kind of alternation seems to provide evidence that the overall level of formality (i.e. attention to speech) may still play a role in affecting the speakers' decision to code-switch or not. Table 99 shows the relative frequencies of direct speech translated into Welsh by contributors (i.e. English to Welsh) or maintained in the original English (English-English). Although the frequency of direct speech in *Post Prynawn* was much lower than in *Dewi Llwyd AFS*, the pattern shows that the norm of *Post*

Prynhawn is to translate direct speech into Welsh, while *Dewi Llwyd AFS* has a more mixed pattern with a trend towards the maintenance of English rather than translation.

Programme	Translated to Welsh	Maintained in English
Post Prynhawn	8	1
Dewi Llwyd AFS	11	29

Table 99 Comparison of frequency of direct speech translated to Welsh or maintained in English

In both *Geth a Ger* and *Tudur Owen*, the original language of direct speech utterances was frequently unclear, making it difficult to calculate how frequently the original language acted as a constraint.

9.5 Humour

Another pattern which has become apparent in the analysis of the data is the frequent co-occurrence of laughter with code-switching. Although we have already established that there seems to be a correlation between the frequency of laughter and that of code-switching in the data, a further question is whether humour might be associated with code-switching in a more intrinsic way. Humour often involves the violation of expectations and breaking of taboos (Siegel 1995, Woolard 1998), so it may not be surprising that on Welsh language radio, which as we have seen sometimes defines itself according to a value system which advocates linguistic purity, code-switching can be a source of humour due to its sheer incongruity and subversiveness.

Tudur Owen is the programme of which this phenomenon is most characteristic. The excerpt below is from a weekly portion of *Tudur Owen* called *Gloywi Iaith gyda Tudur Owen* ('*language improvement with Tudur Owen*'). This is a reference towards the advanced

language improvement courses offered by many Welsh speaking institutions to their staff. In the segment, a female presenter MAN reads a script that mimics a language learning CD or podcast, which the listener expects to lead to a Welsh word of the day. But in place of the expected Welsh word, there is the English equivalent. The juxtaposition of the pompous purist values of *gloywi iaith* with code-switching is used effectively to create tension, and ultimately a subversively humorous contrast:

*MAN: gloywi iaith gyda Tudur_Owen@s:und .

%eng: *language improvement with Tudur Owen*

*TUD: gair Cymraeg y dydd .

%eng: *Welsh word of the day*

*MAN: thank@s:en you@s:en .

*TUD: ddim isio .

%eng: *no need!*

*TUD: (ddi)m isio .

%eng: *no need*

*MAN: thank@s:en you@s:en .

*TUD: &=laugh oedd yna (ddi)m isio sure@s:und .

%eng: *there was really no need*

*TUD: ond er@s:und diolch anyway@s:en ia .

%eng: *but, er, thanks anyway yeah!*

Extract 73 Mocking purist Welsh values in *Tudur Owen*

In extract 73 the presenters discuss their guest GWY who is a producer for *Radio Cymru*. GWY's presence as a potential figure of authority on behalf of *Radio Cymru* is repeatedly used for humour. The presenters here jokingly refer to GWY's selection of Welsh language music for the playlist as a potential shaming tactic, since they have been playing some English language songs up to that point:

*TUD: +< wneud i fi deimlo (y)n euog rŵan .

%eng: *making me feel guilty now*

*GWY: bands@s:und fel (y)na .

%eng: *bands like that*

*TUD: teimlo (y)n euog rŵan .

%eng: *feeling guilty now*

*GAR: +< ia siarad am rai Cymraeg .

%eng: *yeah, talk about Welsh ones*

*TUD: +< <dan ni bod yn> [//] <dan ni yn wneud> [/] ia dan ni yn wneud rai sneusneg²¹

%eng: *yeah we've done some English ones*

*TUD: a mae Gwyddno@s:und wedi rhoid ni yn ein lle .

%eng: *and Gwyddno has put us in our place*

*TUD: cofiwch lle ydach chi boys@s:und .

%eng: *remember where you are lads*

²¹ *Sneusneg* is a humorous adaptation of the word *Saesneg* ("English"), which Tudur frequently uses jokingly when referring to the English language.

*TUD: BBC_Radio_Cymru@s:und .

Extract 74 Satirizing the idea of quotas for Welsh language musical content on *Tudur Owen*

In the following extract, TUD speaks to the sports correspondent Eleri Siôn about a previous interview in which they had conversed through the medium of English for *Radio Wales*, *Radio Cymru*'s English language sister station. After jokingly lavishing hyperbolic praise upon each other for having successfully conducted an interview together in English, Tudur then switches to English in an affected RP accent to claim that they should continue to speak together in English in the future. GAR, who plays the comedic role of the 'straight man', tries to interject to no avail:

TUD: wnes i siarad efo chdi diwrnod o blaen do ar Radio_Wales@s:und yn Saesneg .

%eng: *I spoke with you yesterday on Radio Wales in English didn't I?*

*ELE: do oeddet ti (y)n dda .

%eng: *yes, you were good*

*TUD: +< do .

%eng: *yes*

*GAR: +< Tudur@s:und .

*TUD: eh@s:und ti (y)n dda .

%eng: *you're good*

*ELE: +< oeddet ti (y)n dda (he)fyd .

%eng: *you were good too*

*TUD: +< ti (y)n dda .

%eng: *you're good*

*ELE: +< oeddet ti (y)n dda .

%eng: *you were good*

*TUD: +< dw i yn meddwl bod ni (y)n well yn Saesneg sti .

%eng: *I think we're better in English, you know*

*ELE: oeddet ti (y)n dda .

%eng: *you were good*

*ELE: +< <oeddet ti> [/] oeddet ti doniol .

%eng: *you were, you were funny*

*DYL: &=laugh .

*TUD: <dw i meddwl dylsen ni> [/] dw i meddwl dylsen ni anghofio busnes Cymraeg (y)ma .

%eng: *I think that we... I think that we should forget this "Welsh" business*

*GAR: +< wnewch chi eich dau stopio ?

%eng: will you two stop?

*TUD: <I@s:en think@s:en we@s:en should@s:en> [/] I@s:en think@s:en we@s:en should@s:en do@s:en this@s:en in@s:en English@s:en actually@s:en .

*ELE: +< nagyfe .

%eng: yes

Extract 74 Satirizing norms over the use of Welsh (and avoidance of code-switching) in *Tudur Owen*

This extract highlights the way in which Tudur Owen skirts the line of subversive anti-purist values, by playing with anti-Welsh sentiments: “*I think we should forget this ‘Welsh’ business*”, while also arguably satirizing people who would hold these values. The switch to English here, and the taking on the guise of an anti-Welsh persona, is a crucial part of the joke. This segment can also be appreciated more fully in light of other programmes’ avoidance of code-switching. By pretending that they do not understand these norms, and by transgressing them so brazenly, the presenters hold them up to question, pushing the boundaries of ‘acceptable’ behaviour on Welsh-language radio. The powerful incongruity between the presenters’ assertion of the superiority of English over Welsh, which contrasts sharply with the strong implicit messages of most other programmes (e.g. in *Post Prynawn*, *Dewi Llwyd AFS* and *Geth a Ger* to some extent) and of Welsh standard language ideology in general, leads to a comic effect similar to that described by Siegel (1995) among Fijians when switching to Hindi.

There are also numerous examples in the data where code-switching itself is actually a crucial part of a humorous exchange. In such cases, the joke would fail without the presence of the switch. The example below is from *Tudur Owen*, where DYL plays on the double meaning of the word *bar*, both as a Welsh established loan (a noun meaning a pub or café) and an English transitive verb (meaning to impede or block someone):

*TUD: achos mae y theatre@s:und ym Mangor@s:und yn mynd i fod yn theatr@s:und
Bryn_Terfel@s:und dydy .

%eng: *because the theatre in Bangor is going to be the Bryn Terfel Theatre isn't it*

*DYL: yndy .

%eng: *yes it is*

*GAR: bar@s:und Dyl_Mei@s:und .

%eng: *Dyl Mei's Bar*

*TUD: bar@s:und Dyl_Mei@s:und .

%eng: *Dyl Mei's Bar*

*DYL: dyna ydy (y)r instructions@s:en mae nhw (we)di cael [=! laugh] .

%eng: *those are the instructions they've been given*

*TUD: &=laugh ia .

%eng: *yes!*

*GAR: &=laugh .

*DYL: &=laugh .

Extract 75 Example of pun incorporating a code-switch as a necessary element

9.5.1 'Seriousness'

An interesting counter-example to the association between humour and non-standard language is the relationship between seriousness and standard forms. As an example of this I will discuss the sole standard use of aspirate mutation spoken by the host TUD in *Tudur Owen*. Of the 27 tokens of (a) during the three hours of recorded audio that he produces in *Tudur Owen*, TUD only uses the standard variant once, which is proportionally less than that used by the younger and less experienced hosts of *Geth a Ger*. Looking at the context of the utterance where this

standard token is used is highly revealing, because the tone is markedly different to a conventional segment of *Tudur Owen* (the relevant mutation is highlighted in bold in the second line of transcript):

*TUD: mae isio diolch yn fawr iawn i chi wrandawyr BBC_Radio_Cymru@s:und
a° **phawb** fuodd yn rhan o ddiwrnod Plant_Mewn_Angen ddoe .

%eng: *we need give a big 'thank you' to you, the listeners of BBC Radio Cymru, and
everyone who was part of Children In Need yesterday*

*TUD: dw i (y)n [//] mae (y)n deg i ddeud ei fod o wedi bod yn lwyddiant ysgubol .

%eng: *I'm... I think it's fair to say that it has been a sweeping success*

*TUD: achos mi oedd y cyfanswm +...

%eng: *because the total...*

*TUD: mi glywsoch chi falle ar y newyddion rŵan .

%eng: *maybe you heard on the news now*

*TUD: +, mae y cyfanswm cenedlaethol mae o (y)n torri record@s:und bob blwyddyn
sydd yn anhygoel .

%eng: *the national total breaks the record every year, which is incredible*

*GAR: +< mhm@s:und .

*TUD: pan ti (y)n meddwl amdan y sefyllfa ynde ariannol ar hyd a lled y wlad mae o
yn anhygoel .

%eng: *when you think about the financial situation across the country it's incredible*

Extract 76 The single use of standard mutation following (a) by the host of *Tudur Owen* in context, as he discusses a charity event

Here, the host is describing a charity event that he was involved in hosting. In contrast to the normally subversive and ironic tone, TUD speaks earnestly about the emotional impact that the event has had upon him, and expresses happiness about the success in reaching the donations target. The standard token of (a) appears in an utterance where TUD is thanking all those who took part and donated money. The fact that the sole standard token of (a) appears along with this brief change in tone, which is so distinct from the rest of the programme, may be evidence that (a) acts as a marker – a linguistic resource which speakers can, as with code-switching, draw upon to create stylistic meaning. It is also notable that TUD's whole address to the audience (which is longer than the excerpt presented here) contains no instances of code-switching at all, which provides further proof that both variables work in a similar way to convey contrasting meanings of irreverence and respectfulness. In this case, what TUD is conveying is seriousness, earnestness and authority, values which are not typically part of the show's institutional style, but which can be seen here in abundance.

9.6 Summary

This chapter has attempted to look in closer detail at the specific discourse related factors that shape the more abstract percentage scores that were measured in the previous variationist analysis. Five different discourse strategies were examined, and found to have affected the degree of code-switching used by speakers in different programmes. The next chapter will

provide a discussion of the way in which these strategies allow the indexical meaning of code-switching and (a) to be interpreted and understood.

10. Discussion

Section (10.1) outlines my categorization of each variable as either markers or indicators. Section (10.2) moves on to discuss the behaviour of marker variables in the data, drawing on insight from variationism (10.2.1) speaker centred approaches (10.2.2) and audience design theory (10.2.3)

10.1 Classifying variables

In this section I will discuss my categorization of each variable. I will first discuss the markers code-switching (10.2.1) and AM variables (a) and (â/gyda) (10.2.2). Following this I will describe the indicators (fy) (20.2.3) and (ei.m) (am) and (o) (10.2.4). The final two variables (yn) and (ei.f) were strongly affected by language internal factors, and for this reason I felt unable to categorize them.

10.1.1 Code-switching as a marker

The variable that seems to be most obviously shaped by style is code-switching. There are two main findings that supported the status of code-switching as a style marker. The first is that the quantitative analysis found that the pattern of code-switching seems to correlate strongly with the indices of formality measured in the analysis, including channel cues like laughter, overlapping speech and retracing, as well what I believe to be the institutional values of the programmes. The second was the evidence found in the discourse analysis, where the contributors seemed to exhibit intentional and strategic manipulation of code-switching as a stylistic resource. This included discourse strategies which have been previously attested in the literature, including the use of code-switching for humour (Siegel 1995), to reformulate or retrace norm breaking speech towards situationally 'appropriate' forms (Alfonzetti 1998) and to make speech more comprehensible for listeners (Giles et al 1991).

