

The Acquisition of Sociolinguistic Competence in a Welsh Immersion Context

A thesis submitted to Cardiff University in partial fulfilment of the requirements for the degree of Doctor of Philosophy



Katharine Young

ENCAP/School of Welsh
Cardiff University

December 2024

Abstract

Previous research in Canada shows that French immersion students acquire a limited range of registers available to L1 speakers and do not reach full sociolinguistic competence (Mougeon *et al.* 2010). Welsh immersion differs insofar as pupils from Welsh-speaking and non-Welsh-speaking homes are taught together. However, no studies compare the acquisition of sociolinguistic competence between these groups.

The first aim of the research was to conduct a variationist analysis of the stylistic repertoires of students who speak Welsh at home. The second aim was to compare these patterns with those of students who do not speak Welsh at home to determine whether sociolinguistic competence is acquired. Data were collected from two areas—one with frequent Welsh use in the community and one without—to assess whether sociolinguistic competence is more easily acquired with greater community exposure, as shown in study abroad (Regan *et al.* 2009) and migrant contexts (Ryan 2018).

Data were elicited in three speech contexts of differing formality. To identify style-shifting patterns, three stylistically variable features of Welsh—possessive pronouns, past tense verbs, and intensifiers—were examined.

Findings reveal that by the age of 17, immersion students have acquired many of the same constraints and patterns of use, regardless of home language background. More formal contexts elicited higher rates of more formal variants for each of the three features. Furthermore, I found that those who did not speak Welsh at home were more likely to use informal variants overall, showing that they are more likely to be modelling their speech on their peers, rather than the more formal educational variety, which sets this study apart from other variationist work on international immersion. This finding shows that combining pupils of different home languages is likely to promote the acquisition of sociolinguistic competence among young immersion pupils.

Contents

1	Introduction.....	1
1.1	Overview	1
1.2	Welsh-medium education and immersion	3
1.3	Stylistic variation in Welsh.....	4
1.4	Research questions.....	5
1.5	Variables and analysis.....	6
1.6	Organization of the study.....	7
2	Sociolinguistic competence and style.....	8
2.1	Variationist sociolinguistics.....	8
2.2	Sociolinguistic competence	10
2.2.1	L1 acquiring sociolinguistic competence.....	13
2.2.2	L2 acquiring sociolinguistic competence in immersion education settings.....	16
2.3	Summary and implications	24
2.4	Style	26
2.4.1	Key concepts relating to style	27
2.5	Models of stylistic variation.....	30
2.5.1	Style as attention to speech	31
2.5.2	Style as attention to others	34
2.5.3	Third wave variationist studies of style	35
2.5.4	Summary of stylistic variation	37
2.6	Register	37
2.7	Implications.....	39
3	Stylistic variation and register in modern Welsh.....	42
3.1	Stylistic variation in Welsh.....	42
3.1.1	Introduction.....	42
3.1.2	Written vs spoken Welsh	43
3.1.3	The spoken standard	45
3.1.4	Cymraeg Byw	46
3.1.5	Towards a categorisation of variation.....	48
3.1.6	Sociolinguistic studies on Welsh style according to register.....	50
3.1.7	Styles and registers in Welsh-medium education	52
3.1.8	The stylistic variation of English home language pupils	54
3.2	Summary and implications	56
4	Methodology	58
4.1	Overview of the methodology	58
4.2	The research communities	58

4.2.1	Gwynedd.....	59
4.2.2	Cardiff.....	61
4.2.3	Summary.....	63
4.3	Data collection and analysis.....	64
4.3.1	COVID-19.....	64
4.3.2	Ethics.....	66
4.3.3	Participant recruitment.....	67
4.4	The participants.....	68
4.4.1	Home Language.....	69
4.4.2	Self-reported gender.....	71
4.5	The questionnaire.....	72
4.6	The registers (speech contexts).....	74
4.6.1	The peer-to-peer context.....	75
4.6.2	The mock interview context.....	79
4.6.3	The sociolinguistic interview.....	81
4.7	Data analysis.....	82
4.7.1	Transcription.....	82
4.7.2	The selection of variables.....	83
4.7.3	Comparative sociolinguistics.....	84
4.7.4	Statistical analysis.....	87
4.8	Summary.....	89
5	Possessive pronouns.....	90
5.1	Introduction.....	90
5.2	The feature.....	91
5.2.1	The literary variant.....	91
5.2.2	The sandwich variant.....	93
5.2.3	The colloquial variant.....	95
5.3	Previous work.....	96
5.3.1	Social factors.....	96
5.3.2	Linguistic factors.....	98
5.4	Method.....	100
5.4.1	The feature in the data.....	100
5.4.2	Data extraction.....	101
5.4.3	Social factors.....	102
5.4.4	Linguistic factors.....	103
5.5	Results (distributional analysis).....	112
5.5.1	Region.....	112

5.5.2	Home language	113
5.5.3	Speech context	115
5.5.4	Possessed noun category	117
5.5.5	Lexical frequency.....	119
5.5.6	Grammatical person.....	121
5.5.7	Possessed noun language	122
5.5.8	Mutation constraints.....	125
5.6	Results (inferential analysis).....	126
5.6.1	Individual variation.....	126
5.6.2	Summary of factors.....	127
5.6.3	Overall model.....	128
5.6.4	Summary	134
5.6.5	Gwynedd Welsh home language subset	136
5.6.6	Gwynedd English home language subset	137
5.6.7	Cardiff Welsh home language subset	139
5.6.8	Cardiff English home language subset	140
5.6.9	Hierarchies for the subset groups.....	142
5.7	Discussion	145
5.7.1	Line of evidence 1: Statistical significance	145
5.7.2	Line of evidence 2: Strength of effect.....	146
5.7.3	Line of evidence 3: Constraint hierarchy	148
5.7.4	Insights to the feature beyond sociolinguistic competence.....	150
5.8	Conclusion	151
6	Inflected and Periphrastic verbs.....	153
6.1	Introduction.....	153
6.2	The feature	154
6.2.1	The Simple Past tense	154
6.2.2	Inflected verbs.....	154
6.2.3	Periphrastic verbs.....	157
6.2.4	An historical perspective.....	158
6.3	Previous work	159
6.3.1	Social factors.....	159
6.3.2	Linguistic factors	161
6.4	Method	163
6.4.1	The feature in the data.....	163
6.4.2	Data extraction	163
6.4.3	Social factors.....	164

6.4.4	Internal factors	164
6.5	Results (distributional analysis).....	170
6.5.1	Region	171
6.5.2	Home language	171
6.5.3	Speech context	172
6.5.4	Polarity	175
6.5.5	Interrogative	177
6.5.6	Subject.....	178
6.5.7	Regularity.....	181
6.5.8	Mutation constraints.....	184
6.5.9	Verb language	184
6.6	Results (inferential analysis).....	188
6.6.1	Summary of factors.....	188
6.6.2	Overall model.....	189
6.6.3	Summary	196
6.6.4	Gwynedd Welsh home language subset	197
6.6.5	Gwynedd English home language subset	199
6.6.6	Cardiff Welsh home language subset	200
6.6.7	Cardiff English home language subset	202
6.6.8	Hierarchies for subset groups.....	204
6.7	Discussion	207
6.7.1	Line of evidence 1: Statistical significance	207
6.7.2	Line of evidence 2: Strength of effect.....	207
6.7.3	Line of evidence 3: Constraint hierarchy	208
6.7.4	Insights to the feature beyond sociolinguistic competence.....	210
6.8	Conclusion	211
7	Intensifiers	213
7.1	Introduction.....	213
7.2	The feature	214
7.2.1	Function	215
7.2.2	Borrowing	217
7.3	Previous work	217
7.3.1	Grammaticalization and overuse.....	218
7.3.2	Social factors.....	218
7.3.3	Linguistic factors	220
7.4	Method.....	221
7.4.1	Data extraction	221

7.4.2	Intensifier classification	222
7.4.3	Frequent intensifiers.....	223
7.4.4	The feature in the data.....	224
7.4.5	Social factor coding procedure	229
7.4.6	Internal factor coding procedure	229
7.5	Results (distributional analysis).....	233
7.5.1	Overall distribution of function and context.....	233
7.5.2	Moderators	235
7.5.3	Boosters.....	243
7.5.4	Maximisers.....	251
7.5.5	Summary of descriptive findings	253
7.6	Results (inferential analysis).....	257
7.6.1	Summary of factors.....	257
7.6.2	Overall model.....	258
7.6.3	Summary	262
7.6.4	Gwynedd Welsh home language subset	263
7.6.5	Gwynedd English home language subset	264
7.6.6	Cardiff Welsh home language subset	265
7.6.7	Cardiff English home language subset	267
7.6.8	Hierarchies for the subset groups.....	269
7.7	Discussion	270
7.7.1	Line of evidence 1: Statistical significance	270
7.7.2	Line of evidence 2: Strength of effect.....	271
7.7.3	Line of evidence 3: Constraint hierarchy	273
7.7.4	Insights to the feature beyond sociolinguistic competence.....	274
7.8	Conclusion	276
8	Discussion.....	278
8.1	Introduction.....	278
8.2	Patterns of stylistic variation in WHL speech (RQ1)	279
8.3	The acquisition of sociolinguistic competence for EHL pupils (RQ2).....	284
8.3.1	Use of more informal variants	287
8.3.2	Language contact and switching.....	289
8.3.3	The role of input.....	290
8.3.4	The registers.....	291
8.3.5	Awareness	293
8.3.6	Frequency.....	293
8.3.7	Transformation under transfer	294

8.3.8	Summary of acquisition of sociolinguistic competence	295
8.4	Differences according to Welsh in the community (RQ3)	295
8.5	Implications of the research findings	297
8.5.1	Implications for social use	297
8.5.2	Parent concerns	299
8.5.3	Moving away from ‘the standard’	299
8.5.4	Teaching resources.....	301
8.5.5	Welsh ‘immersion’	303
9	Conclusion	305
10	References	309
11	Appendices.....	332
11.1	Ethics approval.....	332
11.2	Email to approach schools	334
11.3	Information letter for gatekeepers.....	335
11.4	Gatekeeper consent form	341
11.5	Information letter for parents	343
11.6	Information letter for participants	348
11.7	Participant consent form (for students).....	353
11.8	Example interview questions	355
11.9	Sociolinguistic interview questions	356
11.10	Questionnaire	359
11.11	Parent jobs by area	362
11.12	Intensifiers identified in the literature.....	363

List of Tables

Table 2.1: A continuum for acquiring sociolinguistic competence in various settings	26
Table 3.1: Example of written vs spoken usage (north and south) from Ball (1988: 62).....	44
Table 3.2: <i>Cymraeg Byw</i> vs standard forms	47
Table 3.3: Categorising Welsh stylistic variation (based on Young 2019: 10)	49
Table 3.4: Registers of Welsh radio programmes, adapted from Prys (2016: 298).....	51
Table 4.1: Linguistic contrast between Gwynedd and Cardiff	64
Table 4.2: Total number of participants by home language and area	69
Table 4.3: Summary of the sample	71
Table 4.4: Careers module design.....	76
Table 5.1: Possessive pronoun variants summary	90
Table 5.2: Prefixed pronouns for each grammatical person	92
Table 5.3: Infix pronouns for each grammatical person	93
Table 5.4: Pre-nominal pronouns and their corresponding post-nominal pronouns.....	94
Table 5.5: Suffixed pronouns for each grammatical person	95
Table 5.6: Token count for all speaker groups	102
Table 5.7: Coding grammatical person	103
Table 5.8: Frequent nouns in the corpus	106
Table 5.9: Mean raw frequency	107
Table 5.10: Mutation constraints on possessive pronouns for each grammatical person	109
Table 5.11: Possession type and category.....	110
Table 5.12: Overall distribution of possessive pronoun variants.....	112
Table 5.13: Possessive pronouns variants by area and home language	113
Table 5.14: Possessive pronouns variants by speech context, by speaker group	115
Table 5.15: Possessive pronoun variants by alienability sub-category.....	117
Table 5.16: Possessive pronoun variants by alienability and speaker group	118
Table 5.17: Possessive pronouns variants by frequency, by speaker group	120
Table 5.18: Possessive variants used by Gwynedd speakers by grammatical person with highlighted 3 rd person singular and 1 st person plural	121
Table 5.19: Possessive variants used by Cardiff speakers by grammatical person with highlighted 3 rd person singular and 1 st person plural as per Davies (2016).....	122
Table 5.20: Possessive variants used by speaker group and possessed noun language.....	123
Table 5.21: Factors included in the statistical modelling	127
Table 5.22: Final generalised mixed effects model for possessive pronoun variants.....	129
Table 5.23: Analysis of deviance, Wald χ^2 test for maximal model.....	131
Table 5.24: Parsimonious model with only significant predictors of variation	132
Table 5.25: Comparison matrix for possessive by grammatical person	133
Table 5.26: Subset mixed effects model for GWHL	136
Table 5.27: Wald χ^2 test for GWHL colloquial variant use	137
Table 5.28: Subset mixed effects model for GEHL.....	137
Table 5.29: Wald χ^2 test for GEHL colloquial variant use.....	138
Table 5.30: Subset mixed effects model for CWHL.....	139
Table 5.31: Wald χ^2 test for CWHL colloquial variant use	140
Table 5.32: Parsimonious model for CWHL subset	140
Table 5.33: Subset mixed effects model for CEHL	140
Table 5.34: Wald χ^2 test for CEHL colloquial variant use	142
Table 5.35: Parsimonious model for CEHL subset	142
Table 5.36: Line of evidence 1 (possessives)	146
Table 5.37: Line of evidence 2 (possessives)	147