10.1.2 (a) and (â/gyda)

The mutation variables that I consider most likely to be markers are (a) and (â/gyda). The distribution of AM tokens for both variables generally follows the stylistic axis in order of formality, although there is a large gap between *Post Prynawn* and *Dewi Llwyd AFS* on one hand, and *Geth a Ger* and *Tudur Owen* on the other. The two most formal programmes have frequencies of aspirate mutation which strongly diverge from community norms. For me this suggests a diglossic situation, where mastery of this divergent style is almost a prerequisite for participation. This pattern is also reinforced by a gatekeeping effect, as the majority of participants and contributors are to some sense ‘prestigious’ Welsh speakers – i.e. those who are professionally habituated to the use of standard Welsh forms.

The two more informal programmes seem to approximate the community norms of aspirate mutation use to a much greater extent. This is most clearly the case with *Tudur Owen*, where the frequency of standard tokens of (a) on *Tudur Owen* was relatively low, and seemed to most closely parallel the language of younger and less educated Welsh speakers. Furthermore, the only standard token used by any of the core presenters on this programme was used in a very obvious style shift, which clearly departed from the usual style of the programme. Conversely, *Geth a Ger* had a higher standard use of (a) than *Tudur Owen*, which I would argue may represent a conscious attempt by these presenters to craft an authoritative style, just as their avoidance of code-switching does.

Given the very low frequency of use of (â/gyda) in both informal programmes (six tokens each) it is difficult to draw strong conclusions. I would note however, that the single standard token used by GER on *Geth a Ger* is certainly more striking than that used by a non-core presenters on *Tudur Owen*, which was spoken by a sports correspondent. Bearing in mind that there were only two standard tokens of this variable in the whole *Siarad* corpus, that it seems characteristic

of the formal programmes, and that *Geth a Ger* generally used non-standard variants of most variables (excluding the ‘stable’ SM variables and code-switching), I would make the case that (â/gyda) has marker like qualities. Unfortunately, without more data for the three core presenters of *Tudur Owen* (which would allow me to calculate individual scores just as I have done with code-switching), this is not a claim I can make confidently.

10.1.3 (fy) as an indicator

In contrast to code-switching and both AM variables, I argue that (fy) is not subject to style shifting, and may instead be considered an indicator. This is because its distribution seems to reflect the social axis attributes of the presenters – particularly in terms of age - rather than their position on the stylistic continuum. The clearest evidence for this is that *Tudur Owen* had the highest standard use of this variable – even higher than the very formal *Dewi Llwyd AFS* - while having the most nonstandard use of code-switching and AM variables (a) and (â/gyda). In other words, the hosts of *Tudur Owen* have intentionally drawn upon non-standard variants (a) and code-switching to create a style that fits their subversive ethos, but have also used categorically standard tokens of (fy). Conversely, *Geth a Ger*, who have been shown to intentionally adapt their use of code-switching (and arguably have done the same with the AM variables), generally used non-standard variants of (fy).

This leads me to believe that (fy) may not be salient to these speakers, to such an extent that social factors, such as age and level education, deterministically drive their frequency of use by individuals – not stylistic expression. In other words, because speakers are comparatively unaware of the variation within (fy), speakers on *Tudur Owen* did not think (or were unable) to reduce their standard usage, while the presenters of *Geth a Ger* were similarly unable to increase their usage. Given that the *Siarad* corpus analysis in chapter 6 showed that (fy) is strongly stratified by age group, it is likely that the younger ages of *Geth a Ger* was a key factor

here. In order to illustrate this point, figures 38 and 39 below compares presenters' use of (fy) and code-switching against their own age groups, and the difference in status of the variables:²²

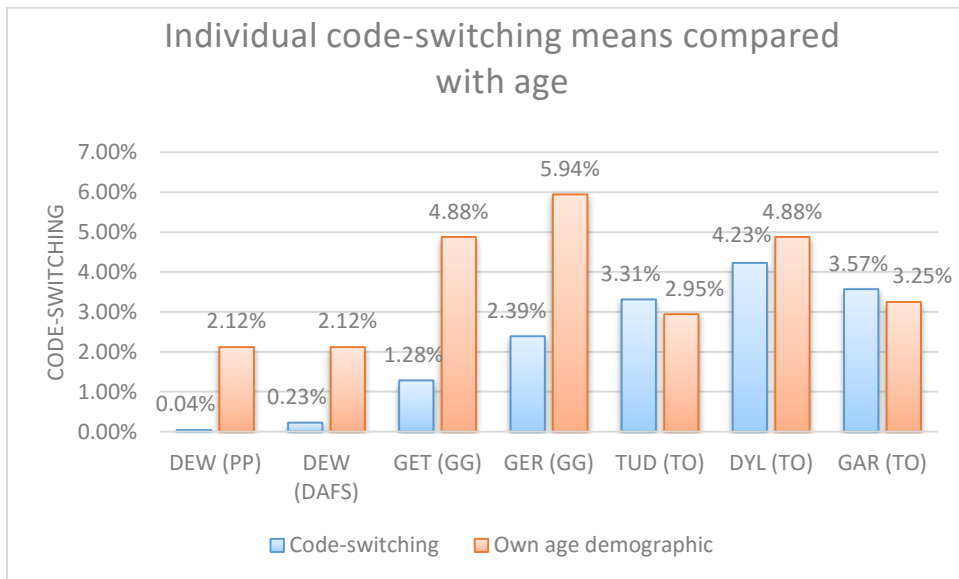


Figure 38 Each presenter's mean Code-Switching use compared their age group average

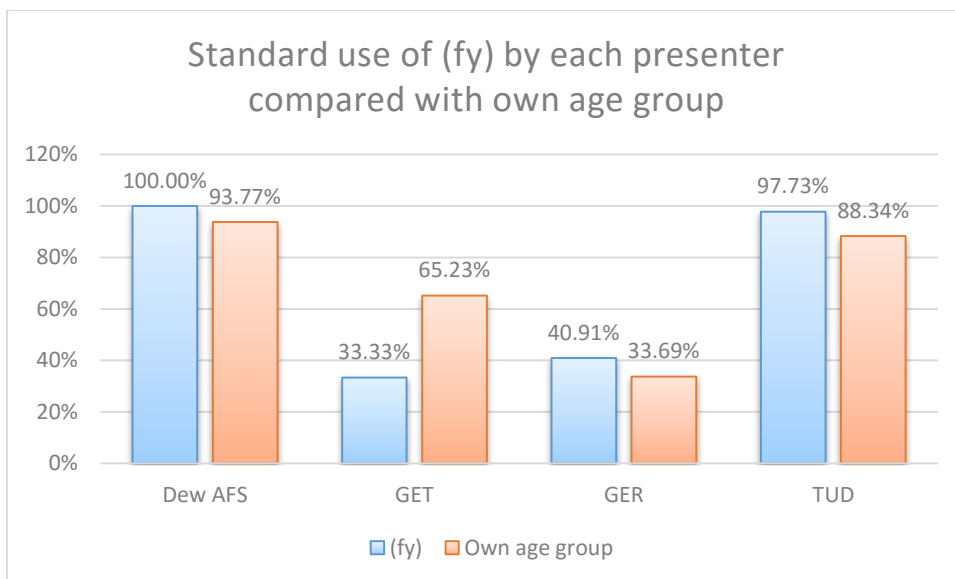


Figure 39 Each presenter's use of (fy) compared with their age group

²² Given that the Siarad corpus data was collected during the late 2000s, I decided to try and compensate for the fact that speakers' language might be expected to reflect the speech patterns of the age group they were in at that time, rather than currently. The 'own age group' scores above thus represents a mean calculated from the means of each speakers' current age group score, and the mean score of the age group below them.

GET and GER's use of code-switching clearly diverge from the vernacular use of their own age group, while their use of (fy) is very non-standard, and arguably reflects the speech of younger speakers. The core cast of *Tudur Owen* on the other hand, converge on their age group's vernacular usage of code-switching, while TUD's²³ own use of (fy) reflects his older age group's much higher norm. Thus *Geth a Ger* and *Tudur Owen*'s use of (fy) perhaps mainly reflects age differences (although there could also be other social factors at work).

There are some qualifications to this argument however. Firstly, this is based on a relatively small number of speakers, and conceivably may not reflect the behaviour of other speakers. Secondly, as I do not have a Labovian continuum of style shifting data for each presenter, there is no way of knowing how each participant's radio speech actually corresponds with their unmonitored vernacular speech. Thirdly, there are some reasons for thinking that (fy) should be a good candidate as a stylistically driven marker. Hatton's (1988) study found evidence that (fy) is a marker, and the non-standard (noun) + fi variant, which typically in this corpus accompanies non-mutation of (fy), has historically been singled out repeatedly in the literature for scorn by academics (Jones and Thomas 1977, PW. Thomas 1984), which suggests that there may be some social stigma associated with it. Despite this however, I still feel that the best explanation is that (fy) is relatively low in the level of awareness of these participants, and that they are thus primarily reflecting vernacular norms in their speech.

10.1.4 (ei.m) (am) and (o)

The variables (ei.m) and (o) both show a relatively flat stylistic distribution across the programmes, with *Post Prynawn*, *Dewi Llwyd* and *Tudur Owen* all reflecting a broadly similar level of use, although *Geth a Ger* had a slightly lower mean for (ei.m) and (o) than the other

²³ There were unfortunately not enough tokens of (fy) for DYL and GAR to calculate individual scores. This is particularly disappointing because DYL, who is younger than his co-presenters, could have provided an interesting contrast to the others. Would he have converged on his older co-presenters' use of (fy) or reflected his age group's norm?

programmes, probably due to *Geth a Ger*'s lower position on the social axis compared with the presenters of the other programmes. This suggests that both these variables can be considered to be below the level of consciousness, as I have argued (fy) may be. In the case of (am), *Post Prynawn* and *Dewi Llwyd AFS* have a slightly higher standard mean than *Geth a Ger* and *Tudur Owen*, which may indicate that the variable is slightly more sensitive to stylistic factors than (o) and (ei.m).

10.1.5 (yn) – An ambiguous case

One of the more ambiguous variables is (yn). The variable follows a distribution which is a hybrid of the others that have been covered thus far. Like (a) and (â/gyda), both *Post Prynawn* and *Dewi Llwyd AFS* can be considered a group together, showing a high frequency of standard usage at 185 and 188 respectively. This suggests that the variable is stratified stylistically. However, unlike the case of the aspirate mutation, *Tudur Owen* has a higher standard usage of (yn) than *Geth a Ger* at 124 index points, in comparison with an index score of 74 for *Geth a Ger*. I have argued above that *Geth a Ger* have intentionally tried to create a formal style, so the fact that *Geth a Ger* has the lowest standard usage may imply that social, rather than stylistic factors are driving the distribution of this variable. For the other indicator variables, *Tudur Owen*'s scores were generally closer to *Post Prynawn* and *Dewi Llwyd AFS*, if not higher in some cases. This again complicates the picture, and suggests that some other factors may be involved.

I have two potential explanations for this pattern. In contrast to the apparently strong effect of frequency of place names in the vernacular speech of the *Siarad* corpus, speakers in *Post Prynawn* and *Dewi Llwyd AFS* seemed to mutate even infrequent place names. This could be due to pressures from the formal context overriding the language internal constraints which normally prevent speakers from mutating less frequent place names. In contrast, *Geth a Ger*

and *Tudur Owen* did seem to follow the community norm of generally using the radical form with place names. Consequently, the difference between the more formal and less formal programmes may be due to the amount of attention which needs to be brought to bear to mutate less frequent place names, but it may also be due to speakers' habituation towards doing so – in other words, their social prestige. This is suggested by the frequencies at which speakers mutated place names after (o) – *Geth a Ger* again had the lowest standard mean, and the general pattern resembled that of (yn).

The second factor which may have shaped *Tudur Owen*'s use of (yn) is that the presenters show a marked preference for using soft mutation following the trigger. An intriguing possibility is that soft mutation following (yn) could have stylistic connotations of its own. Using the soft mutation following (yn) arguably opens up a speakers' stylistic options. This is because it allows speakers to avoid the potentially alienating use of the most standard variant, but also allows them to avoid non-mutation and the possible negative evaluation associated with this (e.g. as being childish or uneducated). One piece of evidence supporting this is that Dewi Llwyd's sole use of a non-standard variant in the *Radio Cymru* corpus is a token of SM following (yn), which appears in a light hearted exchange, accompanied by laughter. This suggests that very standard speakers may step down to SM variants rather than to the radical in informal situations. Ultimately, (yn) is clearly a mutation trigger that manifests in a way quite distinct to the others examined here. As discussed in chapter six, the variable's apparent time profile is unique among mutation variables. I would suggest that its singular distribution here reflects a combination of factors, including its diglossic nature (in terms of the difficulty for most speakers of mutating some place names standardly), its relatively high saliency and the strong use of the 'intermediate' SM variant in the data.

10.1.6 (ei.f) Language internal factors

Ball (1984) considered (ei.f) to be an indicator in his analysis of Cwmtawe Welsh, but the *Radio Cymru* corpus data presents a complicated picture. In three hours of *Geth and Ger* there were only two tokens of this variable. One explanation for this is that *Geth and Ger*, as young men oriented in their conversations towards masculine culture (e.g. action films, football) in the conversations analysed in the data, tended to avoid references to women. Most of the examples of the pronoun *ei* in the data, whether male or female, seem to have animate and mainly human possessors throughout the dataset, so the lack of allusion to women may have had an effect on the data in this case. In terms of other three programmes, the distribution is relatively flat, with an index score of around 150 for each one. The main factor in bringing together the scores of each programme is the high frequency of the intermediate SM mutation, which was frequently applied despite the antecedent possessors being feminine. This is particularly surprising in the case of *Post Prynawn* and *Dewi Llwyd AFS*, as it reveals a frequent lack of application of the gender appropriate mutation. While the participants in these programmes have shown that they use standard mutation variants at a rate that is highly divergent from that of vernacular Welsh speech, their standard use of gender seems less categorical. Unfortunately it appears that in this case the data does not provide us with sufficient evidence to plausibly decide how much of a role is played by style in its distribution.

10.1.7 Summary

In conclusion, taking into account the analysis of the *Siarad* corpus combined with the stylistic analysis above, here are the main findings of the *Radio Cymru* corpus study:

- Code-switching seems to be strongly stylistically driven, and very sensitive to shifts in formality. This is attested in the apparently intentional use of code-switching in

discourse strategies involving humour, flagging and reformulating speech, and style shifting between topics.

- Mutation variables present a more complex picture, with each variable behaving differently:
 - (a) and (â/gyda) were broadly stylistically stratified, but showed a sharp division between the two most formal and two least formal programmes.
 - (am) (o) and (ei.m) all essentially seemed to reflect the speakers backgrounds, i.e. younger speakers used them less standardly than older speakers. This matches the prediction made above, that stable variables would not be markers.
 - (fy) appears to be an indicator, since it seems to reflect social information about the ages of the speakers, rather than being stylistically stratified.
 - The analysis of (ei.f) was generally inconclusive, mainly because of the low frequency of tokens of the variables in *Geth a Ger*. The other three programmes had roughly equivalent use of the variable, which suggests that it is not primarily stylistically driven, and this coincides with Ball's (1984) description of the variable as an indicator. I suggest that the comparatively low standard use in the two more formal programmes may be in part due to the previously reported erosion of the gender system in the Welsh speech community (Gathercole and Thomas 2007), with even prestigious speakers not using the historically appropriate AM mutation with feminine non-human possessors.

10.2 Stylistic analysis

Having decided upon code-switching and the AM variables (a) and (â/gyda) as the best candidates for marker status, this section will now discuss how the data can be understood in the light of current sociolinguistic theories of style.

10.2.1 Applying variationist theory to the data

Where this study reflects a traditional variationist conception of style most clearly is in the finding that some Welsh language style is normatively driven, with speakers responding to ideological pressure by focusing attention on their speech to move towards approved styles. This is perhaps clearest where *Geth a Ger* retrace or reformulate their speech to remove ideologically disfavoured elements of English transfer. It can also be seen in the most formal programmes, where prestigious speakers use discourse strategies of flagging metacommentary to distance themselves from the use of English. Although the programmes here represent very specific ‘performative’ contexts, it may not be too far-fetched to suggest that similar forces could affect ordinary Welsh speakers’ conception of what are Welsh styles appropriate for formal and informal situations. For this reason, I would predict that a traditional Labovian study applied in a modern Welsh speaking community might find a similar outcome, with standard forms of code-switching, (a) and (â/gyda) increasing with greater increments of formality, while certain other variables like (fy), (ei.f) and the soft mutations could remain unaffected, or at least be affected to a lesser extent.

10.2.2 Speaker centred style in the Radio Cymru data

As Bell (1984, 1991) has argued, it is difficult to argue that radio language is not to a great extent planned and strategic. This means that, while attention to speech can explain some phenomena in the data, an interpretation which sees the presenters as deliberately crafting their programme’s style will ultimately be more convincing. But understanding how this is occurring in the data depends on interpreting the indexical meanings of the variants involved, and their connections to contrasting stances taken by presenters and contributors. This in turn demands an interpretation of the ideological framework against which stances are taken, and within which speakers orient themselves. Drawing on the interview which I discussed in the

introduction, on recent research by Robert (2009, 2011) and on the social variation that I found in the data that I analysed in chapter 6, I would describe the discourse frame which provides a backdrop to Welsh style as being primarily based on concerns about linguistic decline, which consequently foregrounds conceptions of age, education and notions of cultural purity.

Some researchers have suggested that class may be a growing factor in shaping the Welsh that speakers use (Madoc-Jones et al 2013). This may or may not be true, and I would agree with Robert (2011) that we need empirical data to establish whether or not this is the case. In the meantime however, I would argue that age is the primary variable affecting the distribution of all the variables in this study – there are significant differences between older and younger speakers for every variable, and I would further argue that there may be a connection between ideas about preserving the integrity of Welsh and the ways Welsh speakers conceptualize age, particularly as younger speakers could be considered ‘responsible’ for the changes which Welsh are undergoing. For example, where working class speakers may bear the brunt of stigmatization for breaking linguistic norms in Anglo-American contexts, in the Welsh context it is younger speakers who are the most likely to violate norms against non-standard use. This is simply because they are the most likely to use these incoming forms, for mechanical and deterministic reasons largely beyond their control.

This is the context within which the style work on *Radio Cymru* takes place. For example, the presenters and contributors on *Tudur Owen* build their style by repeatedly takes stances which are anti-authority, anti-institutional, and which overtly challenge Welsh Standard Language Ideology (Robert 2011). In doing this they build solidarity with younger and less prestigious speakers, and create an atmosphere which includes socially attractive qualities like humour and inclusiveness. Examples include Tudur Owen and a co-presenter violating central Welsh SLI norms by pretending that they prefer to speak English together, substituting English words for the Welsh ‘word of the day’ and mocking the practice of Welsh language music quotas. I argue

that these stances are connected to the indexical meanings of those linguistic features I have described as markers.

Code-switching, for example, seems to index rebelliousness, subversiveness and solidarity. Drawing on Bell's (1984) notion of the roots of style in social variation, one might argue that some of these indexical meanings could have developed partly from the actual greater use of code-switching by younger speakers. Kiesling (2009), who argues for a primary role for stances in forming indexicalities, might also argue that the semiotic connection could have been formed in normative educational contexts in Wales, where the use of English is typically punished in formal classroom situations, and where English has been described as the "language of inclusion" in the vernacular context of the schoolyard (Thomas and Roberts 2011: 92). The broad social stigma levelled against code-switching, predominantly by middle aged and older activists like *Dyfodol I'r Iaith* and *Cylch yr Iaith*, has undoubtedly also contributed to these meanings.

Stances taken by presenters and contributors can also be connected to standard variants. As I have mentioned above, Tudur Owen's single use of (a) occurs where he's taking an unusual stance (for the programme) - He is speaking authoritatively and earnestly, and the use of the AM variant serves to reinforce this. In the case of *Geth a Ger*, the programme contains a mixture of contrasting stances, and this could explain both the slightly elevated use of AM and relatively infrequent code-switching. For example, as I discussed in chapter nine, both GET and GER alternate between apologizing for using English terms and mocking Welsh terms, in essence taking alternating pro-institutional and subversive stances. GER in particular, who used the most AM tokens, frequently makes statements which are not marked epistemically in his didactic *Gwers Ger* segment, which makes them sound like "eternal truths" (Kiesling 2009: 180). Like the fraternity member Mack in Kiesling (1998), he thus may be drawing on the indexical meanings of (a) and (â/gyda) to assume institutional power, and to give his words a

more convincing and authoritative weight. The radical non-standard variant of (a) may also have its own broadly parallel indexical meanings, perhaps indexing qualities of solidarity as well as anti-institutional stances. One of the quotes in my epigraph: “*ffwcio eich gwaith a° cyfalafiaeth*” (“fuck your work and capitalism”) from a song by hip-hop group *Pep Le Pew* also aptly sums up this idea, presenting a strong anti-institutional, rebellious stance alongside the radical variant of (a). The low frequency of AM after (a) in *Tudur Owen* could thus feasibly be due to deliberate use of the non-standard variant, rather than simply low attention to speech.

10.2.3 Audience design and referee design on Radio Cymru

In terms of audience design theory, there is evidence that the contrasting ways in which presenters and contributors orient their linguistic styles towards their audiences is significant in some sense. For example, the highly standard use of language in both *Post Prynawn* and on *Dewi Llwyd AFS* can be most accurately described as *long term outgroup referee design* (Bell 1984). In this paradigm, speakers maintain the use of a linguistic variety which belongs to another group, rather than their own vernacular, and which is not commonly used within the community, but which members of the community (ostensibly) agree upon as a prestigious code. Bell associates this “institutionalised” (Bell 1984: 190) type of referee design with diglossia, and states that the outgroup code could belong to either a colonizing power, or it could be an older version of the community’s own code, as in the case of classical Arabic and Greek (Ferguson 2007). In the Welsh context I would argue that the outgroup code is standard Welsh, which represents the use of Welsh features at an earlier point in history (Fife 1986), i.e. - more frequent standard forms of (â/gyda) and (a), as well as substantially less frequent code-switching.

Tudur Owen is effectively formulated in opposition to this principle, as the whole ethos of the programme is built around a rejection of prestige norms and an embrace of solidarity. In

audience design theory terms, the use of language on *Tudur Owen* is most reminiscent of the classic audience design strategy used by New Zealand radio stations in Bell's (1977) study, where the presenters converged on the actual vernaculars of their listeners. This is most apparent in the frequent use of code-switching and general aversion to aspirate mutation in *Tudur Owen*, even while indicator variables that I have argued are below the level of consciousness (and thus not available for stylistic use), like (fy) (o) and (ei.m), are used at a very high rate of standardness. *Geth a Ger* represents an intriguing intermediate position, where the presenters seem to alternate between referee and audience design strategies. This is apparent in the way both speakers flagged English code-switches as well as low frequency Welsh words, which makes it unclear where the institutional values of the programme lie.

In summary, contributors to both formal programmes use outgroup referee design, which leads them to diverge their speech radically from that of their audience, while *Tudur Owen* largely converges upon the vernacular of the audience. Both approaches may have merit, and may reflect the wishes of their audience in different ways. For example, Bell (1984) described outgroup referee design in the context of the prestigious national broadcaster in his New Zealand study, which used RP accented English despite this being a foreign variety for most listeners. In this case listeners arguably appreciated the standard variety's use in the context of the news, because of its association with authority, intelligence and other relevant values. It could be that listeners of *Radio Cymru* see the use of standard Welsh on *Post Prynawn* and *Dewi Llwyd AFS* in a similar way, and that they evaluate it positively despite the objective linguistic distance from their own speech – although whether this is true remains unclear.

11. Conclusion

This study set out to test whether a set of linguistic features in Welsh could be considered sociolinguistic variables. I believe that I have gone some way towards answering this question. In terms of social factors, all variables in the study were stratified by age, while six were stratified by level of education, and two were correlated with gender. In terms of style, I have argued that only three of the variables were likely to be markers, with code-switching, (a) and (â/gyda) seeming to respond to normative pressures, as well as being actively drawn upon as resources for stance taking behaviour. I've also suggested some of the indexical meanings that might be associated with the marker variables, with the use of code-switching perhaps reflecting values of rebellion and solidarity, while aspirate mutation may convey a learned, earnest authority. Finally, I have also made the case that age may be a central social variable underpinning style in Welsh. This is both because it is the factor that seems to most consistently differentiate the speech styles of speakers, and because discourse on Welsh often seems to be suffused with concerns surrounding the increasing adoption of elements of English transfer into Welsh speech, which also involves a perception of progressive linguistic decline in each subsequent age group. In the next three sub-sections, I will discuss drawback and advantages of the method I used, and the potential implications of the project.