Table 5.38: Line of evidence 3 (possessives)	148
Table 6.1: Inflected verb constructions by subject	155
Table 6.2: Irregular verb constructions (past) in spoken Welsh by subject.....	156
Table 6.3: Periphrastic verb constructions by subject	157
Table 6.4: Soft mutations triggered by the periphrastic variant.....	162
Table 6.5: Subject coding for verb variants	168
Table 6.6: Overall rates of periphrastic and inflected verbs	170
Table 6.7: Verb variant by area.....	171
Table 6.8: Verb variant by home language and area	171
Table 6.9: Verb variant by speaker group and speech context	173
Table 6.10: Verb variant by polarity and speaker group	175
Table 6.11: Verb variant by interrogative and speaker group	177
Table 6.12: Distribution of Gwynedd inflected and periphrastic variants by subject	179
Table 6.13: Distribution of Cardiff inflected and periphrastic variants by subject	179
Table 6.14: Summary of subject patterns observed by WHL and replicated by EHL.....	180
Table 6.15: Six most frequent verbs in the dataset	181
Table 6.16: Periphrastic vs Inflected distribution of irregular verbs	182
Table 6.17: Distribution of verbs by speaker group and verb regularity	182
Table 6.18: Nonce borrowings in the data and the raw count in Simple Past verb constructions	185
Table 6.19: Distribution of verbs by speaker group and verb language	186
Table 6.20: Factors included in the statistical modelling	188
Table 6.21: Final generalised mixed effects model for verb variants.....	191
Table 6.22: Analysis of deviance, Wald χ^2 test for maximal model.....	192
Table 6.23: Parsimonious model with only significant predictors of variation.....	193
Table 6.24: Comparison matrix for verbs by subject.....	194
Table 6.25: Subset mixed effects model for GWHL	197
Table 6.26: Wald χ^2 test for GWHL periphrastic use	198
Table 6.27: Parsimonious model for GWHL subset.....	198
Table 6.28: Subset mixed effects model for GEHL.....	199
Table 6.29: Wald χ^2 test for GEHL periphrastic use	200
Table 6.30: Parsimonious model for GEHL subset	200
Table 6.31: Subset mixed effects model for CWHL.....	200
Table 6.32: Wald χ^2 test for CWHL periphrastic use	201
Table 6.33: Parsimonious model for CWHL subset	202
Table 6.34: Subset mixed effects model for CEHL	202
Table 6.35: Wald χ^2 test for CEHL periphrastic use.....	203
Table 6.36: Parsimonious model for CEHL subset	203
Table 6.37: Line of evidence 1 (verbs)	207
Table 6.38: Line of evidence 2 (verbs)	208
Table 6.39: Line of evidence 3 (verbs)	209
Table 7.1: Position of intensifiers in Welsh.....	214
Table 7.2: Modifiers adapted from Quirk et al. (1985: 589-603)	216
Table 7.3: Classification of intensifiers in the dataset	222
Table 7.4: Intensifier rate per 10,000 words	223
Table 7.5: Etymology of frequent intensifiers in the dataset	225
Table 7.6: Intensifiers by speaker group.....	229
Table 7.7: Rates of intensification with English and Welsh adjectives.....	230
Table 7.8: Most frequently intensified adjectives.....	230
Table 7.9: Rate per 10,000 words of intensifier function	233

Table 7.10: Intensifier function in the three speech contexts per 10,000 words	234
Table 7.11: Intensifier rate per 10,000 words by context	235
Table 7.12: Proportionate use of moderators in different contexts.....	236
Table 7.13: Proportionate use of moderators with Welsh or English adjectives	238
Table 7.14: Proportionate use of moderators with attributive or predicative adjectives	240
Table 7.15: Proportionate use of boosters in different contexts	243
Table 7.16: Proportionate use of boosters with Welsh or English adjectives.....	245
Table 7.17: Proportionate use of boosters with negative or positive adjectives.....	247
Table 7.18: Proportionate use of boosters with attributive or predicative adjectives	249
Table 7.19: Proportionate use of maximisers in different contexts	252
Table 7.20: Factors included in the statistical modelling	257
Table 7.21: Final generalised mixed effects model for boosting intensifier variants	259
Table 7.22: Analysis of deviance, Wald χ^2 test for maximal model.....	260
Table 7.23: Parsimonious model with only significant predictors of variation	261
Table 7.24: <i>Really</i> regression analysis for GWHL	263
Table 7.25: Wald χ^2 test for GWHL colloquial variant use	264
Table 7.26: <i>Really</i> regression analysis for GEHL	264
Table 7.27: Wald χ^2 test for GEHL colloquial variant use.....	265
Table 7.28: <i>Really</i> regression analysis for CWHL	265
Table 7.29: Wald χ^2 test for CWHL colloquial variant use	266
Table 7.30: Parsimonious model for CWHL subset	266
Table 7.31: <i>Really</i> regression analysis for CEHL	267
Table 7.32: Wald χ^2 test for CEHL colloquial variant use.....	268
Table 7.33: Parsimonious model for CEHL subset	268
Table 7.34: Line of evidence 1 (intensifiers).....	271
Table 7.35: Line of evidence 2 (intensifiers).....	272
Table 7.36: Line of evidence 3 (intensifiers).....	273
Table 8.1: Summary of lines of evidence for context across features (WHL)	279
Table 8.2: Summary of lines of evidence for context across features (WHL vs EHL)	286
Table 8.3: Tokens per feature	293
Table 8.4: An (adapted) continuum for acquiring sociolinguistic competence in various settings	304

List of Figures

Figure 4.1: Gwynedd in the north-west of Wales	59
Figure 4.2: Cardiff in the south-east of Wales	61
Figure 4.3: Screenshot of prompt slide for Discussion Session in Workshop 2: University Life (also translated into English)	77
Figure 5.1: Rates of possessive pronoun variants in Gwynedd and Cardiff	113
Figure 5.2: Rates of possessive pronoun variants by area and home language	114
Figure 5.3: Rates of possessive pronoun variants in each speech context by speaker group	117
Figure 5.4: Rates of possessive pronoun variants by alienability and speaker group	119
Figure 5.5: Rates of possessive pronoun variants by frequency and speaker group	120
Figure 5.6: Rates of possessive pronoun variants by noun language and speaker group	124
Figure 5.7: Log-odds likelihood of colloquial use	134
Figure 5.8: Log-odds likelihood of colloquial pronoun use across four speaker groups	143
Figure 6.1: Rates of verb variant by home language and region	172
Figure 6.2: Rates of verb variant by speech context and speaker group	174
Figure 6.3: Rates of verb variant by polarity and speaker group	177
Figure 6.4: Rates of verb variant by regularity and speaker group	184
Figure 6.5: Rates of verb variant by verb language and speaker group	187
Figure 6.6: Log-odds likelihood of periphrastic use	196
Figure 6.7: Log-odds likelihood of periphrastic use across 4 speaker groups	205
Figure 7.1: Rate per 10,000 words of intensifier by function	234
Figure 7.2: Proportionate use of moderators in different contexts	237
Figure 7.3: Proportionate use of moderators with Welsh or English adjectives	239
Figure 7.4: Proportionate use of moderators with attributive or predicative adjectives	241
Figure 7.5: Proportionate use of boosters in different contexts	244
Figure 7.6: Proportionate use of boosters with Welsh or English adjectives	246
Figure 7.7: Proportionate use of boosters with negative or positive adjectives	248
Figure 7.8: Proportionate use of boosters with attributive or predicative adjectives	250
Figure 7.9: Proportionate use of maximisers in different contexts	253
Figure 7.10: Log-odds results for <i>really</i> use	262
Figure 7.11: Log-odds likelihood of colloquial pronoun use across four speaker groups	269
Figure 8.1: Reported use of Welsh at school and outside school (Welsh Language Use Survey 2019-2020)	298

Acknowledgements

First of all, thanks to the ESRC and Welsh Government for co-funding this project. I make frequent reference in this thesis to the need for more research into the use of Welsh in Welsh-medium schools, and I sincerely hope that funding will continue to be available for future work in the field.

To my supervisors, Professor Mercedes Durham (ENCAP) and Dr Jonathan Morris (School of Welsh), I give huge thanks for the support, encouragement and patience you have shown me whilst I have undertaken this work. I have learnt so much over the years and consider myself very lucky to have worked with you.

Diolch, also, to Catrin Redknap formerly of KAS at Welsh Government, for showing such enthusiasm for my project, and for the secondment opportunity in 2021 which has opened many doors into research on immersion education in Wales (and beyond) for which I am extremely grateful.

To the participants at *Coleg y Fro* and *Ysgol y Ddinas* – heb eich parodrwydd i gymryd rhan yn yr ymchwil a chaniatáu imi glustfeinio ar eich sgysiau, ni fyddai'r ymchwil hwn wedi bod yn bosibl. Diolch mawr i bob un ohonoch chi am roi o'ch amser ar gyfnod mor brysur a phwysig yn eich bywydau.

Thanks are also due to my employer, Cardiff Metropolitan University, who took a chance on me in offering me a lecturing position before I completed the PhD, and for their understanding of my need to finish the thesis whilst I've been working full-time. To my office mates, colleagues and friends there, I thank you for your constant cheerleading as I've been writing up. Particular thanks to Dr Mirain Rhys, whose enthusiasm for collaboration has given me a running start in my new career. Diolch am bopeth.

To staff at Cardiff University, I also give thanks. Thanks to Rhian Rattray for always having an answer to my tedious questions and for her support throughout the MA and PhD. To Professor Dawn Knight, I thank you for the numerous work opportunities which have taught me so much and have often ended up being fun. I also thank Dr Iwan Rees for being available to provide sporadic support with elements of Welsh morphosyntax (and for including me in three Triangulation workshops in Bremen). Diolch hefyd i Dr Siôn Jones am gadw cwmni a chynnig cefnogaeth barhaol wrth imi ymlwybro yn dy gysgod.

Thanks also to my peers and postgraduate office mates. Thanks to Ellie for her excellent company on this journey since the beginning of our MA right through to the end of the PhD. Diolch i Jack am fod yn *conference companion* o fri dros y blynyddoedd. Diolch hefyd i Ianto am fod yn glust dros y ffôn ac am daflu golwg munud olaf dros darnau o'r gwaith hwn, ac i Dewi, Kaisa ac eraill am eu caredigrwydd a'u parodrwydd am sgwrs dros baned o Rooibos.

Thanks to my wonderful friends and family for their loving support, for knowing when to ask and when *definitely not* to ask how the thesis was going. Finally, to Martin, thank you. You inspire me to keep growing, one day at a time.

1 Introduction

1.1 Overview

This thesis investigates the acquisition of sociolinguistic competence among children in Welsh immersion education. In order to do this, I examine the stylistic repertoires of young Welsh speakers from both Welsh-speaking and non-Welsh-speaking homes, to assess the extent to which their production of three morphosyntactic features varies between different speech contexts. I then examine the extent to which these patterns of variation are influenced by speakers' geographical area and home language background in order to ascertain (1) whether those from non-Welsh-speaking homes are acquiring similar patterns of variation compared to their peers from Welsh-speaking homes and (2) whether there are differences in the acquisition of sociolinguistic competence among those from non-Welsh-speaking homes in two different communities which differ in the extent to which Welsh is a community language.

Sociolinguistic competence refers to the awareness of linguistic, social, and stylistic factors (Mougeon *et al.* 2010) which influence speech production in the wider community (Holmes and Brown 1976: 423). Previous work on the acquisition of sociolinguistic competence in immersion education contexts has shown that L2 speakers may be limited in terms of their acquisition of sociolinguistic competence, because they are not exposed to the same amount of authentic input as L1 speakers, which is consistently reported in the myriad international studies (e.g. Dewaele 2004; Regan *et al.* 2009; Mougeon *et al.* 2010; Ryan 2018). In these studies, L2 learners tend to over-use formal variants, which has been linked to limited variation in the input (Dewaele 2004: 313). Where the authentic input is increased, the sociolinguistic competence of speakers is thought to improve.

The Welsh context differs from those mentioned above. Firstly, little is known about stylistic variation in the speech of L1 Welsh speakers. Although there is no standard spoken

Welsh variety, there are a number of registers for which different variants could be argued to be expected (B. M. Jones 1993; Rottet and Morris 2018: 8). This assertion is problematic, however, given the development of Welsh-medium education and claims of dialect levelling, obsolescence (M. C. Jones 1998) and the adoption of *Cymraeg Byw* [Living Welsh], a standardized spoken variety (Fife 1986: 144).

Secondly, children from Welsh-speaking families and those from non-Welsh-speaking families are taught together in Welsh-medium schools, with the education system acting as immersion education (Estyn 2022)¹. This is discussed further in section 1.2 below.