11.1 Drawbacks and advantages of the method

One of the main advantages of the radio corpus method is that it allowed the efficient collection of a relatively large corpus of natural speech. I mean natural here in the sense that the speech was not elicited through reading tasks or through any other methods. The stratification of programmes by level of formality using paralinguistic features provides an efficient way of applying stylistic analysis to natural speech, which is useful for a language whose stylistic features perhaps do not lend themselves well to traditional Labovian analysis. Another

advantage of this method was that it allowed the same dataset to be used for multiple analyses. For example, the method allowed for both code-switching and mutation to be coded for the same utterances, something which would have been more challenging to achieve using elicitation with an experimental or survey method. Furthermore, it also allowed for a quantitative approach to be complemented by a qualitative analysis, which allowed me to back up the claims about the stance taking behaviour of the presenters.

In terms of the drawbacks of the methodology, one possible criticism is that it relies on assumptions about the way in which each presenters' radio language differs from their vernacular use. As I have no data for the presenters' unmonitored vernacular speech, the evidence that they're style shifting for each programme is indirect, and is largely based around suppositions about their speech in relation to average community use, and group use of each variable. It seems very likely, for example, that DEW is adapting his speech significantly for the radio, given his almost categorical use of variables like (*â/gyda*), which are nearly absent in the vernacular conversations of the *Siarad* corpus. But in the case of *Geth a Ger* and *Tudur Owen* the possibility remains that my observations about the way in which they have style shifted could be mistaken. For example, perhaps the presenters of *Geth a Ger* really do use the same low rate of code-switching seen in the programme in their day to day vernacular speech. The advantage of Labovian approaches is that they can provide concrete data for multiple individuals across multiple styles, while the method used here is reliant on indirect assumptions about speakers based on age group averages.

Another methodological problem was the low frequency of some variables, which arguably impoverished certain aspects of the analysis. In particular, the AM trigger (*â/gyda*) was sufficiently infrequent in the informal radio programmes that drawing confident conclusions about its use in *Geth a Ger* and *Tudur Owen* is impossible. Although more episodes of each programme could have been transcribed, it is not clear how much additional material would

have been needed. This is of course a problem that studies which use e.g. elicitation methods are able to avoid. However, the low frequency of tokens from some variables can be considered a small price to pay for the ability to balance a broad range of linguistic variables and to use mixed methodologies with naturalistic data. Additionally, the size of the radio corpus analysis can be considered a substantial progression from the study upon which this is based. Compared with a total of 32 mutation tokens analysed in Ball et al (1988), my study has involved the analysis of 1,851 mutation tokens in the *Radio Cymru* corpus analysis and 4,020 tokens in the *Siarad* corpus analysis (with a combined total of 5,871), undoubtedly a significant expansion. This study's methodology has allowed for a large scale population study to complement a more local and speaker centred design, with the result that the latter was strengthened by crucial insights. These included the findings on the influence of internal factors such as *place name*, the impact of age and education on the use of mutation, and the community wide distribution of different variables.

11.2 Future Research

Because of its broad scope, this thesis has opened up multiple potential avenues for future research. Some aspects of the analysis yielded rather ambiguous results – for example, why did (yn) pattern so differently (when place names were included) to the other variables in apparent time? What role is played by factors such as token frequency and place names? Further research could clarify this, perhaps by comparing the effects of frequency and other factors using a more controlled methodology. Another little understood aspect of Welsh is the role of internal constraints on the use of code-switching. My methodology used the dictionary criterion as an efficient device for comparing the relative use of English switches in a large data set. This had the advantage of being simple and easy to operationalize, but may have missed some nuances in terms of completely circumscribing the variable context. Further research could refine the procedure by looking more closely for differences between the types of linguistic forms which

are code-switched most often. Another useful avenue of research could be to explore other variables which reflect variation between Welsh speakers. Here I would draw attention to the *Siarad* corpus once more – the research potential for further study within the dataset remains immense, and *Siarad* offers an efficient means of identifying additional sociolinguistic variables worthy of study.

Finally, I feel quite strongly that future sociolinguistic studies in the Welsh context should aim to correct for the widespread assumption among Welsh speakers that non-standard Welsh only consists of mistakes, or ‘bad Welsh’. Because of the cultural perception (which I described in the first chapter) that changes in Welsh are solely associated with decline and language death, this may be an uphill struggle. However, empirical research which describes social and stylistic variation in these kinds of features may allow speakers to make more measured and objective judgements about the phenomenon, and could even give less standard speakers more confidence to use Welsh. One challenge for future research on Welsh sociolinguistics, I believe, may be to bring to light ‘vernacular’ social contexts outside of the reach of, or in opposition to, Welsh standard language ideology, and to take them seriously in their own right. Further research could find such patterns in community of practice (Eckert 2000) type studies set among Welsh speakers who situate themselves as distant from state sanctioned versions of Welshness, perhaps including the kind of disadvantaged working class Welsh speakers described by Madoc-Jones et al (2013). Ultimately, a crucial priority for future sociolinguistic researchers in Wales should be to engage with the diverse heteroglossic (Bakhtin 1981), as opposed to monologic, potential of Welsh.

11.3 Implications of the project

As discussed in chapter two, *Radio Cymru* is currently in the midst of a change in format, which is prioritising inclusiveness and a range of Welsh ‘voices’. Perhaps Tudur Owen himself could

be considered a forerunner of this process, particularly since his popularity (which has led to television appearances and a high visibility in the Welsh media) may be due at least partly to his populist approximation of vernacular speech and norms, and defiant adoption of what Smith calls “mongrel speech” (Smith 2000). *Radio Cymru* has to balance its output between two conflicting forces. On the one hand, it is bound according to its charter to serve the public, and this has been seized upon by purist pressure groups like *Cylch Yr Iaith* who argue that the corporation has a duty to maintain standards (and in particular to block the use of English), and to provide the standard Welsh style that these listeners prefer. On the other hand, *Radio Cymru* also have a responsibility to create a commercially successful channel, and it is clear that the management do generally believe that this means opening up the range of presenters to include speakers of non-standard Welsh.

The movement towards de-standardization is part of a gradual change that has taken place over the last decade or more, perhaps for a mix of commercial and cultural reasons. In this way, Wales is no different from many other contexts which have seen this trend (Coupland and Kristiansen 2011). In Ball et al’s (1988) original study on Welsh radio, even the most informal programme *Stondin Ddyddiol* used highly standard mutation patterns – with a mean standard AM mutation use at 30% substantially higher than that of the *Siarad* corpus, and this contrasts strongly with the occurrence of the variable in modern informal radio programmes, like *Geth a Ger* and *Tudur Owen*. During the late eighties, programmes like *Geth a Ger* and *Tudur Owen*, which powerfully satirize purist norms and include presenters who are both young and inexperienced in the use of standard Welsh, would have been unimaginable within the institution as it then was. What the future holds for the station is not clear, but as time goes on, the commercially driven movement towards de-standardization seems to be winning traction over standardizing forces, and it would seem to be a sound prediction that the linguistic norms

of *Radio Cymru* will continue towards a more vernacular style which reduces the sharp diglossia seen historically on the channel.

Given the very strong sentiments from corners such as *Cylch yr Iaith* and *Dyfodol i'r Iaith* concerning the need to restrict participation on the radio to proficient users of standard Welsh, the negative implications of this include that many Welsh speakers who lack access to standard Welsh may be excluded from participation in many domains of Welsh life, with little prospect of being able to acquire it effectively. This raises a dilemma for those who wish to promote the use of the language. There are clearly many Welsh speakers who do value the standard variety of Welsh, but it is equally clear that others find it distracting and alienating - which is not surprising when the strongly divergent nature of the code in relation to vernacular Welsh is considered. Given that education and age seem to be factors that strongly affect a speakers' use of these features, this may be even more problematic, since younger and less educated speakers are likely to be those who feel this alienation more than others.

As Robert (2009) has noted, in order for Welsh to survive in the future, attention needs to be paid as much towards maintaining a lively vernacular culture which is maximally inclusive, as towards maintaining linguistic standards. Arguably both sides of the coin, the prestige and solidarity capabilities of Welsh, can be maintained successfully using positive strategies, which do not discriminate against or reject Welsh speakers who use non-standard features. For instance, attention could be focused upon empowering speakers to use standard forms, perhaps using novel techniques such as games, apps and instructional websites, rather than simply excluding a large and valuable segment of the Welsh speaking population from public life. As Gareth Potter has argued, there are many Welsh speakers from a diverse variety of backgrounds who want to contribute towards the vibrancy of modern Welsh culture, and it would be a tragic missed opportunity to shut such people out due to preciousness about elements of English transfer in their speech. *Radio Cymru*'s strategy of involving speakers who speak a broad range

of different types of Welsh thus seems to me a laudable attempt at opening up participation in Welsh-speaking society to an increasingly cosmopolitan Wales.

References

Aitchison, J. W. and Carter, H. (1987). The Welsh language in Cardiff: a quiet revolution. *Transactions of the Institute of British Geographers*, 12(4), 482-492.

Aitchison, J. W. and Carter, H. (2000). *Language, economy, and society: The changing fortunes of the Welsh language in the twentieth century*. Cardiff: University of Wales Press.

Alfonzetti, G. (1998). The conversational dimension in code-switching between Italian and dialect in Sicily. In P. Auer (Ed.), *Code-switching in conversation: Language, interaction and identity* (pp 180-210). London: Routledge.

Álvarez-Cáccamo, C. (1996). The power of reflexive language (s): Code displacement in reported speech. *Journal of pragmatics*, 25(1), 33-59.

Auer, P. (1984). *Bilingual conversation*. Amsterdam: John Benjamins Publishing.

Auer, P. (Ed.). (1998). *Code-switching in conversation: language, interaction and identity*. London: Routledge.

Awbery, G. M. (1986). Moves towards a simpler, binary mutation system. In H. Andersen (Ed.), *Sandhi phenomena in the languages of Europe (Trends in Linguistics, Studies and Monographs, 33)* (pp. 161-166) Berlin: Mouton de Gruyter.

Baker, C. (1993). Bilingual Education in Wales. In H. B. Beardsmore (Ed.), *European models of bilingual education* (pp. 7-29). Clevedon: Multilingual Matters.

Bakhtin, M. M. (1981). *The dialogic imagination: four essays* (C. Emerson and M. Holquist, Trans). Austin: University of Texas Press.

Ball, M. J. (1984). *Sociolinguistic aspects of the Welsh mutation system*. Unpublished doctoral dissertation, University of Wales.

- Ball, M. J. (1986). The reporter's test as a sociolinguistic tool. *Language in Society*, 15(03), 375-386.
- Ball, M. J. (1988). Variation in the use of initial consonant mutations. In M. J. Ball (Ed.), *The use of Welsh: A contribution to sociolinguistics* (pp. 70-81). Clevedon: Multilingual Matters.
- Ball, M. J. and Müller, N. (1992). *Mutation in Welsh*. London: Routledge.
- Ball, M. J., and Williams, B. J. (2001). *Welsh phonetics*. Lewiston: The Edwin Mellen Press.
- BBC (2011). *Carchar i aelod o Gylch yr Iaith*. Retrieved from: http://news.bbc.co.uk/welsh/hi/newsid_9490000/newsid_9497600/9497685.stm [Accessed 04 July 15]
- BBC (2013). *Newidiadau Mawr i Radio Cymru*. Retrieved from: <http://www.bbc.co.uk/cymrufyw/25038081>. [Accessed 04 July 15].
- Beebe, L. M. (1981). Social and situational factors affecting the communicative strategy of dialect code-switching. *International Journal of the Sociology of Language*, 1981(32), 139-149.
- Bell, A. (1977). *The language of radio news in Auckland: A sociolinguistic study of style, audience, and subediting variation*. Unpublished PhD thesis, University of Auckland.
- Bell, A. (1984). Language style as audience design. *Language in society*, 13(02), 145-204.
- Bell, A. (1991). *The language of news media*. Oxford: Blackwell.
- Bell, A. (2001). *Back in style: Reworking audience design*. In P. Eckert & J. R. Rickford (2001) *Style and sociolinguistic variation* (pp. 139-169) Cambridge, UK: Cambridge University Press.
- Benor, S. B. (2001). The learned /t/: phonological variation in Orthodox Jewish English. *University of Pennsylvania Working Papers in Linguistics*, 7(3), 2.