Thirdly, the extent to which Welsh is used in the wider community differs markedly between communities, due to the language contact situation with English and historical, social factors such as loss of domains of use (Aitchison and Carter 1994: 24) and heavy migration following industrialisation.

This thesis therefore contributes to our understanding of sociolinguistic competence by examining stylistic variation in minority language revitalisation contexts. It contributes to our understanding of language variation in this context and, particularly, the role of home language variation (Morris 2017; Mayr *et al.*, 2017; Mennen *et al.*, 2020; Gruffydd 2022). Studies on how Welsh is acquired and used is also particularly important at a time where the language is undergoing a significant shift in terms of its revitalisation. A number of recent national policy targets have prioritised the growth of Welsh in the education sector and beyond (see Welsh Government 2017), which places a particular emphasis on creating new speakers of the language by means of language immersion in schools. The immersion model of Welsh education is discussed below.

¹ ¹ Estyn (2022) describe the Welsh-medium education model as serving as a heritage language model as well as immersion education. However, I recognise the difficulty with equating situations of ‘heritage education’ between different countries such as the United States and Wales.

1.2 Welsh-medium education and immersion

The first large-scale Welsh-medium school was introduced in Llanelli (M. C. Jones 1998: 17), initially designed for Welsh-speaking families (Owen 2018: 25) but later attracting non-Welsh-speaking families as well (Redknap 2006: 4-5), supported by the Education Act of 1944 which gave learners the right to receive education according to their parents' wishes (Duggan et al. 2014: 24). Welsh became a compulsory subject in schools with the National Curriculum in 1988, with children in Welsh-medium education sitting the same examinations regardless of language background.

In these Welsh-medium classrooms, where home language backgrounds are combined, although the education is the same for all learners, those who do not speak Welsh at home are receiving immersion education. Thus, the term 'immersion' largely refers to the type of learner, rather than the education model itself (Jones et al. 2024: 10). Given the focus in this thesis on English home language pupils acquiring sociolinguistic competence, I therefore refer to Welsh-medium education as immersion education hereon in.

The proportion of learners who are being immersed in Welsh-medium schools (i.e. who come from non-Welsh-speaking homes) differs throughout Wales. Cardiff Council claim that 70% of their current Welsh-medium pupils do not speak Welsh at home (Cardiff Council 2024). In Gwynedd, where the number of Welsh speakers is much higher, it is likely that the rate of immersion learners is lower. Revitalisation immersion education, where the majority of pupils do not speak the language at home, is similar to minoritized language contexts beyond Wales, such as the case of the immersion models for Breton (see Goalabré 2015), Scottish Gaelic (see Birnie 2022), Basque (see Gondra 2024), Catalan (see Arnau 2000) and Irish (see Murtagh 2007).

As in other revitalisation contexts, the immersion educational domain in Wales provides some pupils with their only substantial exposure to the minoritized language, whereas they are

exposed to English “in almost all domains [...] by a wide variety of individuals, and across numerous contexts of language use” (Rhys and Thomas 2013: 650).

The current government’s Cymraeg 2050 strategy aims to double the number of Welsh speakers to one million by 2050 and increase daily use of the language (Welsh Government 2017: 11). This would lead to a greater number of children and young people in Welsh-medium education from non-Welsh-speaking homes. Although numbers of speakers are rising, there are concerns that young people’s use of the language is often limited to educational domains, especially when they do not speak Welsh at home and it is not widely spoken in the community (Morris 2014). By comparing patterns of variation between speakers in a majority Welsh-speaking community and those in a community where the use of Welsh is much more restricted to social networks, this thesis sheds light on the extent to which the acquisition of sociolinguistic competence is influenced by patterns of language use in the wider community.

The new national curriculum, *Cwricwlwm i Gymru* (Welsh Government 2022) explicitly states its aim to teach the “ability to use and adapt languages in a range of roles, genres, forms, media and styles and in a suitable register”. However, little is known about the extent to which education settings are preparing Welsh-speakers to use the language outside of the school domain, and in various different settings. It is increasingly important to explore how these new learners acquire sociolinguistic competence.

1.3 Stylistic variation in Welsh

Welsh is characterized by an extremely broad range of registers (Rottet & Morris 2018: 8), associated with domains of use, types of interlocutors, and spoken versus written distinctions. Very little research has examined the extent to which speakers style-shift in different contexts, regardless of the register differences being well-attested in the literature (see B. M. Jones 1993). Work by M. C. Jones (1998) on the morphosyntactic and phonological variables in Welsh speech has found that many features of Welsh have been increasingly adapted and

simplified over the decades, which she has linked to the influence of L2 learners of Welsh on the speech community as a whole, particularly in more anglicized areas. This has also been found to be the case in minoritized and lesser used languages beyond Wales where the loss of stylistic variation is tied to language obsolescence (Kasstan 2020: 79). Thus far, it is unclear whether patterns of stylistic variation are as likely to be eroded in less anglicized areas than those considered by M. C. Jones (1998), where Welsh is heard more often in the community.

Having said that, stylistic variation in Welsh immersion schools has been found in more recent variationist work. Morris (2013) and Gruffydd (2022) found stylistic phonological variation in the language of pupils at Welsh immersion schools in contrasting areas of Wales, providing evidence that these young speakers do have access to different styles and registers in Welsh. Little, however, is known about patterns of morphosyntactic stylistic variation in Welsh immersion schools, and whether patterns of variation are different between those who speak Welsh at home and those who do not. I next summarise the aims of the research based on what has been introduced in this chapter so far.

1.4 Research questions

The project will answer the following:

- (1) What are the patterns of stylistic variation in the speech of pupils from Welsh-speaking homes?
- (2) Do pupils from non-Welsh-speaking homes acquire similar patterns of use across styles, and therefore acquire similar sociolinguistic competence?
- (3) How does the acquisition of sociolinguistic competence among L2 speakers differ depending on the extent to which Welsh is spoken in the wider community?

1.5 Variables and analysis

In order to answer these research questions, the current study will focus on three markers of stylistic variation. The first is the use of possessive pronouns, from the formal literary possessive pronoun (Davies 2016) to the more informal colloquial variant. The second morphosyntactic variable is the inflection of verbs (from the informal periphrastic variant (B. M. Jones 1993) to the more formal inflected variant). The third feature is boosting intensifiers (Tagliamonte 2016). The use of these features of Welsh (which are known to vary stylistically) will be analysed across various speech contexts, as well as the patterns and constraints that underlie their use, as an indication of the sociolinguistic competence of participating students; that is, whether they vary their speech to a variety of interactions, both formal and informal, and thus have acquired sociolinguistic competence (Bayley and Regan 2004). This will be the first variationist work of its kind to examine patterns of sociolinguistic competence in Welsh immersion pupils.

Considering the aims of this thesis, the current research will examine stylistic variation across different speech contexts (are informal variants being used in informal settings, etc.), in order to examine broad patterns of sociolinguistic competence. It is widely accepted that there are operational challenges in distinguishing casual from careful speech in interviews (Eckert and Rickford 2001), thus participants were recorded in three speech contexts of differing formality; an informal peer group context, a sociolinguistic interview, and a mock job interview, in order to determine whether their use of morphosyntactic variants varied between contexts. Two schools participated in the research. Within both schools, a mix of Welsh home language and English home language pupils were recruited to participate in order to address the second main aim of the research, i.e. do English home language pupils follow the patterns of their Welsh home language peers? The patterns of use for these three features will then be compared across groups from two different home languages in order to assess to what extent

they match. Speech data were elicited from pupils who were all between 16 and 18 years old, because, at the end of their school career, they were deemed to have acquired a full range of variation.

1.6 Organization of the study

The thesis will be laid out as follows. Firstly, I review the literature on sociolinguistic competence in international immersion contexts, and then I provide the reader with the required context into how examining style through a variationist lens can help understand whether it has been acquired. Then, I present the context of stylistic variation and register in modern Welsh, offer a categorisation of different styles in the language which then points to the evolution of potential stylistic practice in Welsh-medium education. The fourth chapter outlines the research design and comparative methodology employed to compare the patterns of variation in different groups of speakers. Chapters five through seven present the analysis of each of the three features under study: possessive pronouns, past tense verbs, and intensifiers. The discussion chapter (chapter 8) draws together the main findings, and answers the research question outlined above (section 1.4). I then provide a conclusion to the work, which points to implications and future directions for research.

2 Sociolinguistic competence and style

This chapter discusses previous research on the study of sociolinguistic competence and how those studies relate to the current work on Welsh-medium education. In order to understand how sociolinguistic competence is analysed in this study, I also present models of stylistic variation; that is, various methods of observing how speakers shift their style to match the requirements of the speaking context. First of all, however, I present variationist sociolinguistics, the framework within which sociolinguistic competence is analysed.

2.1 Variationist sociolinguistics

The variationist sociolinguistic paradigm focuses on understanding the systematic variability and change in language and, particularly, the influence of linguistic and social factors thereon. Developed initially by William Labov in the 1960s (see Labov, 1966), this paradigm has become a “cornerstone” in the field of sociolinguistics (Milroy 1995: 435). The central premise is that language varies systematically rather than randomly, and occurs at all levels of linguistic structure, including phonology, morphology, syntax, and the lexicon. Variation is often correlated with social factors such as socioeconomic class, age, gender, ethnicity, and social networks, revealing how social structures influence language production (Labov, 1972).

The variationist paradigm has evolved through different waves, with each new wave building upon the previous ones rather than displacing them. The first wave focused on broad social categories (e.g., class, age, gender) and used quantitative methods to correlate these categories with linguistic variables. Notable examples of first wave variationist studies include Labov’s New York City study, which investigated the pronunciation of the post-vocalic /r/ across different social classes, showing how this variable correlated with socioeconomic status (Labov, 1966). Trudgill’s Norwich study examined linguistic variation in Norwich, England,

correlating variables such as h-dropping and the pronunciation of /ng/ with social class and gender (Trudgill, 1974).

The second wave emphasized the role of local categories in language variation; for example, gender was analyzed not just as an isolated social category, but as an important factor within the specific local context under study. This wave also focussed on social networks and communities of practice, introducing more qualitative methods and focusing on how these social structures influence language use. For example, Milroy's Belfast study focused on the role of social networks in maintaining linguistic norms, showing how tightly knit communities preserve local dialect features (Milroy, 1987).

The third wave highlights the social meaning of variation and how individuals use language to construct and negotiate identities in specific interactions, focusing on indexicality and the stylistic use of language (Eckert 2012). Among the most notable works in third wave variationist study are Eckert's analysis of how vowel variation among Detroit teenagers indexed different social identities and stances (Eckert 1989; 2000), and Podesva's (2007) study of how a gay medic uses 'falsetto' in different contexts to construct his professional and personal identity.

The study of sociolinguistic competence is firmly rooted in the first and second waves of variationist research. That is, establishing a systematic relationship between linguistic variation and broad social categories (such as age, gender, class) through quantitative methods, and further focusing on the intricate social dynamics within smaller, more localized communities.

The current study aligns closely with second wave variationist sociolinguistics, as it recognizes that locally-relevant facts about participants play an important part in understanding patterns of variation, beyond broad social categories. Previous work on sociolinguistic competence (e.g. Rehner *et al.* 2003; Regan *et al.* 2009; Mougeon *et al.* 2010) has focused on a range of sociolinguistic variants which differ in terms of their stylistic markedness, through

a system of variant categorisation (variants which range from formal through to casual). Further information about studies of sociolinguistic competence is provided below.

2.2 Sociolinguistic competence

Patterns of sociolinguistic variation in one's first language are acquired in the home from caregivers and develop further as the child matures and their social experience broadens (Labov 2013: 248). The study of stylistic variation in L1 contexts is rooted within the field of language variation and change, and studies of sociolinguistic competence use methods from this field to explore the range of linguistic and extralinguistic variation acquired in both L1 and L2 (which includes later-learned and learner language) speakers. L1 speakers are said to know the rules of sociolinguistic variation, whereas the proficiency of later learners of a language has traditionally been measured as competence against native norms.

The study of sociolinguistic competence sets out to understand patterns of variation in L2 speech, which sets it apart from traditional Second Language Acquisition (SLA) research which has focused on aspects of the target language where native speakers display invariant linguistic usage (Rehner *et al.* 2003). Variationist studies of L2 speech have shown that variation in interlanguage, or learner speech, like variation in native speech, is highly systematic and subject to a range of linguistic and social constraints (Regan and Bayley 2004: 3). Indeed, in using the methods of L1 variationist research, researchers into the sociolinguistic competence of L2 speakers discovered that some of the same factors that influence variant choice in L1 speech also had an impact on the alternation of native versus non-native forms in learners' L2, such as attention to form, communicative task, interspeaker accommodation, and medium of communication (see Rehner *et al.* 2003 for an overview). Equally, factors that are unique to L2 learners were also found to be influential, such as input, time spent learning the target language, and transfer from the learners' L1 to their L2 (Rehner *et al.* 2003: 129).

In order to avoid confusion between sociolinguistic variation in L1 speech and L2 speech, variationist studies of sociolinguistic competence have distinguished between the former as variation along the horizontal continuum and the latter as variation along the vertical continuum. Where speakers move along the ‘vertical continuum’ this refers to an increase in proficiency (Bayley and Regan 2004), where variation is viewed as non-acquisition of target language features, which can, in part, be explained by the influence of a speaker’s L1. An example of this from Welsh could be phrasal verbs directly translated from English (e.g. *tyfu i fyny* [to grow up]), which has been found not be significantly stylistically marked, but qualitatively associated with ‘incorrect’ use (Young 2019). The horizontal continuum, on the other hand, refers to a continuum of “social dialects in the speech community” (Regan *et al.* 2009: 16), where L2 speakers of French, for example, can variably use *on* and *nous* variants, which are associated with more and less prestigious varieties (e.g. Rehner *et al.* 2003). In Welsh, a strongly stylistically marked feature is the colloquial possessive pronoun construction *car fi* [my car]².

Durham (2014) categorises these types of variation in a similar way, but by different names; she refers to learning-related variation, and target-based variation, respectively. In a similar vein, Mougeon *et al.* (2010), Dewaele (2004) and others, distinguish between Type 1 (as vertical) and Type 2 (horizontal) variation. Nance *et al.* (2016) further apply notions of accent aim, identity construction, and learning motivation to their analysis, and advocate for ‘Type 3’ variation in new speakers.

In this thesis, I focus on Welsh language variation among speakers who have acquired the language at home, and those who have acquired it at school, but who are taught in the same classes. I focus on stylistically constrained variation, rather than the non-acquisition of target language features. The sociolinguistic variables in this study will be examined as markers of

² The first analysis chapter in this thesis covers this feature.

style (that is, variables will have more casual and more formal variants), and not as an indication that a speaker has made a ‘mistake’. My focus, therefore, is on Type 2 (Mougeon *et al.* 2010), target-based (Durham 2014) or horizontal (Regan *et al.* 2009) variation. Research on the acquisition of sociolinguistic variation in both home languages (e.g. Romaine 1984) and later-learned languages (e.g. Regan *et al.* 2009; Mougeon *et al.* 2010; Durham 2014) considers sociolinguistic competence to be an essential component in communicative competence, which Hymes (1972: 281) describes as “whether (and to what degree) something is appropriate [...] in relation to the context in which it is used and evaluated”. Sociolinguistic variants can be used as markers of style and register, social status or group membership, and exist as a “crucial property of all human languages” (Mougeon *et al.* 2010: 4).

Despite descriptive accounts of registers in Welsh, little is known about the stylistic repertoires of younger speakers in Welsh-medium education and particularly the acquisition of sociolinguistic competence among those from non-Welsh-speaking backgrounds. The current study will therefore explore the stylistic repertoires of pupils from English home language (EHL)³ backgrounds in Welsh-medium education in a number of registers compared with pupils from Welsh home language (WHL) backgrounds.

The comparison between pupils of different home language backgrounds who are taught in the same classrooms sets the current work apart from previously discussed work on the acquisition of sociolinguistic competence. As discussed, pupils in Welsh immersion education are taught in the same class, regardless of home language background, and similar levels of competence are therefore expected to be acquired across speakers of both backgrounds by the end of compulsory school age, because their exposure to authentic input (Rehner and Mougeon 2003) from L1 speakers in the class is thought to level out any differences (Mayr *et al.* 2017;

³ All participants in the current study were from either Welsh home language or English home language backgrounds. This will be explained further in the methods chapter.

Nance 2020). The extent to which this expectation is realised, in relation to sociolinguistic competence and stylistic repertoire, is yet to be explored. I will present the literature on variation acquired through home language transmission, and the competence reached through formal education, as well as the factors which can influence the acquisition in both contexts in following sections.

2.2.1 L1 acquiring sociolinguistic competence

The process of acquiring sociolinguistic competence in a home language is seen as a natural process; “when children acquire their mother tongues, they evidently acquire the local variants and the norms of their usage too” (Chambers 2003: 174). Children’s language development is said to begin with patterns transmitted to them from caregivers, and they are thus able to replicate the form of their parents’ generation’s language (Labov 2007: 346). It is also well known that aspects of linguistic structure can be modified in speech addressed to children (known as child directed speech, or CDS), including simplified syntax, shorter utterances, wider pitch range and slower speaking rates (Foulkes et al 2005: 178-179)⁴. Moreover, it has been found that the production of variants is “highly variable” in CDS (Roberts 2005: 157), with caregivers moving between standard and local forms and back again according to the situational context of the interaction (Smith and Durham 2019: 11). Children are therefore subject to a wide range of variation in the speech of their caregivers, some of the patterns of which will be discussed below.

According to Kerswill (1996: 187), lexically complex rules are more difficult for children to acquire than low-level phonetic rules. Further to this, although certain linguistic variables may be acquired by children based on caregiver speech, the complex linguistic constraints which govern their use may not mirror the use of adults until later. Kovac and

⁴ It is acknowledged that CDS may be a Western phenomenon, rather than a universal, as discussed by Roberts (2005: 155).

Adamson (1981: 409) explain that the children “may have the feature by age five, even although the various constraints on the behaviour of that feature may not be in order until considerably later”.

One such constraint is gender. In a study by Foulkes *et al.* (2005) looking at glottal stop use by caregivers found that boys were far more frequently exposed to non-standard variants than girls in CDS, thus showing that caregivers provide boys and girls as young as 2 years old with opportunities to learn the social-indexical values of certain sociolinguistic variables (Foulkes *et al.* 2005: 177). This exposure to social-indexical sociolinguistic variation occurs at a very young age in home language acquisition, where CDS is conditioned by the age and gender of the child, but also by the context in which the utterance takes place (Smith and Durham 2019: 12).

Another constraint (more relevant to the current study) is style. In an exploration of the acquisition of sociolinguistic variation, Smith *et al.* (2007; 2013) found that children’s grasp of shifting styles from CDS (although different from community norms) was mediated through their parents’ behaviour; that is, children learn to associate their parents’ use of formal variants with punishment and instruction and more casual speech with fun (Labov 2001: 437). The literature discussed here seems to point to caregivers’ central role in providing input which correlates strongly with the child’s output across variables, although the rate at which these are acquired varies (Smith and Durham 2019: 197).

However, at a certain age, children’s variation shifts to being influenced more by peers than their caregivers. Kerswill and Williams’ (2000) work in Milton Keynes studied forty-eight children in different age groups (four, eight and twelve years old), and found caregiver influence to be much stronger within the group of four-year-olds than in other age groups (Kerswill and Williams 2000: 106). Their study was able to demonstrate that as the child grows

older, not only does their variation move away from that of caregivers, but it also becomes increasingly systematic.

This has been attributed to the fact that awareness of stylistic norms develops at the same time as an awareness of social norms (Bell, 1984); that children develop the capacity to address different audiences as their social experience develops (this is supported by Romaine 1984, Kerswill and Williams 2000, and Labov 2001 to name but a few). Romaine (1984: 182) writes that “the transition from home to school brings children into a wider sphere of social activity and involves learning new styles of speaking and writing to cope with new communicative tasks and functions”. Thus children do not learn patterns of variation from caregiving adults alone; they also acquire those patterns from peers and from whomever they come into contact with.

Sociolinguistic studies have shown the importance of peer group influence on individual patterns of variation. For example, Eckert (1988) looked at the social networks of *jocks* and *burnouts* in a school in Detroit and found linguistic differences between the groups which were attributed to their cultural differences. The (non)use of the variant such as (uh) backing and lowering had no correlation with parents’ socioeconomic status; rather the students’ *jock/burnout* identities and social network clusters pointed to a significant correlation, with burnouts exhibiting the highest frequency of the extreme backing and lowering of (uh). Adolescents in this study assimilated with their peers; their linguistic repertoires tied to their social practice at school, rather than their family background (see also Milroy 1987).

The acquisition of sociolinguistic competence in a home language has been proven to begin early, although there is much variation in the rate of acquisition depending on caregiver input, linguistic and social constraints. There are also different phases within which children acquire and use language; the first is influenced by caregivers and the latter in by peers. It can be surmised that caregiver speech – although influential in the early acquisition of patterns of

sociolinguistic variation – does not account for the variation seen in adolescent speech, which becomes markedly different as the child develops. The participants in the current study are in the phase of acquiring variation from peers, which is likely to draw speakers from different home languages closer together in terms of their patterns of use.

2.2.2 L2 acquiring sociolinguistic competence in immersion education settings

As previously stated in section 2.2, acquiring variation in another language (other than one's home language) is also thought to be highly systematic and subject to a range of linguistic and social constraints (Bayley and Regan 2004: 324). Now, I turn to introduce the research on how sociolinguistic competence is acquired in a number of different immersion education settings. Recall that one of the observed limitations of immersive education programs (which aim to create bilingual speakers (see Johnson and Swain 1997) is that the enrolled students acquire limited competence in more casual variants. Their stylistic variation has been found to differ from those who acquire the language at home. Studies of sociolinguistic competence in other L2 settings (beyond the immersion classroom) are presented later in this section, but first I introduce the findings on the non-acquisition of sociolinguistic competence in immersion contexts.

2.2.2.1 The non-acquisition of sociolinguistic competence in immersion contexts

Research on the acquisition of sociolinguistic competence in immersion settings has largely found that pupils attending such education do not grasp the full range of styles available to L1 speakers. In the Canadian immersive education setting, for example, all students acquire French as a second language and speak a language other than French at home. This is referred to as a 'one-way' immersion model, where a majority language group acquires the same second language; that is, all students move together in one direction towards proficiency in the second

language (Fortune and Tedick 2008: 5). The descriptor ‘one-way’ references the non-target-language home background of the students attending the immersion program.

Studies by Mougeon and Rehner (2001), Rehner and Mougeon (1999) and Mougeon *et al.* (2010) have shown the limitations in participants’ acquisition of sociolinguistic competence in such a context. Specifically, students in the immersive classroom in this study were found to overuse formal and hyperformal variants, and to dramatically under-use marked informal variants, when compared with native-speaking peers. Examples of the types of features analysed in the French immersion context were the use of *on* instead of *nous* [we], the deletion of the negative particle *ne*, and the use of inflected, periphrastic and futurate present tenses. In each of these cases, local native patterns showed a great deal of variation, whilst immersion learners varied less and used higher rates of the more formal variants. Lexical items such as comparing the use of *seulement* vs *juste* [only] also showed that immersion students were less likely to opt for the more informal variants overall.

Closer to home, work on Welsh by Hatton (1988: 251) compared pupils’ ability to shift between a formal and an informal setting by measuring their use of nasal mutation after the possessive pronoun *fy* [my] and the locative preposition *yn* [in]. Hatton discovered that the group lacking Welsh exposure at home exhibited minimal stylistic variation across conditions for *fy*. This was attributed to the exclusive influence of formal school patterns on their speech, rendering them “monostylistic” (Hatton 1988: 251). Unlike children with further access and exposure to Welsh outside of education settings, those who do not speak Welsh at home lacked a diverse range of registers (Hatton, 1988: 251). According to Hatton, pupils’ ability to style-shift is largely contingent on their access to registers outside of the classroom environment.

However, not all research has found immersion students to over-use formal variants. In the case of work on the 2nd person address pronoun *tu* and *vous* [you] in French, Swain and Lapkin (1990) found that immersion students used the informal pronoun of address *tu* much

more frequently than native speakers, where the formal pronoun *vous* was expected. Although the stylistic practice of immersion students in this study is in contrast with other immersion literature showing the use of hyperformal variants more often (e.g. Mougeon and Rehner 2001), Swain and Lapkin (1990) attributed their results to the same root cause. That is, the immersion classroom is too narrow a language learning context to allow for the development of native-like norms of variation.

A number of factors have been shown to effectively promote the acquisition of sociolinguistic competence beyond this limited learning context, and I present them below. The six factors thought to improve the acquisition of sociolinguistic competence are: extra-curricular activities, explicit instruction, increased variation in teacher talk, further education, study abroad and the influence of the peer group. At the end of this section (section 2.2.2.8) I also present some considerations about how different measures of the (non)acquisition of sociolinguistic competence can impact the extent to which it is observed in L2 speech.

2.2.2.2 *Extra-curricular activities*

As stated in the above research, immersion students' language reflects the formal nature of the classroom and shows less usage of more casual variants which are known to exist in wider community use. Or, where casual variants are used, their patterning is not the same as their native counterparts'. The sociolinguistic competence of L2 students in immersion is therefore considered to be limited. Mougeon *et al.* (2010: 164) found, on the other hand, that immersion students who engaged in extra-curricular activities in the target language made less use of formal variants, and more use of informal variants than their peers who did not engage with French as much outside of school. This shows that exposure to the language outside of school can positively affect the acquisition of sociolinguistic competence in non-native communities and that students can – and do – encounter different kinds of variation outside of the classroom (see also Sundquist's (2009) work on extramural language learning).