- Bevan, J. (1970). *Astudiaeth Seinyddol o Gymraeg llafar Coety Walia a Rhuthun ym Mro Morgannwg*. Unpublished M. A. thesis, University of Wales.
- Bialystok, E. (1991). *Language processing in Bilingual children*. Cambridge, University Press.
- Blom, J. P. and Gumperz, J. J. (2007). Social meaning in linguistic structure: code-switching in Norway. In Wei, L. (Ed.) *The Bilingualism Reader* (pp. 75-96). London: Routledge.
- Bloomfield, L. (1933). *Language History*. New York: Holt, Rinehart & Winston.
- Bourdieu, P. (1991). *Language and symbolic power*. Cambridge: Polity.
- Bourhis, R. Y. and Giles, H. (1977). In H. Giles (Ed.), *Language, Ethnicity and Intergroup Relations* (pp. 119-135). London: Academic Press.
- Bourhis, R. Y., Giles, H., Leyens, J. P. and Tajfel, H. (1979). Psycholinguistic distinctiveness: Language divergence in Belgium. H. Giles and R. St. Claire (Eds.), *Language and social psychology* (pp. 158-185). Oxford: Blackwell.
- Borsley, R. D., Tallerman, M. and Willis, D. (2007). *The syntax of Welsh*. Cambridge University Press.
- Brend, R. M. and Pike, K. (1976). *Tagmemics*. The Hague: Mouton.
- Bucholtz, M. (1996). Geek the girl: language, femininity and female nerds. In N. Warner, J. Ahlers, L. Bilmes, M. Oliver, S. Wertheim, M. Chen (Eds.), *Gender and Belief Systems* (pp. 119–31). Berkeley, CA: Berkeley Women Lang. Group.
- Campbell-Kibler, K. (2007). Accent, (ING), and the social logic of listener perceptions. *American speech*, 82(1), 32-64.

Morris, C. (2011). *Y drafodaeth radio bythgofiadwy gyda Peter Hughes Griffiths* . [Blog post] Retrieved from: <http://ytwill.com/2011/06/gareth-potter-peter-hughes-griffiths/>. [Accessed 05 July 15]

Chambers, J. K. (1995). *Sociolinguistic theory*. Cambridge: Blackwell.

Chambers, J. K. (2000). Universal sources of the vernacular. *Sociolinguistica*, 14, 11-15.

Chambers, J. K. (2003). *Sociolinguistic Theory: Linguistic Variation and Its Social Significance*, 2nd edition. Maiden, MA. Blackwell Publishers.

Chambers, J. K. and Trudgill, P. (1980) *Dialectology*. Cambridge, Cambridge University Press.

Cheshire, J. (1982). *Variation in an English dialect: A sociolinguistic study*. Cambridge: Cambridge University Press.

Coupland, N. (1985). 'Hark, hark, the lark': social motivations for phonological style-shifting. *Language & Communication*, 5(3), 153-171.

Coupland, N. (2001). *Language, situation, and the relational self: Theorizing dialect-style in sociolinguistics*. In P. Eckert & J. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 185-210). Cambridge: Cambridge University Press.

Coupland, N. (2007). *Style: Language variation and identity*. Cambridge: Cambridge University Press.

Coupland, N. (2012). Bilingualism on display: The framing of Welsh and English in Welsh public spaces. *Language in Society*, 41(01), 1-27.

Coupland, N. and Ball, M. J. (1989). Welsh and English in contemporary Wales: Sociolinguistic issues. *Contemporary Wales*, 3, 7-40.

- Coupland, N., Bishop, H., Williams, A., Evans, B. and Garrett, P. (2005). Affiliation, engagement, language use and vitality: Secondary school students' subjective orientations to Welsh and Welshness. *International Journal of Bilingual Education and Bilingualism*, 8(1), 1-24.
- Coupland, N. and Kristiansen, T. (2011). SLICE: Critical perspectives on language (de) standardisation. In Tore Kristiansen and Nikolas Coupland (Eds.), *Standard languages and language standards in a changing Europe* (pp. 11-35). Oslo: Novus Press.
- Clyne, M. (1967). *Transference and triggering: Observations on the language assimilation of postwar German-speaking migrants in Australia*. The Hague: Martinus Nijhoff.
- Cukor-Avila, P. and Bailey, G. (1995). An approach to sociolinguistic fieldwork. *English World-Wide*, 16, 1-36.
- Cunliffe, D. J. (2012). The use of the Welsh language on Facebook: an initial investigation. *Information, Communication and Society*, 13(2), 226-248.
- Cunliffe, D., Morris, D. and Prys, C. (2013). Young bilinguals' language behaviour in social networking sites: the use of Welsh on Facebook. *Journal of Computer-Mediated Communication*, 18(3), 339-361.
- Davies, E. J. (1955). *Astudiaeth Gymharol o Dafodieithoedd Dihewyd a Llandygwydd*. Unpublished M. A. thesis, University of Wales.
- Davies, J. (2007). *Hanes Cymru*. London: Penguin.
- Davies, J. (2014). *The Welsh language: a history*. Cardiff: University of Wales Press.
- Davies, L. (1968). *Astudiaeth seinyddol gan gynnwys geirfa o dafodiaith Merthyr Tudful a'r Cylch*. Unpublished M. A. thesis, University of Wales.

Davies, P. and Deuchar, M. (2010). Using the Matrix Language Frame model to measure the extent of word-order convergence in Welsh-English bilingual speech. In A. Breitbarth, C. Lucas, S. Watts and D. Willis (Eds.), *Continuity and Change in Grammar* (pp. 77-81). Amsterdam: John Benjamins Publishing.

Davies, P. and Deuchar, M. (2014). Auxiliary deletion in the informal speech of Welsh-English bilinguals: A change in progress. *Lingua*, 143, 224-241.

Deuchar, M. (2005). Minority Language Survival: Code-mixing in Welsh. In Cohen, K. T. McAlister, K. Rolstad, and J. MacSwan (Eds.), *ISB4: Proceedings of the 4th International Symposium on Bilingualism* (pp. 608-620). Somerville, MA: Cascadilla Press.

Deuchar, M. (2006). Welsh-English code-switching and the Matrix Language Frame model. *Lingua*, 116(11), 1986-2011.

Deuchar, M. and Davies, P. (2009). Code switching and the future of the Welsh language. *International Journal of the Sociology of Language*, 195, 15-38.

Deuchar, M. (2014). *Documentation File for the Bangor Siarad Corpus*. Retrieved from: http://bangortalk.org.uk/docs/Siarad_doc.pdf

Deuchar, M., Davies, P., Herring, J., Couto, M. P., and Carter, D. (2014). In Building bilingual corpora. In E. M. Thomas and I. Mennen (Eds.), *Advances in the Study of Bilingualism*, (pp. 93-111). Bristol: Multilingual Matters

Deuchar, M., Muysken, P. and Sung-Lan Wang (2007). Structured variation in code-switching; towards an empirically based typology of bilingual speech patterns. *International Journal of Bilingual Education and Bilingualism* 10(3), 298-340.

Dorian, N. C. (1981). *Language death: The life cycle of a Scottish Gaelic dialect*. Philadelphia: University of Pennsylvania Press.

Dorian, N. (1994). Stylistic variation in a language restricted to private-sphere use. In D. Biber and E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 217-232). New York: Oxford University Press.

Dressler, W. and Wodak-Leodolter, R. (1977). Language preservation and language death in Brittany. *Linguistics*, 15(191), 33-44.

Dyfodol i'r Iaith (2014). *May Day Challenge for Radio Cymru*. [Blog post] Retrieved from <http://www.dyfodol.net/2014/05/01/her-calan-mai-i-radio-cymru/?lang=en#more-1261>

[Accessed 04 July 15]

Eckert, P. (2000). *Language variation as social practice: The linguistic construction of identity in Belten High*. Oxford: Blackwell.

Eckert, P. (2001). Style and Social Meaning. In P. Eckert and J. R. Rickford (Eds.), *Style and Sociolinguistic Variation* (pp. 119-126). Cambridge: Cambridge University Press.

Eckert, P. (2003). *The meaning of style*. Proceedings of the Eleventh Annual Symposium about Language and Society. Texas Linguistic Forum, Austin 47: 41-53.

Eckert, P. (2005). *Variation, convention and social meaning*. Paper presented at the Annual Meeting of the Linguistic Society of America. Oakland, California.

Eckert, P. (2008). Variation and the indexical field. *Journal of sociolinguistics*, 12(4), 453-476.

Eckert, P. and Rickford, J. R. (Eds.) (2001). *Style and sociolinguistic variation*. Cambridge: Cambridge University Press.

Ferguson, C. A. (2007). Diglossia. In L. Wei (Ed.), *The Bilingualism Reader* (pp. 33-46). New York: Routledge.

Fife, J. (1986). Literary vs. colloquial Welsh: problems of definition. *Word*, 37(3), 141-151.

Finegan, E. and Biber, D. (2001). Register variation and social dialect variation: the Register Axiom. In P. Eckert and J. R. Rickford (Eds.), *Style and Sociolinguistic Variation* (pp. 235-267). Cambridge: Cambridge University Press.

Finegan, E. and Biber, D. (1994). Register and social dialect variation: An integrated approach. In E. Finegan and D. Biber (Eds.), *Sociolinguistic perspectives on register* (pp 315-347). New York: Oxford University Press.

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

Fishman, J. A. (1980). Bilingualism and biculturism as individual and as societal phenomena. *Journal of Multilingual & Multicultural Development*, 1(1), 3-15.

Fynes-Clinton, O. H. (1913). *The Welsh vocabulary of the Bangor district* (Vol. 1). Milford, Oxford University Press.

Gal, S. (1979). *Language shift: Social determinants of linguistic change in bilingual Austria*. London: Academic Press.

Gardner-Chloros, P. (2009). *Code-switching*. Cambridge: Cambridge University Press.

Gardner-Chloros, P. and Finnis, K. (2004). How code-switching mediates politeness: gender related speech among London Greek-Cypriots. *Estudios de Sociolingüística*. Special Issue: Language and Gender: An interdisciplinary perspective. 4(2), 505-533.

Gathercole, G., V. C., Mon Thomas, E. and Hughes, E. (2008). Designing a normed receptive vocabulary test for bilingual populations: A model from Welsh. *International Journal of Bilingual Education and Bilingualism*, 11(6), 678-720.

- Gathercole G. and Thomas E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition* 12 (2), 213–237.
- Giggs, J. and Pattie, C. (1991). *Croeso i Gymru. Welcome to Wales. But welcome to whose Wales?* Working paper 10. Dept. of Geography: University of Nottingham.
- Giles, H., and Powesland, P. F. (1975). *Speech style and social evaluation*. London: Academic Press.
- Giles, H., Coupland, J. and Coupland, N. (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge: Cambridge University Press.
- Giles, H., Taylor, D. M. and Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in society*, 2(02), 177-192.
- GPC Online (2014). University of Wales centre for advanced Welsh and Celtic studies. Available at: <http://www.geiriadur.ac.uk/>
- Gumperz, J. J. (1982). *Discourse strategies* (Vol. 1). Cambridge: Cambridge University Press.
- Gweini (2007). *Faith in Wales: Counting for communities*. Retrieved from: <http://www.gweini.org.uk/faith-in-wales/full-report.htm>
- Greenbaum, C. W. and Cooper, R. L. (1988). *Speech Accommodation: Model for effects on language learning and interpersonal attraction*. Jerusalem: Hebrew University.
- Hatton, L. (1988). The development of the nasal mutation in the speech of schoolchildren. In M. Ball (Ed.), *The Use of Welsh* (pp. 239-257). Clevedon: Multilingual Matters.
- Hechter, M. (1978). Group formation and the cultural division of labor. *American Journal of Sociology*, 84(2), 293-318.

Honey, J. (1997). Sociophonology. In F. Coulmas (Ed.), *The Handbook of Sociolinguistics* (pp. 92-106). Oxford: Blackwell.

Houston, A. (1985). *Continuity and Change: The Variable (ing)*. Unpublished PhD thesis, University of Pennsylvania.

Hymes, D. H. (1971). Sociolinguistics and the ethnography of speaking. In E. Ardener (Ed.), *Social anthropology and language* (pp. 47-93). London: Tavistock Publications

Hymes, D.H. (1972) On Communicative Competence. In J.B. Pride and J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Harmondsworth: Penguin,

Irvine, J. T. (2001). Style as distinctiveness: the culture and ideology of linguistic differentiation. In J. K. Chambers and P. Eckert (Eds.), *Style and Sociolinguistic Variation*. (pp. 21-43). Cambridge: Cambridge University Press.

Davies, J. (2014). *The History of the Welsh Language*. Cardiff: University of Wales Press.

Jenkins, D. (1962). Aberporth: A Study of a Coastal Village in South Cardiganshire. In E. Davies and A. D. Rees (Eds.), *Welsh Rural Communities*. (pp. 1-63). University of Wales Press, Cardiff.