2.2.2.3 *Explicit instruction of stylistic variation*

Some studies have also found that explicit instruction of variation can mitigate the limited grasp of sociolinguistic competence in the immersion classroom. For example, further research on French Canadian immersion L2 speakers (French & Beaulieu 2016) evaluated the effects of explicit instruction on advanced L2 learners' production of two stylistic variables (/l/ deletion vs retention and *ne* deletion vs retention). Explicit L2 instruction of sociolinguistic variation was found to impact the stylistic variation of students, triggering changes in the in the students' use of the target feature, particularly in contexts where they were able to plan their production. In the other type of task (unplanned, spontaneous speech), students tended to overuse formal variants. It was found that pre-task planning reduced constraints on students' attentional resources, thus facilitating their language processing; on the other hand, removing the planning time on oral production tasks increased the cognitive load for speakers, which adversely affected their processing skills, leading to their use of more formal variants (French and Beaulieu 2016: 65). Type of task as well as pre-task planning can therefore influence individual style-shifting in some cases of language immersion. The aim of that study was to explicitly teach students the difference between types of style, in order to improve their understanding and production of variation in contexts outside of the classroom.

Preliminary work on the Welsh context (Young, 2019) has shown that primary and secondary school students on the whole are reported as being less formal than their teachers in their use of a number of features. However, according to teacher reports, marked or stigmatised casual features are increasingly corrected by teachers when the register of a context becomes more formal, demonstrating that the teachers have an expectation for the sociolinguistic competence of students to develop in the classroom. Other than teachers' reports of correcting students' markedly casual style in formal registers (Young 2019), little is known about the

explicit teaching of morphosyntactic variation in Welsh classrooms, and how much variation students are exposed to in the classroom.

Although teaching in SLA settings is different from immersion settings, considering that it does not aim to produce additive bilinguals (Johnson and Swain 1997), there is a danger in SLA, too, that L2 learners can become mono-stylistic communicators, because of limited exposure to authentic language in the classroom (Dewaele 2001). Some work has explored the effects of explicit instruction in SLA on L2 students' development of sociolinguistic competence. Work by Geeslin and Long (2014), Beaulieu and French (2016) and earlier work by Lyster (1993; 1994) demonstrated the positive effect of explicitly teaching stylistic variation on the development of sociolinguistic competence in learners of French as an L2. Further afield, studies such as Yu's (2008) study on English learners in Taiwan have found that implicit instruction methods do not improve stylistic variation whereas the explicit enrichment intervention for teaching Spanish in the USA (Van Compernelle *et al.*, 2016) saw a marked change in students' ability to justify their use of *tú* vs *usted* 2nd person pronoun system.

2.2.2.4 *Increased variation in input from teachers*

The sociolinguistic variation in the teacher input has also been found to positively affect how immersion learners acquire sociolinguistic competence. The stylistic practice of teachers in French immersion classrooms in Canada has been considered to be limited and not reflective of broader community norms (Mougeon and Rehner 2001). A further study by Mougeon and Rehner (2019) shows that the teachers use more standard variants in the immersion classroom, at rates above even those found in the speech of the highest social strata within the wider community, which could explain the limited variation encountered in French immersion learners. In contrast, in the study of a Mandarin-English immersion education setting in the United States, Starr (2017) set out to examine sociolinguistic variation and acquisition. Although from different areas (Taiwan and Northeastern mainland China), the teachers'

Mandarin speech followed common patterns of variation by event type (teaching vs non-teaching, instructional vs non-instructional and behaviour management). The stylistic variation of teachers was found to reflect the social meaning of the variants, and they were seen to construct appropriate personae around the classroom situations that arose (Starr 2017: 47).

2.2.2.5 *Further education*

Beyond the age of compulsory immersion education, a study by Rehner (2011) sought to address whether enrolling at a bilingual university would improve the sociolinguistic competence of post-immersion learners. It was hypothesised that the higher level of education in the language and increased opportunities to connect with native speakers of French inside and outside of the classroom could expose students to greater degrees of variation. In contrast with the previously-cited studies on their younger immersion peers, university learners who had previously attended French immersion programs showed higher use of the more informal variants of the lexical variables *habiter/vivre/rester/demeurer* [to dwell] and *travail/emploi/job/poste/ouvrage* [work]. This finding suggests that university immersion programs offer students a better understanding of ‘natural’ language than their high school immersion counterparts, and those following a traditional core L2 program. The limitations of the secondary school immersion program (where no students spoke the target language at home) appears to be mitigated through exposure to a higher level of education in an immersion setting which combines students from home language backgrounds.

2.2.2.6 *Study abroad*

Study abroad refers to the opportunity afforded to university students to stay in “the community of the L2 they wish to acquire” (Regan *et al.* 2009: 2). The study abroad context places the learner (although on a more temporary basis) in a context where the target language is the majority language, in actual linguistic immersion. Previous research has found that opportunities to engage in informal interactions with native speakers can contribute to an

improvement in sociolinguistic competence as it widens learners' access to input. Regan *et al.* (2009) studied various morphosyntactic features (such as *ne* deletion and *nous/on* [we] use) of Irish-English speakers of French L2 during their year abroad at university. Learners of French were found to be sensitive to native speaker patterns of variation and contact with native speakers (which provides varied input of quality) was found to positively affect the process of acquiring native patterns which the Irish French speakers did not have prior to their year abroad.

A study by Terry (2017) examined how the experience of participating in target language interactions and engaging in the negotiation of meaning with native speakers of French during study abroad contributes to French L2 learners' acquisition of sociolinguistic competence. The author found that the length of time spent studying abroad correlated positively with the acquisition of target-like patterns of variation, although the findings were not statistically significant (Terry 2017: 569). Indeed, the only significant extralinguistic factor affecting acquisition of target-like variation was a social network analysis, which looked at the effect of density and multiplexity of relationships with native speakers during the year abroad on /l/ elision by learners (Terry 2017: 570).

2.2.2.7 *Peer group influence*

Some research has found peer groups to significantly influence the acquisition of sociolinguistic competence in L2 speakers. Recent research (Howley, 2015; Ryan, 2018) considered the acquisition of local variants in migrant adolescents⁵. Ryan's (2018) study examined the language of Polish migrants at school in three speech contexts of varying formality, compared with the linguistic variation used by local Glaswegian peers. The Polish migrants in this study were shown to replicate some of the native variability, as well as innovating in the case of certain features, sometimes in the form of hypercorrection. Howley's (2015) study, on the other hand, investigated adolescent Roma migrants' acquisition of local

⁵ Considered here as 'immersion' learners as they are being immersed in the majority school language.

phonetic variants in a Manchester school and found that very few of the Roma migrant participants replicated their L1 peers' variation across three phonological variants (2015: 203). Both studies found friendship networks to have a statistically significant effect on patterns – more so than the age at which the language was acquired, and the length of residency in the area. This shows that although both studies were conducted in schools, social networks and the communities of participants and their relationships with other peers whose home language was English was seen to play a major role in their variation, rather than teacher input or classroom materials. In other words, in the cases of these migrant speakers, peer norms were acquired more than wider school norms. Work by Schleef *et al.* (2011) also found friendship networks to be a significant social constraint on the style of locally-born vs migrant teens' realisation of (ing) in London and Edinburgh, showing that combining home language backgrounds (which includes L1 speakers) can have a positive effect on the development of sociolinguistic competence.

From a Welsh perspective, researchers have compared speakers from different home languages and found few differences in phonological and phonetic variation (Morris 2017; Mayr *et al.*, 2017; Mennen *et al.*, 2020; Gruffydd 2022). On the other hand, Morris (2013; 2021) did find significant differences between home-language groups in the production of /r/ in both English and Welsh; these differences were attributed to the tendency for linguistic background to be an important marker of peer-group membership in the community, that is, peer group dynamics can override differences in variation between students from different home language backgrounds.

2.2.2.8 *Measuring the (non)acquisition*

It is worth noting that the (non)acquisition of sociolinguistic competence in immersion settings could be related to the type of variant being measured. French and Beaulieu (2016) found that it was easier to acquire L1 norms of stylistic variation in morphosyntactic variables (*ne*

deletion/retention), rather than phonological variables (/l/ retention/deletion) in immersion settings. According to the authors (French and Beaulieu 2016: 29), morphosyntactic variables are more categorical in nature, and their (non)use happens in the same way, regardless of phonological context. This is at odds however, with work by previous authors (Mougeon and Rehner 2001; Rehner and Mougeon 1999; Mougeon *et al.* 2010, etc.), who found that even lexical items, which are supposedly easier yet to learn, were not being acquired in immersion settings. Work by Gruffydd (2022) has found that similar patterns of sociolinguistic variation are found in phonological variants in Welsh immersion schools, but it is yet to be seen where morphosyntactic variables also pattern similarly.

Furthermore, L2 speakers have been found to replicate constraint patterns and also to innovate new patterns (Meyerhoff and Schleef 2012: 406; Schleef *et al.* 2011). In these studies, the Polish participants find the non-linguistic constraints more difficult to acquire than linguistic ones, which leads to cases of innovation in learner speech. It is unknown so far whether L1 and L2 pupils in Welsh immersion schools acquire linguistic constraints more easily than non-linguistic constraints (such as style).

2.3 Summary and implications

The common thread which runs between all contexts of acquiring sociolinguistic competence is the level of input to which language users are exposed. Whether in a home language (through parental transmission, and later peer influence) or an immersive setting (at school or in a study abroad context) the greatest amount of variation, or that which conforms most to first language norms, is acquired where there is more input. In contexts where language is classroom-based, research has shown that not enough is done in order to address the explicit teaching of sociolinguistic variation. In contexts where the language is not acquired within family and community settings, frequent contact with native speakers has been shown to positively affect learners' sociolinguistic variation.

A question remains regarding the difference between sociolinguistic competence and stylistic choice. Some speakers choose to avoid using stylistically marked features in their language – not for lack of sociolinguistic competence or awareness – but because they want to maintain their learner identity (French and Beaulieu 2016). The extent to which this might be the case in young speakers of Welsh is unclear; considering the favourable attitude held towards traditional and standard models of Welsh (Thomas 1996) and a rejection of ‘new speaker’ varieties (Robert, 2009), the notion that, among certain circles, “ideal Welsh is pure Welsh” (Robert 2011: 140) may prove different to the French context of learner identity. This question, however, is beyond the scope of the current study.

The different contexts considered in the literature review above are summarised in Table 2.1. This shows the research presented above along a continuum from implicit acquisition (where caregiver norms are intrinsically transmitted) to explicit learning (where specific interventions are required to introduce sociolinguistic variation). I would argue that the Welsh immersion setting, in terms of acquiring sociolinguistic competence, resides in the middle of the continuum; speakers from Welsh and non-Welsh language backgrounds occupy the same classes, and the language is present (to varying degrees) outside of the classroom. The study abroad context, though effective in terms of exposing learners to L1 norms, is a finite experience. A gap exists in the research, however, as to the way in which sociolinguistic competence in these contexts (where Welsh is prominent in the community and where it is not) is acquired, which emphasises the need for the current research.

Context of acquiring sociolinguistic competence	Home language	Migrant language	Combined home language immersion contexts	Study abroad	Immersion education	SLA
Existing key research	Smith <i>et al.</i> 2013; Smith and Durham 2019	Schleef <i>et al.</i> 2011; Howley 2015; Ryan 2018	Rehner 2011; Morris 2017, 2021; Gruffydd 2022	Regan <i>et al.</i> 2009; Terry 2017	Mougeon and Rehner 2001; Rehner and Mougeon 1999; Mougeon <i>et al.</i> 2010	Lyster 1993, 1994; Geeslin and Long 2014; Beaulieu and French 2016)
	<p>more sociolinguistic competence acquired</p> <p style="text-align: right;">less sociolinguistic competence acquired</p>					

2.4 Style

26

to use the standard (ai) realisation in reading list tasks, and the more informal realisations /ɛ/ and /a/ in natural conversation.

Before turning to how stylistic variation has been analysed in the past, three key concepts (standard language, language change and grammaticalization) in the sociolinguistic study of style relative to the current thesis are introduced briefly below.

2.4.1 Key concepts relating to style

2.4.1.1 Standard language

Standard variants are often associated with formal speech, and non-standard variants with more casual speech. These findings also relate to social class, where formal standard speech is associated with higher social classes, and casual non-standard speech with lower social classes. Standard varieties are associated with historically entrenched elites and often associated with prestigious written registers, and they are legitimized by standard language ideologies (SLIs) which promote negative attitudes towards deviations from the idealised standard norm “which names as its model the written language, but which is drawn primarily from the spoken language of the upper middle class” (Lippi-Green 1997: 64).

Although formal registers might call for language closer to the standard, formality and standardness are not necessarily related. According to Trudgill (2002: 84) standardness is neither a language, an accent, a style nor a register. Standard language is a variety in itself; one which some describe as a minority variety, but one nonetheless against which other varieties of a language are measured. Speakers are never fully consistent in how they use features of language, and their speech will often contain a mixture of what is referred to as ‘standard’ and ‘non-standard’ forms of the same variant.

Labov (1966) linked global prestige to formal speech and global stigma to casual speech, establishing stylistic variation as a nexus between individual, cognitive, and social aspects. The

stylistic activity of a speaker was therefore deemed to be “directly connected to the speaker’s place in, and strategies with respect to, the socio-economic hierarchy” (Eckert and Rickford: 2001: 2). For many stylistic markers, however, speakers shift regardless of class, which shows that there are known stylistic differences between formal and informal variants.