Jilg, T. (2003). Cadarnle'r Gymraeg? Astudiaeth Eirfaol o'r Iaith Gymraeg ym Mlaenau Ffestiniog. Unpublished PhD thesis, Cardiff University.

Jones, A.E. (1988) Age-related variation in the Welsh dialect of Pwllheli. In M, Ball. (Ed.), *The Use of Welsh* (pp. 104-122). Clevedon: Multilingual Matters.

Jones, D. G. (1988). Literary Welsh. In M, Ball. (Ed.), *The Use of Welsh*. (pp. 125-171) Clevedon: Multilingual Matters.

- Jones, K. (1995). Code-switching, intertextuality and hegemony: exploring change in bilingual discourse. In *Proceedings of the Summer School, Code-Switching and Language Contact* (pp. 108-118). Ljouwert/Leeuwarden: Fryske Akademy.
- Jones, M. (2005). Some structural and social correlates of single word intrasentential code-switching in Jersey Norman French. *French Language Studies*, 15, 1–23.
- Jones, M. C. (1998). *Language obsolescence and revitalization: Linguistic change in two sociolinguistically contrasting Welsh communities*. Oxford: Oxford University Press.
- Jones, R. O. (1976). Cydberthynas amrwyiadau iaith â nodweddion cymdeithasol yn y Gaiman Chubut, sylwadau rhagarweiniol. *Bulletin of the Board of Celtic Studies*, 27, 51-64.
- Kerswill, P. (1996). Children, adolescents, and language change. *Language variation and change*, 8(02), 177-202.
- Kiesling, S. F. (1998). Men's identities and sociolinguistic variation: The case of fraternity men. *Journal of Sociolinguistics*, 2(1), 69-99.
- Kiesling, S. F. (2009). Style as stance. In A. Jaffe (Ed.), *Stance: sociolinguistic perspectives* (pp. 171-194). Oxford: Oxford University Press.
- King, G. (2005). *Modern Welsh: a comprehensive grammar*. London: Routledge.
- Labov, W. (1963). The social motivation of a sound change. *Word*, 19(3), 273-309.
- Labov, W. (1966). *The Social Stratification of English in New York City*. Washington: Centre for Applied Linguistics.
- Labov, W. (1971). The notion of "system" in creole languages. In D. Hymes (ed.), *Pidginization and Creolization of Languages* (pp 447-472.). Cambridge: Cambridge University Press.

Labov, W. (1972). *Sociolinguistic Patterns*. Oxford: Blackwell.

Labov, W. (2001.a). A dissection of style-shifting. J. R. Rickford and P. Eckert (Eds.), *Style and sociolinguistic variation* (pp. 85-105). Cambridge: Cambridge University Press.

Labov, W. (2001.b). *Principles of linguistic change, II: social factors*. Massachusetts: Blackwell.

Labov, W. (2007). Transmission and diffusion. *Language*, 83(2), 344-387.

Lavandera, B. R. (1978). Where does the sociolinguistic variable stop? *Language in society*, 7(02), 171-182.

Lenneberg, E. H., Chomsky, N. and Marx, O. (1967). *Biological foundations of language* (Vol. 68). New York: Wiley.

Lepage, R. B. and Tabouret-Keller, A. (1985). *Acts of Identity, A creole based study of language and ethnicity*. Cambridge: Cambridge University Press.

Lewis, W. G. and Andrews, H. (2014). Bilingualism in Higher Education. In Thomas E. M., & Mennen, M. (Eds.), *Advances in the Study of Bilingualism* (pp. 171-200). Bristol: Multilingual Matters.

Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. London: Routledge.

Lloyd, S.W. (2008). *Variables that affect English language use within Welsh conversations in North Wales*. Unpublished M.A thesis, Bangor University.

MacWhinney, B. (2000). *The CHILDES project: Tools for analysing talk*, 3rd edition. New Jersey: Lawrence Erlbaum Associates.

Madoc-Jones, I., Parry, O. and Jones, D. (2013). The 'Chip Shop Welsh': Aspects of 'Welsh Speaking' Identity in Contemporary Wales. *Studies in Ethnicity and Nationalism*, 13 (3), 394–411.

Mcallister, F., Blunt, A., Prys, C., Evans, C., Jones, E. and Evans, I. (2013). *Exploring Welsh speakers' language use in their daily lives*. Beaufort Research. Accessed from: <http://www.beaufortresearch.co.uk/BBQ01260eng.pdf>

Meyerhoff, M. (2011). *Introducing sociolinguistics*. New York: Routledge.

Milroy, J. and Milroy, L. (1985). *Authority in language: Investigating language prescription and standardisation*. London: Routledge.

Milroy, L. and Milroy, J. (1992). Social network and social class: Toward an integrated sociolinguistic model. *Language in society*, 21(01), 1-26.

Milroy, L. (1980). *Language and Social Networks*. Oxford: Blackwell.

Morin, Y. C. and Kaye, J. D. (1982). The syntactic basis for French liaison. *Journal of linguistics*, 18(02), 291-330.

Muysken, P. (2000). *Bilingual Speech. A typology of code-mixing*. Cambridge: Cambridge University Press.

Myers-Scotton, C.M. (1993). *Social motivations for codeswitching: evidence from Africa*. Oxford: Clarendon Pres.

Myers-Scotton, C. (1998). A way to dusty death: the Matrix Language turnover hypothesis. In L.A. Grenoble and L.J. Whaley (Eds.), *Endangered languages: language loss and community response* (pp. 289-316). Cambridge: Cambridge University Press.

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

Nishimura, M. (1986). Intrasentential code-switching: The case of language assignment. In J. Vaid (Ed.), *Language processing in bilinguals: Psycholinguistic and neuropsychological perspectives* (pp. 123-143). Hillsdale: Lawrence Erlbaum Associates.

Ochs, E. (1988). *Culture and Language Development: Language Acquisition and Language Socialization in a Samoan Village*. Cambridge: Cambridge University Press.

Ochs, E. (1992). Indexing gender. In A. Duranti and C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 335-358). Cambridge: Cambridge University Press.

Parafita Couto, M. D. C., Carter, D., Davies, P. and Deuchar, M. (2011). Factors determining code-switching patterns in Spanish-English and Welsh-English communities. *Revista Española de Lingüística Aplicada*.

Phillipps, D. (2008). Effaith newidiadau diweddar ar hynodrwydd ieithyddol y Gymraeg. *Gwerddon*, 1(3), 94-117.

Podesva, R. (2004, April). On constructing social meaning with stop release bursts. In *Sociolinguistics Symposium* (Vol. 15).

Pope, J., Meyerhoff, M., & Ladd, D. R. (2007). Forty years of language change on Martha's Vineyard. *Language*, 83(3), 615-627.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en Espanol: toward a typology of code-switching. *Linguistics*, 18 (7-8), 581-618.

Poplack, Shana. (1981). Syntactic structure and social function of code-switching. In R. Duran (Ed.), *Latino discourse and communicative behavior* (pp. 169-184). New Jersey: Ablex.

Poplack, S. (1987). Contrasting patterns of code-switching in two communities. In M. Heller (Ed.), *Codeswitching: Anthropological and Sociolinguistic Perspectives* (pp 215-244). New York: Mouton de Gruyter.

Poplack, S. (1993). Variation theory and language contact. In D. R. Preston (Ed.), *American Dialect Research* (pp. 251-286). Amsterdam: John Benjamins.

Poplack, S. (2012). What does the Nonce Borrowing Hypothesis hypothesize? *Bilingualism: Language and Cognition*, 15(03), 644-648.

Poplack, S. and Meechan, M. (1998). How Languages Fit Together in Codemixing. *International journal of bilingualism*, 2(2), 127-38.

Preston, D. R. (2001). Style and the psycholinguistics of sociolinguistics: the logical problem of language variation. In P. Eckert and J. R. Rickford (Eds.), *Style and Sociolinguistic Variation* (pp. 119-126). Cambridge: Cambridge University Press.

Putman, W. B. and Street, R. L. (1984). The conception and perception of noncontent speech performance: Implications for speech-accommodation theory. *International Journal of the Sociology of Language*, (46), 97-114.

Rampton, B. (1995). *Crossing: Language and identity among adolescents*. Manchester: Northampton.

Rhys, M. and Thomas, E. M. (2013). Bilingual Welsh-English children's acquisition of vocabulary and reading: implications for bilingual education. *International Journal of Bilingual Education and Bilingualism*, (16)6, 633-656

Rickford, J. R. and McNair-Knox, F. (1994). Addressee-and topic-influenced style shift: A quantitative sociolinguistic study. In D. Biber and E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 235-276). New York: Oxford University Press.

- Robert, E. (2009). Accommodating “new” speakers? An attitudinal investigation of L2 speakers of Welsh in south-east Wales. *International Journal of the Sociology of Language*, 195, 93-115.
- Robert, E. (2011). Standardness and the Welsh language. In T. Kristiansen and N. Coupland (Eds.), *Standard Languages and Language Standards in a Changing Europe* (pp. 223-239). Oslo: Novus.
- Romaine, S. (1980). Stylistic variation and evaluative reactions to speech: problems in the investigation of linguistic attitudes in Scotland. *Language and speech*, 23(3), 213-232.
- Ross, S. and Berwick, R. (1992). The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition*, 14(02), 159-176.
- Rosignoli, A. (2011). *Flagging in English Italian code-switching*. Unpublished PhD thesis, Bangor University.
- Samuel, O. (1970). *Astudiaeth o dafodiaith Cymraeg Cylch y Rhigos*. Unpublished M. A. thesis, University of Wales.
- Sankoff, D. and Leberge, S. (1978). The linguistic market and the statistical explanation of variability. In D. Sankoff (Ed.), *Linguistic Variation: Models and Methods* (pp. 85-96). New York: Academic Press.
- Sankoff, D., Poplack, S. and Vanniarajan, S. (1990). The case of the nonce loan in Tamil. *Language variation and change*, 2(01), 71-101.
- Sankoff, G. and Blondeau, H. (2007). Language change across the lifespan: /r/ in Montreal French. *Language*, 83(3), 560-588.
- Saunders, L. (1937). *Canlyn Arthur*. Denbigh: Gwasg Gee.

Sayers, D. (2014). The mediated innovation model: A framework for researching media influence in language change. *Journal of Sociolinguistics*, 18(2), 185-212.

Sebba, M. and Wooton, T. (1998). We, they and identity. In P. Auer (Ed.), *Code-switching in conversation: Language, interaction and identity* (pp. 262-289). London: Routledge.

Schilling-Estes, N. (2005). Investigating Stylistic Variation. In J. K. Chambers and N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 376-401). Malden: Blackwell.

Siegel, J. (1995). How to get a laugh in Fijian: Code-switching and humor. *Language in Society*, 24(01), 95-110.

Slobin, D. I. (1977). Language change in childhood and in history. In J. Macnamara (Ed.), *Language Learning and Thought*. (pp. 185-214). New York: Academic Press.

Smith, J., Durham, M. and Fortune, L. (2007). "Mam, my trousers is fa'in doon!": Community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change*, 19(01), 63-99.

Smith, R. (2000). Broadcasting and the Welsh Language. In Jenkins, G. (Ed.), *"Let's do our best for the ancient tongue": the Welsh language in the 20th century* (pp 311-342). Cardiff: University of Wales Press

Stammers, J.R. (2010). *The integration of English-origin verbs into Welsh: a contribution to the debate over distinguishing between code-switching and lexical borrowing*. Saarbrücken: VDM.

Stuart-Smith, J. (2011). The view from the couch: Changing perspectives on the role of the television in changing language ideologies and use. In T. Kristiansen and N. Coupland

(Eds.), *Standard Languages and Language Standards in a Changing Europe* (pp. 223-239). Oslo: Novus.

Street, R. L. (1991). Accommodation in medical consultations. In Coupland, J., Coupland, N. & Giles, H. (Eds.) *Contexts of Accommodation* (pp. 131-156). Cambridge: Cambridge University Press.

Tagliamonte, S. A. (2012). *Variationist sociolinguistics: change, observation, interpretation*. Oxford: Blackwell.

Tagliamonte, S. A. and Hudson, R. (1999). Be like et al. beyond America: The quotative system in British and Canadian youth. *Journal of Sociolinguistics* 3(2): 147-172.

Thakerar, J. N., Giles, H. and Cheshire, J. (1982). Psychological and linguistic parameters of speech accommodation theory. In C. Frasier (Ed.), *Advances in the social psychology of language* (pp. 205-255). Cambridge: Cambridge University Press.

Tallerman, M. O. (1987). *Mutation and the syntactic structure of modern colloquial Welsh*. Unpublished PhD thesis, University of Hull.

Thomas, A. R. (1982). Change and decay in language. In D. Crystal (Ed.), *Linguistic controversies: Essays in linguistic theory and practice in honour of F. R. Palmer* (pp. 209–219). London: Edward Arnold.

Thomas, A. R. (1987). A spoken standard for Welsh: Description and pedagogy. *International Journal of the sociology of Language* (66), 99-114.

Thomas, E. M., Apolloni, D. and Lewis, G. (2013). The learner's voice: exploring bilingual children's selective language use and perceptions of minority language competence. *Language and Education*, 28(4), 340-361.

Thomas, E.M. and Roberts, D. (2011) Exploring bilinguals' social use of language inside and out of the minority language classroom. *Language and Education*, 25(2), 89-108.

Thomas, E. M. and Gathercole, V. C. M. (2007). Children's productive command of grammatical gender and mutation in Welsh: An alternative to rule-based learning. *First Language*, 27(3), 251-278.

Thomas, E.M., Lewis, W.G. and Apolloni, D. (2012). "Variation in Language Choice in Extended Speech in Primary Schools in Wales: Implications for Teacher Education." *Language and Education* 26 (3): 245–261.

Thomas, P. W. (1996). *Gramadeg y Gymraeg*. Cardiff: Gwasg Prifysgol Cymru.

Thomas, P.W. (1984). Variation in South Glamorgan consonant mutation. In M. J. Ball and G. E. Jones (Eds.), *Welsh Phonology* (pp. 208-236). Cardiff: University of Wales press.

Thomas, S. E. (1988). A study of *Calediad* in the Upper Swansea Valley. In Ball, M. (Ed.) *The use of Welsh: A contribution to sociolinguistics* (pp. 85-96). Clevedon: Multilingual Matters.

Thorne, D. A. (1971). *Astudiaeth seinyddol a morffolegol o dafodiaith Llangennech gan gynnwys geirfa*. Unpublished M. A. Thesis, University of Wales.

Torras, M. C. and Gafaranga, J. (2002). Social identities and language alternation in non-formal institutional bilingual talk: Trilingual service encounters in Barcelona. *Language in Society*, 31(04), 527-548.

Trudgill, P. (1974). *The social differentiation of English in Norwich*. London: Cambridge University Press.

Trudgill, P. (1981). *Linguistic accommodation: Sociolinguistic observations on a sociopsychological theory*. Papers from the Parasession on Language and Behavior, 218-237. Chicago Linguistics Society, Chicago.

Trudgill, P. (1983). Language Contact in Greece: Reduction and Simplification in the Albanian Dialects of Attica and Biotia. In P. Trudgill (Ed.), *On Dialect: Social and Geographical Perspectives*. (pp.108-126). Oxford: Blackwell.

Trudgill, P. (1992). *Sociolinguistics: An introduction to language and society*. Penguin UK.

Trudgill, P. (2014). Diffusion, drift, and the irrelevance of media influence. *Journal of Sociolinguistics* 18 (2), 214–222.

Vagg, T., Reid, S. and Harrison, J. (2011). *BBC Nations Radio Review: Quantitative audience research assessing BBC Radio Cymru's delivery of the BBC's Public Purposes*. Retrieved from:

http://downloads.bbc.co.uk/bbctrust/assets/files/pdf/our_work/nations_radio/2011/radio_cymru_research.pdf

Welsh Government (2011). *Welsh Language Measure 2011*. Retrieved from:

<http://www.legislation.gov.uk/mwa/2011/1/contents/enacted>

Welsh Government (2013). *One language for all: Review of Welsh second language at Key Stages 3 and 4*. Retrieved from:

<http://gov.wales/topics/educationandskills/publications/wagreviews/one-language-for-all/?lang=en>

White, L., and Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second language research*, 12(3), 233-265.

Williams, C. H. (2007). Marketing Welsh in an ambivalent context. *Noves SL.: Revista de sociolingüística*, (1), 5. Retrieved from:

<http://www.gencat.cat/llengua/noves/noves/hm07tardor-hivern/docs/williams.pdf>

Williams, C. H. (2010). From act to action in Wales. In D. Morris (Eds.), *Welsh in the twenty-first century*. (pp. 36-60). Cardiff: University of Wales Press.

Williams, E. (2009). Language attitudes and identity in a North Wales town: “something different about Caernarfon”? *International Journal of the Sociology of Language* (195), 63-91.

Williams, G. (1987). Bilingualism, class dialect, and social reproduction. *International Journal of the Sociology of Language*, 1987(66), 85-98.

Wolfram, W. A. (1969). *A Sociolinguistic Description of Detroit Negro Speech*. Washington DC: Center for Applied Linguistics.

Wolfram, W., and Fasold, R. W. (1974). *The study of social dialects in American English*. New Jersey: Prentice Hall.

Wolfram, W. and Schilling-Estes, N. (2006). *American English*. Oxford: Blackwell.

Woolard, K. A. (1985). Language variation and cultural hegemony: Toward an integration of sociolinguistic and social theory. *American ethnologist*, 12(4), 738-748.

Woolard, K. A. (1988). Codeswitching and comedy in Catalonia. In M. Heller (Ed.), *Codeswitching: Anthropological and sociolinguistic perspectives* (pp. 53-76). Berlin: Mouton de Gruyter.

Y Ddraig Goch (1927). Pwnc yr iaith eto: rhai ystyriaethau ar gyfrif 1921 (1927, May). *Y Ddraig Goch*, 1(12), 3-7.

Zentella, A. C. (1990). Integrating qualitative and quantitative methods in the study of bilingual code switching. *Annals of the New York Academy of Sciences*, 583 (1), 75-92.

Appendix A

This is a screenshot of the Excel file which contains all mutation tokens for the *Siarad* corpus. Each mutation token receives two lines – the upper line describes the target word, and the type of mutation applied, while the bottom line includes the original utterance from which the token was extracted, along with a line number. I colour coded each token to reflect whether it was standard (green), radical (red), auburn (intermediate SM) or undetermined (grey).

A	B	C	D	E	F
Davies 1	SAR	NON			Davies 2 GRE
	PREP AM		PREP AM		PREP AM
1009	cariad-cariad, und following prep am (missing PRONOUN)	1462	mynd-fynd, s following prep am		76 tridiau-dridiau, s
1009	oh@s:und na oedd hwnna am cariad hi	1462	fel tymor nesa y fi meddwl am (.) fynd adre (.) xxx +/.		76 ac fydda chdi (di
2197	cyfnod-gyfnod, s following prep am	2185	owpl-gwpl, und following prep am (borrowing - couple@s:und)		1271 gweddill-weddi
2197	oh@s:und ody (.) am gyfnod hir	2185	gyda ni twel bydden ni (y)n mynd am gouple@s:und o ddrinks@s tan ddeuddeg		1271 dw i just@s:und
2699	cymaint-gymaint, s following prep am	2216	pump-bump, s following prep am		2371 mynd-fynd, s fol
2699	no@s way@s mae mynd i xxx a aras yn surgery@s:und (.) am gymaint â (hyn)ny o	2216	so@s:und o'n i mynd adre gyda (y)n chwaer (.) am (.) bump		2371 oh@s:und yeah
	PREP O	2376	rhan-ran, s following prep am		2747 blwyddyn-flwyd
56	bod-fod, und following prep o (fossil - sure o fod)	2376	so@s:und os o't ti (y)n y corws ti just@s:und yn sefyll f(an) yna (.) a am ran fwya o f(2747 wnoeth um@s:u
56	bydd e nice@s:und sure@s:und o fod gobeithio		PREP O		2847 cymaint-gymain
272	pobl-pobl, ns following prep o	26	bod-fod, und following prep o (fossil - sure o fod)		2847 pam ti (we)di [?
272	and os fi (y)n gwybod bod loads@s [?] o pobl arall i gael i edrych ar_öl y person@s:	26	oh@s:und sure@s:und o fod		3274 dechrau-ddech
419	bod-fod, und following prep o (fossil - sure o fod)	32	bod-fod, und following prep o (fossil - sure o fod)		3274 o'n i (y)n meddu
419	oh@s:und (.) xxx [//] (.) ti sure@s:und o fod gallu dychmygu fi (y)n	32	sure@s:und o fod		3393 tro-dro, und foll
452	pobl-bobl, s following prep o	550	pawb-pawb, ns following prep o		3393 tyrd awn ni am e
452	neithiwr (.) cymaint o babi (y)na i edrych ar_öl Del@s:und	550	fi [//] fi eitha lico (y)r syniad o (.) pawb yn aras yn Dol_y_Cothi@s:und [?]		3367 sgwennu creadi
746	gwaith-waith, s following prep o	1477	diwrnodau-ddiwrnodau, s following prep o		3903 mynd-fynd, s fol
746	faint o waith s(y) (gy)da ti ?	1477	moe o y'n eitha (.) hectic@s:und as ti mynd adre (.) am (y)chydig o ddiwrnodau		3903 oh@s:und yeah
1732	bod-fod, und following prep o (fossil - sure o fod)	1852	merched-merched, ns following prep o		4185 cael-gael, s follc

Data is arranged for each speaker moving downwards, while scrolling rightwards reveals more speakers. A similar file exists for the *Radio Cymru* corpus data.

Appendix B

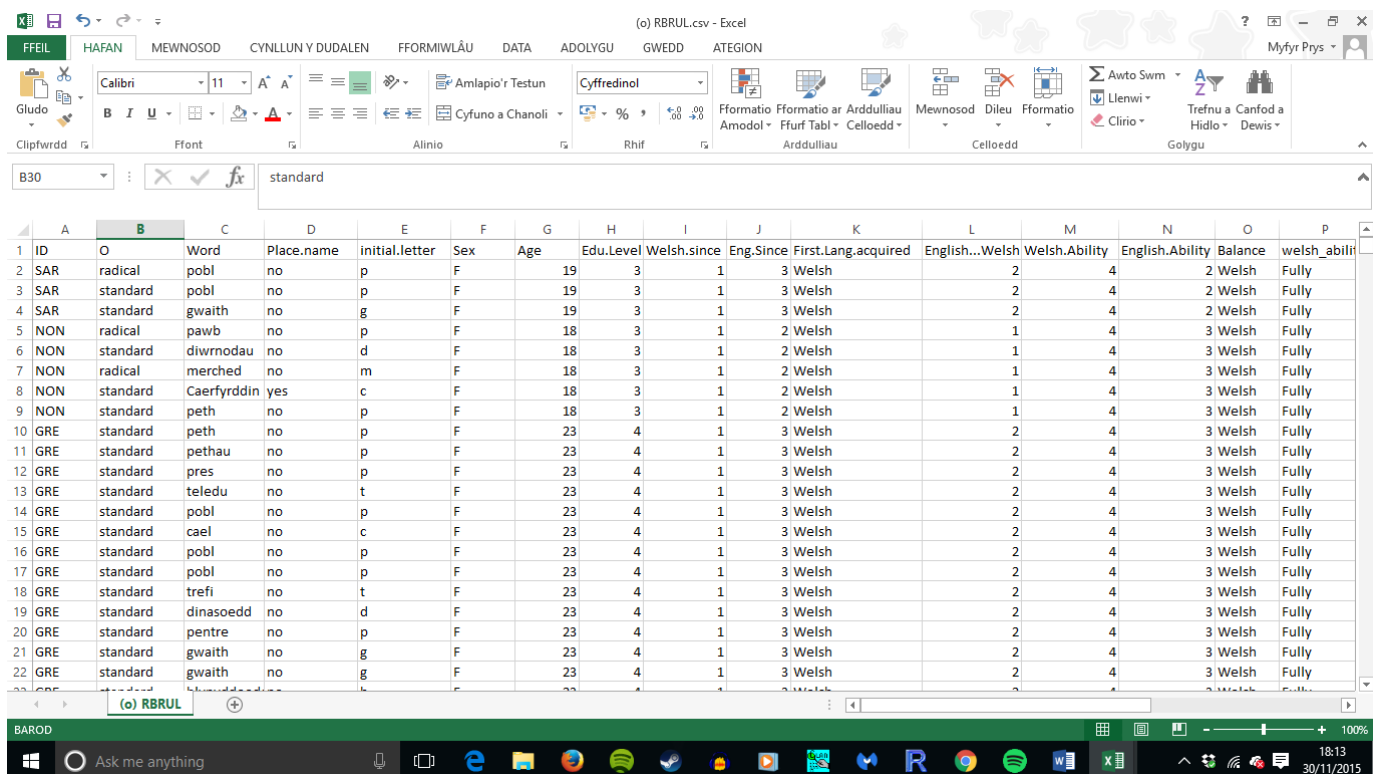
This is a screenshot of the excel file which contains individual mutation and code-switching scores for each speaker in the *Siarad* corpus. The number of standard and radical tokens were inputted based on the information in the mutations token file described above. Speakers and file names (all anonymised) are listed downwards on the left, while variables are listed rightwards at the top of the screen.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Transcript	Speaker	Age	Gender	CS.Average	am.s	am.ns	am.without.place.names.s	am.without.place.names.ns	am.place.names.only.s	am.place.names.only.ns	o.s
140	robert5	LIN	59	female	0.61%	0	0	0	0	0	0	3
141	robert6	EIR	56	female	3.99%	0	1	0	1	0	0	7
142	robert7	CLR	57	female	11.04%	2	0	2	0	0	0	5
143	robert7	HUV	66	male	2.37%	0	0	0	0	0	0	0
144	robert8	CLV	77	male	2.67%	0	0	0	0	0	0	0
145	robert8	EML	79	male	1.69%	2	0	2	0	0	0	4
146	robert8	GOR	81	male	1.68%	4	2	4	1	0	1	5
147	robert8	INT	82	male	2.42%	2	0	2	0	0	0	4
148	robert8	STN	86	male	2.75%	0	0	0	0	0	0	0
149	roberts3	MED	56	female	1.57%	5	0	4	0	1	0	8
150	stammers1	EIF	61	male	0.06%	2	0	2	0	0	0	6
151	stammers1	GTH	72	male	2.49%	2	1	2	1	0	0	20
152	stammers8	CAR	67	female	1.50%	0	0	0	0	0	0	3
153	stammers8	ISL	66	male	2.06%	2	1	2	1	0	0	11
154	stammers9	ENF	67	female	0.12%	5	0	5	0	0	0	6
155	stammers9	RNW	70	female	0.10%	1	0	1	0	0	0	4
156	SUM:					415	159	406	130	8	23	1070
157	AVG:				4.03%		72.30%		75.75%		25.81%	
158						FREQ	574					FREQ