These stylistic differences can also relate to notions of overt and covert prestige. Overt prestige is associated with a variant that speakers are aware of and which is broadly related to speakers of higher status. The variant [ɪŋ], for example, is associated with prestige (being of the standard variety), whereas [ɪn] is stigmatised. The notion of overt and covert prestige is discussed in Trudgill’s early work (1974) on Norwich English. Participants in this study placed greater value on supralocal prestigious pronunciation of ‘tune’ as [tju:n] and recognised the [tʰu:n] variants as distinctively local and non-standard. Having said that, Trudgill found participants to use the local variant most often in their interviews, showing a clear discrepancy between what the speakers say they do and what they actually do, pointing to a covert prestige attached to the local variant. This shows us that speakers do not necessarily orient themselves to standard notions of prestige; indeed, there are those who navigate between institutionally based prestige and more locally based prestige. As will be discussed in further detail in chapter 3, Welsh lacks a spoken standard, which is a key difference from the English example provided here. However, the standard language ideology of Welsh speakers discussed in 2.3 suggests that many Welsh speakers may have a clear notion about what they should be doing in relation to standard language use, even if they do not necessarily do it themselves.

2.4.1.2 Changes in style

The styles used by speakers are thought to change over time. Chambers (2003: 171) presents a model in three stages, indicating a movement from vernacular to standard, that accompanies the transition from adolescence to young adulthood:

First in childhood, the vernacular develops under the influence of family and friends...Second, in adolescence, vernacular norms tend to accelerate beyond the norms established by the previous generation, under the influence of dense networking...Third, in young adulthood, standardization tends to increase, especially for the subset of speakers involved in language-sensitive occupations in the broadest sense of the term.

This change has been observed in longitudinal studies such as that by Rickford and Price (2013). These authors recorded interviews with two female speakers of African American Vernacular English (AAVE) in the mid 1908s and then again in the mid 2000s. In their first interviews, they were “avid exploiters of the vernacular” (Rickford & Price, 2013: 161), and their non-standard features outnumbered their male counterparts. Whereas in the second round of interviews, after having started jobs, families and changing social networks, this use dropped significantly, and both speakers displayed far greater use of more standard features. Where there is stability at the community level, but change in the individual’s stylistic repertoire, the change is described as age-grading (Sankoff and Blondeau 2007: 562). In relation to these outlined stages, the participants of the current study will be expected to replicate styles of the peers, because of the outlined “dense networking” (Chambers 2003: 171).

In the case of a change in progress, individuals may “retain their childhood patterns, with each age cohort of speakers registering an increasing [or decreasing] use of the variant upon entering the community” (Sankoff and Blondeau 2007: 562). Some research into Welsh has found evidence of language change in progress. In her analysis of casual Welsh speech in two communities of Wales, M. C. Jones (1998) took significant differences between age groups in each locality as evidence of language change. It is important to note, too, that changes over time in Welsh may, in some part, be influenced by the education variety of Welsh used in Welsh-medium schools. Indeed, Sayers (2009: 293) notes that if education is the principal reason for increasing Welsh use, then it stands to reason that the kind of Welsh being used is likely to be influenced by that education. The extent to which this is the case in relation to the

stylistic variation encountered at school by the participants of this study will be discussed further in chapter 3 of the thesis.

2.4.1.3 *Grammaticalization*

At times, language change does not relate to style, but is instead led by a process of grammaticalization. In historical linguistics, grammaticalization refers to the process by which words or constructions evolve from more concrete or lexical forms into grammatical elements with specific grammatical functions (Nevalainen and Palander-Collin 2011). This transformation often involves a shift towards increased grammatical complexity, bleached lexical meaning, and heightened structural dependence. Sociolinguistics explores how such changes occur in language over time, and how social factors influence the grammaticalization process. Grammatical changes have been observed and documented at length in majority languages, such as in varieties of English (e.g. Tagliamonte et al., 2014) and French (e.g. Tristram, 2021). On grammaticalization in Welsh, there is a fair bit of research (e.g. D. Willis 2007, 2019). Webb-Davies and Shank (2020) present an analysis of the grammaticalization of the future temporal reference *mynd i* [going to]. Although the current work only considers the speech of young people, captured at a single point in time, it is recognised (and acknowledged further in following chapters) that the stylistic patterns of these young people exist within an ever-shifting landscape involving change and grammaticalization over time.

2.5 Models of stylistic variation

In sociolinguistic research, the concept of style is associated with Labov's variationist program (e.g. Labov 1972), where style referred to the tendency for speakers to adjust their speech in a linear fashion between formal and informal depending on the requirements of certain social settings. Since the conception of these early ideas, the style-shifting across individuals in a speech community has been attributed to a) the attention people pay to their speech (e.g. Labov

1972); b) the idea that speakers speak with an audience in mind (e.g. Bell 1984) and c) they design their speech to suit that audience (e.g. Coupland 1988). These ideas will be discussed in further detail below.

Despite its significance, there are operational challenges in distinguishing casual from careful speech in interviews, and researchers have often increased their focus on linguistic and social constraints in more recent variation studies (Eckert and Rickford 2001: 3). Indeed, although style is seen as an important independent variable in variationist sociolinguistic research, it has not been the focal point of such research (Chambers 2003: 6). Indeed, Coupland (2007: 216) notes that it has been considered a relatively minor dimension of language variation. The current section presents the historical lenses through which stylistic variation has been explored in sociolinguistic work. From large scale quantitative survey methods to ethnographic or speaker-centred approaches, the current section will provide a summary of the relevance of each model to the current thesis.

2.5.1 Style as attention to speech

The primary view in the attention to speech model is that the degree of attention paid to what we are saying controls our speech. With a great deal of attention paid, speech is increasingly formal, whereas where little attention is paid to speech, the output is more casual and closer to the vernacular (which is considered the most natural form of speech)⁶. Variationist sociolinguistic studies of this model analyse stylistic variation by introducing different tasks which act as a proxy for formal and informal registers, by implementing artificial strategies that aim to give the researcher control over the amount of attention their participants paid to their speech, particularly to draw the speech closer to the vernacular (e.g. Labov 1972; Trudgill

⁶ The notion of the vernacular was imperative to addressing the challenge of trying to record naturally-occurring speech data (as though the researcher weren't observing), but only being able to do so by observing and recording in the first place; the observer's paradox.

1974). Such studies combine a series of activities which allow the researcher to test whether attention to speech is an important constraint on variation.

Formal styles are elicited through reading tasks and casual styles and the vernacular are elicited via sociolinguistic interviews. In sociolinguistic interviews the physical situation does not change, and the interviewer remains the same, but types of speech activity are created to introduce different levels of attention to speech by interviewees as a proxy for different “situations” (Coupland 2007: 14). Speech style is then measured by quantitative means based on how frequently particular speech variants are used by speakers in specific contexts, providing a statistical definition of style for a sociolinguistic interview context. The interview was designed to be the most casual situation (involving least attention to speech) and a series of reading tasks followed which focus the interviewees on the linguistic variables in increasing increments by having participants read a passage aloud, then a word list which increases the level of formality again because reading isolated words increases participants’ focus on pronunciation. Trudgill’s (1974) work on Norwich English also used word lists, reading passages and casual and formal interviews as a proxy for varying register to elicit different styles from participants. As the register for each task became more formal, the speech style of participants became more formal. Regardless of participant social class, Trudgill noticed that the use of standard and prestige varieties increased with the formality of speech style, and the use of non-standard and non-prestige varieties increased with the informality of speech style.

Even within sociolinguistic interviews, certain topics are considered more casual than others (as they elicit the vernacular to different extents). Labov (1972; 2001) identified eight contexts pertaining to either careful or casual styles. These can include a “trigger situation” (Díaz-Campos 2005: 59), previously known as the ‘danger of death’ question, where the interviewer aims to gain their participants’ emotional involvement. In so doing (as hypothesised in Labov’s 1972 work), leading to the production of more casual speech. The

different topics discussed in the literature often form the basis of the structure of sociolinguistic interviews. The contents of the interviews in my own study will be discussed in the Methodology chapter.

However, the attention to speech model has been widely contested in the literature. It has been argued that speakers can focus attention towards producing any style – regardless of whether it is perceived as a prestige or vernacular style (Bell, 1984). The unidimensional continuum alluded to in Labov’s attention to speech model has been criticised for equating different tasks with levels of attention to speech, without taking external factors relating to register into account. As Finegan and Biber (2001:204) claim, mode of communication, setting, topic, purpose and relative status of interlocutors and their shared familiarity are factors contributing to the register of communication, therefore it seems overly simplistic to look at stylistic variation through attention to speech alone. Indeed, this model’s inability to account for interlocutors (interview and reading tasks cannot provide a proxy for different types of audience) led to Bell’s developing of audience design theory (1984) which is presented in section 2.5.2 below.

That being said, there are clear benefits to using the attention to speech model as a measure of stylistic variation as it allows the researcher to pattern the variation across styles in a way that is comparable and replicable in different speech communities, by looking at how people respond to situations devised to control their attention paid to their speech. Importantly, according to Labov (2001: 85) comparable interviews from representative studies of communities provide the most “solid and replicable” findings of intra-speaker variation (that is where the interlocutors and the social context are roughly constant). The meaning of stylistic variation stems from its quantitative patterning across speech communities, rather than in individual usage of variables under study (Labov 2001).

2.5.2 *Style as attention to others*

Bell's (1984) Audience Design model is based on the theory that speakers' speech style-shifts in response to their listeners. This model was considered less unidimensional than the attention to speech model – rather than a speaker adapting their style along a formality continuum towards the notion of prestige, this model presents the speakers as 'designers' of utterances for different audiences. Audience design assumes that persons respond mainly to other persons, and "that speakers take most account of hearers in designing their talk" (Bell 1984: 159). Bell based this framework (1984) on the perceived style-shift in New Zealand newsreaders on two publicly owned radio stations (Bell 1977). The same newsreaders, reading the same news were more likely to use conservative variants of /t/ voicing and /t,d/ deletion on the radio station which had an audience which was generally older and of a higher socioeconomic status, than the station associated with lowest socioeconomic status and youngest audience. Seeing as the audience was the only variable factor in what was essentially a natural matched guise situation (Bell 2001), the shifts in style were attributed to the listeners, rather than attention to speech.

In Bell's (1984) view, attention to speech was a mediating variable and that more focus should be on the situational factors which *cause* speakers to pay more or less attention; that is, the register. The most important register factor influencing individuals' style according to Bell's Audience Design model was the addressee. This notion was tested in Rickford and McNair-Knox's (1994) study of "Foxy Boston" (also mentioned in section 2.4.1), and her use of a number of sociolinguistic variables when talking to an African American and a White American English interlocutor about similar topics. Foxy was found to use more variants associated with African-American speakers when talking to the African-American interlocutor, and more variants associated with White American Vernacular English when addressing the White American English interviewer. The authors showed that addressee was an important element of register which is likely to affect the stylistic production of the participant, and that

the differences could be due to familiarity with the addressee as well as differences in ethnicity. However, they noted that variants varied more between topics than between the interviews overall, showing that more than one element of register can affect stylistic variation in any given situation. In the current study, different registers will be combined with varying addressees (in contrast with the Foxy Boston study), in order to fully understand the stylistic repertoires of pupils in Welsh-medium schools. Representation will need to be drawn from a number of pre-existing school-based registers; talk between students and adults and peer group talk etc.

In a recent expansion on his theory, Bell (2001) argues that even spontaneous adoptions of style on the part of a speaker are undertaken with an audience in mind, which downplays the role of a speaker's agency in creating and modifying style to construct meaning. Coupland's work (1980; 1985; 2001) shows how the link between context and style is "less direct, less determined and more subject to speakers' and listeners' creative agency" (Coupland 2007: 13), by providing an exploration of the construction of personae, as is presented in the section below.

2.5.3 Third wave variationist studies of style

The acquisition of sociolinguistic competence has been linked to the first and second waves of variationist research (see section 2.1). It should be noted, however, that style has also been analysed in third wave approaches to language variation, and the speaker design and repertoire approaches are briefly mentioned below.

2.5.3.1 Style as speaker design

Under the speaker design model, the study of style moves closer to the study of individuals' style-shifting practice, and away from large-scale quantitative analyses of speech communities. These studies of intraspeaker variation describe speakers' style-shifting in terms of creating an

identity or stance (Coupland 1985; Schilling-Estes 2002; Podesva 2007). An example of this type of research into stylistic variation is Coupland's work on stylization which is typically associated with high performance events such as public speaking and radio host speech (see Coupland 1985; 2001).

The focus of work under the speaker design model is on an individual's identity construction through style-shifting, rather than investigating the direct links between style and register as set out in variationist work. Labov - the pioneer of first wave variationist sociolinguistic research - himself (2001: 85) notes that the naturalistic ethnographic phenomenon of the second and third waves of research appears to be the most "immediately appealing and satisfying", enabling the researcher to observe real-life cases of variation in speakers' day-to-day interactions. The speaker design model, however, does not help to address the main aims of this research, which set out to find broad generalizable sociolinguistic patterns in the stylistic repertoires of a large sample of students from different home language backgrounds. As a consequence, this thesis will not attempt to collect qualitative ethnographic data on a small number of students in order to explore issues of identity and stance construction.

2.5.3.2 Style as a repertoire

Another approach is to examine the stylistic repertoires of speakers in order to move beyond individual speaker agency as seen in speaker design approaches, and to explore how individuals are embedded within wider social structures (Sharma 2011: 465). In her study of British English and Indian English dialect speakers, Sharma (2011) drew a comparison between a standard variationist analysis of common variants of Indian English varieties in a repertoire analysis, which presented each individual's 'life –worlds' as opposed to constituting a formality continuum (Sharma 2011: 474), as would have been presented by early work on style, such as Labov's previously mentioned work (1966, 1972).

Sharma claims that a benefit of using the repertoire analysis approach is being able to look at family domains which have historically been excluded by common interview-based methods (Sharma 2011: 465). The focus on family linguistic practice, however, is not suitable for the current study, considering that half of the participating students do not use Welsh at home with family. Added to the limitation of this model of variationist analysis is the non-comparability of contexts in the repertoire study – the ‘life-worlds’ (Sharma 2011: 474) of participants did not naturally line up, meaning few generalisations about the stylistic practice of participants could be made.

2.5.4 Summary of stylistic variation

The study of style in variationist sociolinguistics has inspired a series of theoretical developments which have become well-established models in the field. The attention to speech model uses different tasks as a proxy for formal and informal registers in order to elicit stylistic variation from participants (e.g. Labov 1966, 1972; Trudgill 1974), while audience design models claim that Labov’s “mechanistic attention variable” (Bell 1984: 150) does not do enough to relate style-shift to specific situational factors which cause it, making it a mediating variable. The current work draws from these two models as it uses tasks and interlocutors to elicit different styles from speakers (discussed further in section 4.5). In reviewing the present literature it has become increasingly apparent that the interaction between style and register are central to exploring patterns of sociolinguistic variation, and thus the acquisition of sociolinguistic competence. Further discussion about register is presented below.

2.6 Register

Here, I present a broad definition of what is meant by register in the context of this work. Recall that this thesis aims to explore the style-shifting practice of pupils in Welsh-medium education. I intend to capture changes in speech style on the basis of register, i.e. “what situational

constraints are operative” (Coupland 2007: 15). As Trudgill (2000: 82) sets out, “registers are an example of a particular kind of language produced by a particular kind of social context”. These “social contexts” can be defined in a number of different ways. Early sociolinguists sometimes used the concept of register as linked to who the speakers are addressing (addressee registers) e.g. ‘baby talk’, ‘foreigner talk’ and ‘elderspeak’ (Coupland 2007: 57), which is similar to Bell’s audience design model (1984) and more recently linked with the speech style paradigm. In fact, as well as addressees, register can be defined by the attention paid to one’s speech during an interaction (see Labov 1972); material conditions of the situational context (Ure and Ellis 1977); the topic or type of subject matter, for example religion *vs.* small talk about the weather; and the type of speech itself, such as Schilling Estes’ (1998) work on performance speech registers, where speakers attempt to display for others a certain language or variety (Schilling-Estes 1998: 53).

Crucially, each of these external situational factors determines the formality of the register, and thus elicits the production of formal and informal styles. According to Chambers (2009: 4), formality increases in direct proportion to the number of social differences between participants; therefore, as the register becomes more formal, so the style used tends to follow. The methodology chapter (chapter 4) provides full details of the registers designed to elicit different speech styles.

Registers are associated with a set of characteristic linguistic features (which can be stylistically marked) and closely identified with specific contexts and uses. A register can be comprised of a number of situational factors such as purpose, audience/addressee, planning time, etc. From the sociolinguistic literature, register variation has been used to explain the style-shifting observed in written and spoken modes (e.g. Tagliamonte 2016) and on the other hand, linguistic characteristics associated with different styles can help to identify registers (Biber and Conrad 2009); thus registers can influence style, and at the same time, style can

determine a register. Furthermore, it proposed that register is nuanced (Tagliamonte 2016) – although they are defined according to situational factors, not all linguistic variables will be directly linked to each register, as that might be dependent on social evaluation, that is, whether they are regarded as prestigious or stigmatised. For this reason, each analysis chapter (chapters 5, 6, and 7) will provide an overview of the anticipated ways in which register may affect each feature under study, particularly in relation to the language experience of the individual, seeing as registers correspond with existing situations of language use (Ure and Ellis 1997: 197). Further discussion about the types of Welsh features likely to be encountered in various registers is discussed later, in chapter 3. By looking at styles in a number of registers, we will be able to establish the extent to which pupils vary their language accordingly in Welsh-immersion schools.

2.7 Implications

The purpose of the current thesis is to observe the stylistic repertoires of students at Welsh-medium schools. Currently, little variationist work exists in the Welsh context (e.g. Prys 2016) and little is known about the stylistic repertoire of those in Welsh-medium education, and less still about the differences which may exist between home language backgrounds in producing this variation. The comparison of home language backgrounds makes the current work particularly suited to quantitative variationist analysis, where ‘speaking differently’ is measured based on how frequently particular speech variants are used by speakers, offering a statistical definition of ‘a style’ (Coupland 2007: 15). This will allow the researcher to establish and analyse broad trends in the context of Welsh on a relatively large scale. Recall that in the case of French immersion, pupils showed a limited grasp of more informal styles (Rehner and Mougeon 1999; Mougeon and Rehner 2001; Mougeon *et al.* 2010).

The Canadian studies mentioned previously used sociolinguistic interviews to elicit lexical and morphosyntactic data in order to point to participants’ sociolinguistic competence.

As discussed previously, Labovian-style interviews aim to use attention to speech as a proxy for different registers, eliciting different speech styles from participants. The current research goes beyond relying solely on sociolinguistic interview data, by also capturing naturally occurring speech in a number of settings outside of the interview, where audience and topic among other factors will vary, potentially causing a shift in speaker style. A more varied methodology might offer access to a more varied repertoire; although crucially, the sociolinguistic interview will allow the data to be comparable across the immersion contexts of Wales and beyond.

This analysis will allow me to draw conclusions about the sociolinguistic competence of students in this minority-language educational context as it will provide evidence of their ability to vary their speech in a variety of interactions, both formal and informal (Bayley and Regan 2004). Previous work has focused on a range of sociolinguistic variants which differ in terms of their stylistic markedness, through a system of variant categorisation based on the following examples of criteria (Mougeon *et al.* 2010: 9).

- The degree to which variants conform to the rules of standard speech or not
- Whether they are associated with speakers from upper or lower social classes
- Whether they are associated with formal or informal registers
- Whether or not their use is stigmatised

This chapter has outlined the variationist framework of the current study, and has detailed the literature on the acquisition of sociolinguistic competence in previous work. I have discussed how I define style and register based on previous studies of sociolinguistic variation. I now turn to the next chapter, where I discuss what is known about varying styles and registers in Welsh. Based on this, the reader will have a good understanding of the types of styles and registers likely to be encountered by pupils in Welsh education, and thus the chapter provides a backdrop against which sociolinguistic competence in that context will be measured.

3 Stylistic variation and register in modern Welsh

The current chapter reviews research on the styles of Welsh and on matters surrounding standard and educational varieties of the language. It is through the review of previous work that we can draw conclusions about the stylistic categorisation of different features. This will provide a key for understanding which types of style are used in different contexts by the participants of this study, and thus point to the extent to which pupils from different home language backgrounds have acquired sociolinguistic competence.

3.1 Stylistic variation in Welsh

3.1.1 Introduction

Style is considered a key factor in Welsh morphosyntactic variation (Ball 1988: 60), making Welsh a prime candidate for variationist research on style-shifting practices. Indeed, Welsh has been described as a language “characterised by an extremely broad range of registers depending on such factors as the level of formality, the medium of communication (writing versus speech), the age of the participants in the interaction... and the setting” (Rottet & Morris, 2018: 8). Patterns of stylistic variation have been mapped by some sociolinguists (see Fife 1986; D. Willis 2016; Prys 2016, for example) and an example of the type of styles possible in Welsh, (extracted from Rottet & Morris’ (2018: 8) account) are presented below. Although not an exhaustive list⁷, these are all possible translations of the English phrase “I will not sing” (from the most formal literary to the most colloquial) and are a good indication of the difference between stylistically-marked variants of Welsh:

Ni chanaf
Ni chanaf i ddim
Chanaf i ddim
Chana’ i ddim

⁷ For example, the list does not include reference to periphrastic variants such as *wna i ddim canu* which are further discussed in Chapter 6.

3.1.2 *Written vs spoken Welsh*

The medium (or register, see 2.6) of communication (writing and speech) is reported above as being a predictor of style-shifting in the paragraph above. The gap between the standard, written variety and spoken Welsh is substantial and even describes these as two strikingly different *languages*, rather than varieties of same language (Fife 1986: 145).

D. G. Jones (1988: 162) points to three features where literary and spoken Welsh differ systematically from each other: a) pre-verbal particles (literary language requires the negative *ni* and *nid*, and interrogative *a* and *ai*, which tend to be omitted from spoken language), b) postnominal possessive pronouns (generally omitted in the literary unless emphatic *dy gar di/dy gar* “your car”), and c) inflected verbs rarely used in the vernacular, which shows a preference for periphrastic constructions⁸. B. M. Jones (1993) also discusses a number of inflected verbs existing in literary Welsh which are hardly used in the vernacular. The extent to which both varieties are used and mastered is thought to vary; colloquial Welsh has been previously described as the native tongue for Welsh speakers, and the written form as an artificial language which requires formal teaching (King, 2016).

Ball, Griffiths and Jones (1988: 192) argue that the difference between literary and colloquial varieties of Welsh constitutes an instance of diglossia. Diglossia refers to a situation where a community uses two highly divergent codes or varieties, comprised of dialects and a superposed formal variety which is learned largely through formal education and not used by any sector of the community in everyday conversation (Ferguson 2007: 42). The situation for Welsh is no different according to Ball *et al.* (1988: 192) who argue that:

The high variety, known as *yr iaith safonol*, 'the Standard language', or *yr iaith lenyddol*, 'the literary language' ... is found, as is usually the case with diglossic situations ... in literature

⁸ As I mention in the introduction chapter of this thesis, possessive pronouns and inflected vs periphrastic verbs are two of the three features analysed in the current study.

and in all manner of publications ... The low variety, known as *yr iaith lafar*, 'the spoken language', is the everyday spoken language of the Community, within the family, with friends, and usually in all contexts where the high variety is inappropriate.

Despite these reported differences, linguists have traditionally disregarded the divergence between the two and have made assumptions about the language as a whole, usually that features of standard, literary Welsh are over-generalized to colloquial, vernacular Welsh (Fife, 1986). Indeed, B. M. Jones (1993) observes that terms like 'correct' and 'incorrect' are frequently used to describe the difference between written and spoken Welsh, respectively and of the colloquial form as being a "reduced" form of the former (Fife 1986: 145).

It should be noted that vernacular forms in Welsh are subject to geographical variation. It is well-attested that Welsh is subject to significant regional variation. Dialectological studies such as B. Thomas and P. W. Thomas (1989) and D. Willis (2014) have mapped these varieties extensively, and more recent variationist sociolinguistic work has used third wave ethnographic methods to understand patterns in local varieties in Wales, such as the Welsh of Cardiff (Gruffydd 2022). Generally, according to A R Thomas' *Linguistic Geography of Wales* (1973: 14) three overall dialect areas (north, midlands and south) each of which is divisible into east and west regions, make up the six main local dialects of Wales, however, minor speech areas can also be identified at a lower level. Often, the main differences are observed between what is considered 'northern' Welsh and 'southern' Welsh. In fact, adults learning Welsh through *Dysgu Cymraeg* will opt to follow a North or South dialect stream. In order to demonstrate the differences between the literary and colloquial forms (both northern and southern), Table 3.1 shows the first-person present tense paradigms of *bod* "to be" (adapted from Ball 1988: 62).⁹

Table 3.1: Example of written vs spoken usage (north and south) from Ball (1988: 62)

Written (literary)	Northern spoken	Southern spoken
Yr ydwyf i	'dw i	rw i

⁹ The traditional southern form is *wi*, although *fi* is common and *dwi* will be heard in the southeast.

There are noteworthy differences in Table 3.1 between the spoken varieties of northern and southern speakers and between the written and spoken varieties. The example above of literary *yr ydwyffi* is unlikely, however, to appear in the writing of everyday Welsh, and further to this, speakers in both the north and the south may be using a more standardized variety in everyday speech, particularly in more formal contexts. Indeed, today, many would deem it perfectly acceptable to use *rydw i* in both those written and spoken contexts¹⁰.

This change in use of different styles is in part following a number of societal developments which have drawn the written and spoken styles of Welsh closer together. The domains in which Welsh is used have increased over time (following an increase in public services available through Welsh and the rise of Welsh-medium education), leading to the development of a standard spoken variety (as opposed to dialectal vernacular forms), and a more common variety of written Welsh (Cymraeg Byw) both of which are somewhat contested in the literature. These are discussed further, below.

3.1.3 The spoken standard

Although well-attested in the literature (e.g. Ball *et al.* 1988), some have recognised the shortcomings of a simple comparison between ‘literary’ and ‘colloquial’ varieties, because of the strategic importance of a common standard spoken form, which bridges some of the dialectal extremes and is understood and accepted in all the regions (D G Jones 1988: 136).