Whole counts and percentages are both calculated at the foot of each variable's column, which can be seen at the bottom of the screen. As with the mutation tokens file on the previous page, a similar Excel file exists containing scores for all *Radio Cymru* corpus speakers.

Appendix C

Below is a screenshot of the RBRUL file containing data for the mutation variable (o). Each token has its own row, and is coded with additional information including the target word's initial consonant, whether it is a place name, as well as various further details about the speaker who used it, such as their age, sex and educational level.



ID	Word	Place.name	initial.letter	Sex	Age	Edu.Level	Welsh.since	Eng.Since	First.Lang.acquired	English...Welsh	Welsh.Ability	English.Ability	Balance	wels_h_abil
2	SAR radical	pobl	no	p	F	19	3	1	3 Welsh		2	4	2 Welsh	Fully
3	SAR standard	pobl	no	p	F	19	3	1	3 Welsh		2	4	2 Welsh	Fully
4	SAR standard	gwaith	no	g	F	19	3	1	3 Welsh		2	4	2 Welsh	Fully
5	NON radical	pawb	no	p	F	18	3	1	2 Welsh		1	4	3 Welsh	Fully
6	NON standard	diwrnodau	no	d	F	18	3	1	2 Welsh		1	4	3 Welsh	Fully
7	NON radical	merched	no	m	F	18	3	1	2 Welsh		1	4	3 Welsh	Fully
8	NON standard	Caerfyrddin	yes	c	F	18	3	1	2 Welsh		1	4	3 Welsh	Fully
9	NON standard	peth	no	p	F	18	3	1	2 Welsh		1	4	3 Welsh	Fully
10	GRE standard	peth	no	p	F	23	4	1	3 Welsh		2	4	3 Welsh	Fully
11	GRE standard	pethau	no	p	F	23	4	1	3 Welsh		2	4	3 Welsh	Fully
12	GRE standard	pres	no	p	F	23	4	1	3 Welsh		2	4	3 Welsh	Fully
13	GRE standard	teledu	no	t	F	23	4	1	3 Welsh		2	4	3 Welsh	Fully
14	GRE standard	pobl	no	p	F	23	4	1	3 Welsh		2	4	3 Welsh	Fully
15	GRE standard	cael	no	c	F	23	4	1	3 Welsh		2	4	3 Welsh	Fully
16	GRE standard	pobl	no	p	F	23	4	1	3 Welsh		2	4	3 Welsh	Fully
17	GRE standard	pobl	no	p	F	23	4	1	3 Welsh		2	4	3 Welsh	Fully
18	GRE standard	trefi	no	t	F	23	4	1	3 Welsh		2	4	3 Welsh	Fully
19	GRE standard	dinasoedd	no	d	F	23	4	1	3 Welsh		2	4	3 Welsh	Fully
20	GRE standard	pentre	no	p	F	23	4	1	3 Welsh		2	4	3 Welsh	Fully
21	GRE standard	gwaith	no	g	F	23	4	1	3 Welsh		2	4	3 Welsh	Fully
22	GRE standard	gwaith	no	g	F	23	4	1	3 Welsh		2	4	3 Welsh	Fully

As should be clear, more factors were available than were used for this thesis, which certainly leaves the door open for further analysis.

Appendix D

Finally, this is an extract of one of the CLAN transcripts that makes up the *Radio Cymru* corpus, along with the header that precedes the transcription. The following segment comes from the *Tudur Owen* radio programme, and the file is labelled *Tudur_Owen_1*:

@Begin

@Languages: cy, en

@Participants: TUD, DYL, GAR, ELE, IWA, OWA

@ID: cy, en, zh|siarad|TUD|male|||Adult||

@ID: cy, en, zh|siarad|DYL|male|||Adult||

@ID: cy, en, zh|siarad|GAR|male|||Adult||

@ID: cy, en, zh|siarad|ELE|female|||Adult||

@ID: cy, en, zh|siarad|IWA|male|||Adult||

@ID: cy, en, zh|siarad|OWA|male|||Adult||

@Media: Tudur_Owen1.WAV, audio

@situation: interview on Welsh radio program Tudur_Owen

@Date: 11-NOV-2013

@Comment: Researcher:Myfyr Prys, Bangor University

@Comment: Filename: Tudur_Owen1.CHA; Soundfile: Tudur_Owen1.WAV

@Transcriber: Myfyr Prys, Bangor University

@Time Duration: 01:16:52

*TUD: diolch Geraint@s:und .

%eng: *thanks Geraint*

*TUD: croeso cynnes iawn i (y)r sioe .

%eng: *a very warm welcome to the show*

*TUD: mae hi (y)n hanner dydd dydd Sadwrn y chweched ar hugain o Hydref .

%eng: *it's mid day on Saturday the twenty sixth of October*

*TUD: mae angen troi clociau heno cofiwch .

%eng: *the clocks need to be changed tonight remember*

*TUD: cofiwch hwnnw pwysig iawn .

%eng: *remember that - very important*

*TUD: yn y studio dyn sydd awr o flaen ei amser Dyl_Mei@s:und .

%eng: *in the studio, a man who is an hour ahead of his time, Dyl Mei*

*DYL: hello@s:und .

%eng: *hello!*

*TUD: &=laugh .

*DYL: da .

%eng: *good one*

*TUD: dyn o (y)r gorfffenol Gareth_Iwan@s:und .

%eng: *a man from the past, Gareth Iwan*

*GAR: &=laugh .

*DYL: &=laugh .

*TUD: &=laugh .

*TUD: a merch sydd mor bell ar ei ôl hi dydy hi (ddi)m hyd yn oed yma .

%eng: *and a woman who is so far behind that she's not even here*

*TUD: Manon_Rogers@s:und ddim yma heddiw .

%eng: *Manon Rogers isn't here today*

*DYL: aww@s:und .

*GAR: +< aww@s:und .

*TUD: cysylltwch â ni .

%eng: *contact us*

*TUD: chwech saith pump cant (.) ydy rhif text@s:en .

%eng: *six seven five hundred is the text number*

*TUD: neu dim tri saith dim tri pump cant pum cant ydy yr rhif ffôn .

%eng: *or, zero three seven zero three five one hundred five hundred is the phone number*

*TUD: neu fe gewch chi drydar .

%eng: *or you can tweet*

*TUD: ar y rhaglen heddiw mi fydda i yn siarad efo Eleri_Sion@s:und am yr hyn fydd ganddo hi (.) a team@s:und (.) Camp_Lawn@s:und ar ein cyfer ni ar_ôl y rhaglen yma am ddau o (y)r gloch .

%eng: *on the program today I'll be speaking to Eleri Sion about what she and the Grand Slam team will have for us after this program at two o'clock*

*TUD: mi fydd yna bennod diweddaraf o Bron_Meirion@s:und &=silly_voice .

%eng: *there will be the most recent chapter of Bron Meirion!*

*TUD: mae (y)na gyfle i chi glywed eich cyfarchion neu eich hysbysebion yn

cael eu darllen gan neb llai na cyn frenin Lloegr Risiart_Y_Trydydd

%eng: *there's a chance for you to hear your greetings or advertisements being read by none less than the former king, Richard The Third*

*TUD: felly gyrrwch eich negeseuon ac eich cyfarchion ac eich hysbysebion i_fewn am hynny .

%eng: *so send your messages and your greetings and your advertisements in for that*

*GAR: &=gasp .

*TUD: felly lle bynnag ydach chi be bynnag ydach chi (y)n wneud ar bnawn Sadwrn eitha gwyllt ar hyn o bryd .

%eng: *so wherever you are, whatever you're doing on a pretty wild Saturday evening*

*TUD: gobeithio wir fedrwch chi aros efo fi Tudur_Owen@s:und yn fyw o Fangor@s:und (.) ar BBC_Radio_Cymru@s:und (.) tan ddau o (y)r gloch

%eng: *I really hope you can stay with me Tudur Owen live in Bangor at Radio Cymru until two o'clock*

*TUD: ooh@s:und mae (y)n sŵn neis .

%eng: *ooh that's a nice sound*

*TUD: ooh@s:und mae hwn yn neis .

%eng: *ooh this is nice*

*TUD: ooh@s:und neis Swnami@s:und .

%eng: *ooh nice, Swnami*

*TUD: ooh@s:und neis te ?

%eng: *ooh nice isn't it?*

*TUD: ooh@s:und neis .

%eng: *ooh nice*

*TUD: neis de .

%eng: *nice isn't*

*TUD: (fa)sai canu rŵan dydy .

%eng: *there should be some music now*

*TUD: ooh@s:und neis .

%eng: *ooh nice*

*TUD: tyd (r)ŵan .

%eng: *c'mon now*

*TUD: (dy)ma ni .

%eng: *here we go*

%com: *song begins*

*TUD: Swnami@s:und Gwreiddiau@s:und .

%eng: *Swnami (with their song) Gwreiddiau*

*TUD: right@s:und mae (y)r wobr er@s:und am y person cynta i gysylltu â ni yn mynd i Julie_Howardson@s:und .

%eng: *right, the award for the first person to get in touch goes to Julie Howardson*

*TUD: haia@s:und Julie@s:und .

*TUD: mae hi (we)di trydar .

%eng: *she has tweeted:*

*TUD: +" ooh@s:und Tudur_Owen@s:und ar Radio_Cymru@s:und mewn chydig o funudau .

%eng: *"Ooh Tudur Owen on Radio Cymru in a few minutes!"*

*GAR: &=laugh .

*TUD: a mae hi (y)n deud mae [/] mae hi (y)n mynd i fod yn gwneud gwaith tŷ .

%eng: *and she's saying that she's going to be doing house work*

*TUD: er@s:und ac yn edrych ymlaen gwranddo .

%eng: *er, and looking forwards to listening*

*TUD: well@s:und falch dy fod ti yn gwranddo Julie@s:und .

%eng: *well I'm glad that you're listening Julie*

*TUD: a os wnewch chi gofio ar un o (y)r rhaglenni os dach chi yn gwyllo S_pedwar_C

%eng: *and if you remember on one of programs, if you watch S_Four_C*

*TUD: um@s:und (.) mi oedd Julie@s:und yn wneud reflexology@s:und arna i ryw chydig o wythnosau yn_ôl .

%eng: *er, Julie was doing reflexology on me a few weeks ago*

*TUD: paid ti â ysgwyd dy ben fel yna .

%eng: *don't you shake your head like that!*

*TUD: dw i (y)n gwybod bod chdi (ddi)m yn gwyllo S_pedwar_C Dyl_Mei@s:und !

%eng: *I know that you don't watch S_Four_C Dyl Mei!*

END OF EXCERPT