Whether or not this common standard exists, however, is unclear. Fife (1986: 145) has described a standard spoken Welsh which is closer to the literary, written form of Welsh, and associated with formal speaking occasions such as preaching, lecturing, broadcast journalism and in the National Eisteddfod (Fife 1986: 145). This description of a standard spoken variety

¹⁰ Indeed, using a search in a large corpus of spoken and written Welsh (CorCenCC, Knight *et al.* 2020) *rydw i* is clearly used in both modes.

is contested, however. In part, this is because in Welsh, unlike in British English, there is no nationally accepted non-regional standard accent (Ball 1988: 50); that is, each regional dialect has its own spoken standard. The evidence for these different spoken standards, however, is lacking. Indeed, very little work has sought to categorise the standard nature of spoken Welsh.

However, some research has pointed to domains or registers where a standard spoken Welsh might be used. Welsh in the media is one such register (Ball, Griffiths and Jones 1988), which has been associated with high varieties of the language, with some broadcasters being regarded as the “purveyors of standard Welsh” (Smith 2000: 339), which has been recently followed by an attempt to de-standardize some content to bring it closer to younger speaker norms (Prys 2016: 51). This supports Rottet and Morris’ (2018: 8) argument that style is largely predicted by the age of the speakers.

What, therefore, do we know about young speakers’ use of a standard spoken variety? Welsh in the education system has seemingly promoted the acquisition of the standard variety of Welsh, which has been argued to be responsible for the demise of local dialects (M. C. Jones 1998)¹¹. I will introduce what is known about the stylistic repertoires of pupils in Welsh medium education in section 3.1.6. But for now, it is important to consider the introduction and establishment of *Cymraeg Byw* [Living Welsh] within the educational domain, which is thought to have an influence on the Welsh used by young people. This is presented in the section below.

3.1.4 *Cymraeg Byw*

As discussed, formal and informal varieties of Welsh have been described as being very distinct. In the 1960s, a committee of education professionals sought to align the literary language with its spoken counterpart, by creating and standardizing a variety of Welsh called

¹¹ The current work will not look specifically at the various dialects of Welsh, but the reader can find more about them in Thomas (1973) and B. Thomas and P. W. Thomas (1989).

Cymraeg Byw [Living Welsh] (Rees 2017). The publication of the three booklets on *Cymraeg Byw* forms (Cymraeg Byw 1964; 1967; 1970) was fundamental to the development of Welsh as a Second Language, where teachers sought to define a standard spoken form (Davies 1988: 200), at a time where teaching resources such as *français fondamental*, which focussed on teaching French to foreign learners based on the reality of its usage (Klinger and Véronique 2006: 2) were becoming increasingly popular.

The standardized written and colloquial *Cymraeg Byw* was met with contempt by many who thought artificial evolution of the language has impacted on the types of linguistic register in Welsh (see Fife 1986 for a discussion of this). However, by the release of the third booklet, it was expected that the variety be used by L2 and L1 Welsh speakers alike. Since, *Cymraeg Byw* is representative of a register which can be seen in the work of modern writers and simultaneously acceptable when used in conversation (Davies 1988: 208) as exemplified in 3.1.2. *Cymraeg Byw* was instrumental in the resurgence of Welsh as a second language and Welsh-medium education in anglicised areas and has inevitably brought about some changes to the nature of the language more broadly. The extent to which *Cymraeg Byw* forms are used in Welsh schools today is under researched, but I present some examples in Table 3.2 of *Cymraeg Byw* forms in relation to the more standard written equivalent (adapted from Bolitho 2016).

Table 3.2: *Cymraeg Byw* vs standard forms

Cymraeg Byw	Standard form	English gloss
Rydw i	Rwyf/yr wyf/yr ydwyf fi	I / I am
Dydw i ddim	Nid wyf	I am not
Es i	Euthum	I went
Fy ngwaith i	Fy ngwaith	My work

Y stori yma	Y stori hon	This story
-------------	-------------	------------

In order to demonstrate how *Cymraeg Byw* forms are likely to fit with other reported styles of Welsh, and how those in turn are likely to be used by young speakers of Welsh, I present a working categorisation of stylistic variation in Welsh below.

3.1.5 Towards a categorisation of variation

The varieties of Welsh have grown in accordance with the profile of the language following attempts to revitalise the language, leading to a wider range of registers needed by speakers (Thomas, 1988a: 19). The contrast of formal and informal Welsh is a major stylistic difference of the language (Ball *et al.* 1988: 192), but having said that, it has also been argued that the variables that distinguish them are not easy to define precisely (Jones, 2010). As previously mentioned, some authors have shown inflected/periphrastic forms and affixed pronouns to be distinctive features of the difference between formal and informal varieties (D. G. Jones 1988: 162; B. M. Jones 1993). Further distinctions in the literature are cited and summarized below.

Thomas (1996: 13) describes the formality of the Welsh language (similar to other languages) as residing on a continuum, between hyper formal and hyper informal, providing a means of varying the form of interaction depending on the social context. In their online guide to writing Welsh, Welsh Government (2015) defines four Welsh registers in the *Arddulliadur*, each supplied with a comprehensive list of the types of grammatical and stylistic characteristics of the language. These range from classical Welsh, to formal, to informal, to very informal spoken styles. In contrast, *Canolfan Bedwyr* (2015) describe ten registers existing in Welsh, ranging from high prestige archaic language (encountered mostly in written texts) to low prestige slang (which is mostly spoken).

Table 3.3 is adapted from Young (2019: 10) and shows the categorisation of registers from the works cited above. It is important to note that both the *Arddulliadur* and Register

Typology Matrix associate highly informal language with spoken language, which supports the assertion that spoken and literary language are markedly distinct from one another in Welsh. Crucially, however, the contents of the table below challenge the assertion that there is a simple distinction between the standard and the spoken diglossia as previously argued by Ball *et al.* (1988: 192). It is possible to map different registers of writing and speaking onto the framework below, and thus it is possible to visualise the expectations about the types of styles likely to be encountered in those registers.

Table 3.3: Categorising Welsh stylistic variation (based on Young 2019: 10)

Formality continuum (Thomas 1996)	<i>Arddulliadur</i> (Welsh Government 2015)	Language Register Typology Matrix (Canolfan Bedwyr 2015)	Expected range
<div>Hyper formal</div> <div>↑</div> <div>↓</div> <div>Hyper informal</div>	Classical	Archaic	<div>↑</div> <div>↓</div>
		Classical	
	Formal	Formal	
		Technical	
	Informal	Neutral	
		Simplified language/Clear Welsh	
		Informal	
	Spoken/very informal	Very informal/ spoken	
		Regional	
		Slang	

The higher end of the formality continuum pertains to religious texts and legal documents (Canolfan Bedwyr, 2015; Welsh Government, 2015). The current study will focus on stylistic variation between different registers likely to be encountered and produced by students at Welsh immersion schools and the sole focus will be on spoken Welsh. Therefore, the styles used by speakers are expected to reside along the more informal end of the continuum of formality.

The previously mentioned *Cymraeg Byw* forms would probably most closely align with neutral or simplified language; the neutral register is thought to be representative of school pupil essays (Canolfan Bedwyr 2015), which is likely to be aligned with more standard, formal spoken varieties of Welsh. Indeed, the spoken varieties encountered in more formal school settings are likely to reside around the neutral/simplified category, whereas more varieties in more informal settings are likely to occupy the lowest echelons of formality. In the fourth column of Table 3.3 I list the expected range of styles to be encountered in the current study.

Before presenting in further detail the styles expected to be taught and used in Welsh-medium education (the descriptions of which are few and far between), I first present previous sociolinguistic studies on Welsh style according to register. In doing so, I will point to the main features that have been considered stylistically marked in Welsh. This will help provide a justification for the features chosen to be studied in the current thesis, as they are markers of stylistic variation.

3.1.6 Sociolinguistic studies on Welsh style according to register

Early sociolinguistic work in Welsh by B. Thomas (1980) analysed the speech of one participant in the Afan Valley near Swansea, South Wales. The use of a number of stylistically marked features was examined across three registers. In the reading register, the participant read a piece she had written for women's institute literature; the formal register consisted of an address to the minister on behalf of her church; and an informal register was comprised of the

local dialect as spoken with the researcher herself in the participant's home. Not only does this work show that the speaker varied her language substantially between one register and the next, but it also shed light on the speaker's awareness of her selection of register-appropriate styles and of patterns of language variation generally (Thomas 1980: 579). The speaker in this study showed both awareness of and confidence in producing stylistically-appropriate language in each register; thus a high level of sociolinguistic competence. However, the number of tokens varied substantially between registers, the recordings of more formal registers being much shorter than the casual spoken register. Indeed, for the most formal language, which came from the church, only one interaction was recorded. According to B. M. Jones (1993), more experimental methods are required in order to gather formal data, which is more difficult to come by in Welsh.

A more recent study by Prys (2016) looked at the stylistic practice of different radio presenters on programmes where registers were distinctly different. The distribution of each programme was placed on a continuum of formality according to influencing factors such as the subject matter, familiarity between interlocutors, amount of laughter and overlapping speech and social factors such as prestige. Table 3.4 shows the registers associated with each radio programme discussed by Prys (2016).

Table 3.4: Registers of Welsh radio programmes, adapted from Prys (2016: 298)

Tudur Owen	Gethin a Ger	Dewi Llwyd Ar Fore Sul	Post Prynawn
Less formal register		More formal register	
(vernacular / naturalistic speech style)		(ideological / careful speech style)	

From the features examined in this study, code-switching between English and Welsh was found to be generally linear, with less formal programmes showing more frequent code-switching than formal programmes (Prys 2016: 341). The mutation variable (a) and (â/gyda) was also found to be stylistically marked, which Prys suggests is representative of a diglossic

situation between the most formal and most informal programmes, where the more formal programmes deviate significantly from community norms and towards high prestige variants (Prys 2016).

Other Welsh work on stylistic variation according to register (Young 2019) has compared the reported use of stylistically marked features in various classroom and outside-of-classroom contexts. According to that study, speaking *vs.* writing contexts were likely to produce significant differences in the use of stylistically marked features, whereas the classroom *vs.* outside of classroom contexts produced statistically significant different uses of variants in other features. Teachers reported that pupils used more formal morphosyntactic and phonological features in more formal contexts (including classroom speech and written work), whereas more informal features were used in less formal contexts, thus pointing to pupils' sociolinguistic competence.

From this work on Welsh, it is clear that registers in Welsh are seemingly well defined, and stylistic markers can be effective indicators of sociolinguistic competence, as is the case of French immersion which is well-attested in the literature (Rehner and Mougeon 1999, Mougeon and Rehner 2001, Rehner *et al.* 2003; Mougeon *et al.* 2010). What remains to be explored, however, is the extent to which pupils actually vary their style according to register (rather than what their teachers report), as well as determining the types of register existing in Welsh-immersion schools. Further to this, the current research will show whether there are any differences in level of sociolinguistic competence between pupils from Welsh-speaking homes and those not from Welsh-speaking homes. In order to understand this, the following section outlines the type of Welsh likely to be encountered in Welsh-medium education.

3.1.7 Styles and registers in Welsh-medium education

It is yet to be seen through a variationist lens whether an educational 'standard' Welsh has developed as a variety following the long history of Welsh and English contact and

revitalisation efforts. The Welsh examination board, Welsh Joint Education Committee (WJEC, 2015:5)¹² states that the curriculum for Welsh as a First Language (in Welsh-medium schools) should help students to “understand the impact of variations in language...adapting speech and writing to different situations”, and Welsh as a Second Language (in English-medium schools) GCSE curriculum aims to promote the use of Welsh in “further studies, in the workplace and in their communities” (WJEC, 2017: 3). Teachers of Welsh in both types of school are expected to teach varieties of Welsh which are appropriate in a number of different registers; conforming to academic, workplace and more casual community norms, thus improving the sociolinguistic competence of their students.

The reality from other international immersion contexts (as discussed in section 2.2.2.1), however, is that immersion schooling fails to prepare students to use more informal varieties of language, limiting their sociolinguistic competence (Mougeon *et al.* 2010). Preliminary work by Young (2019) found that teachers reported students to use standard varieties of the phonological variants plural *pethau* ‘things’ over *petha/pethe*, and were more likely to use (f) endings rather than the more vernacular variety (e.g. *nesaf* over *nesa* ‘next’), across three different contexts of use. This is also supported in other variationist Welsh work by M. C. Jones (1998) which suggests that some standard forms are taking the place of local dialect features. Work on Welsh variation in young people’s speech suggests that for some stylistically marked features, students have been reported to over-use formal variants in informal registers, which is in keeping with previous Canadian research. These claims, which in part are based on teacher perceptions rather than student use, will be confirmed in the current study.

Other work such as Gruffydd (2022), has found that pupils in Welsh immersion schools to significantly vary their speech according to context. In his analysis of a reading task and a

¹² Although this guidance pre-dates the new national curriculum, *Cwricwlwm i Gymru*, this guidance was in place at the time where participating students were in the education system.

sociolinguistic interview, he found stylistic phonological variation in the language of pupils in a Cardiff Welsh-medium school, and claims as a consequence that pupils have access to different styles and registers in Welsh (2022: 188). Whether this is also the case for morphosyntactic features, however, is yet to be seen.

There are a number of questions which remain about the varieties of Welsh present in the educational domain. It is not known whether speakers of Welsh who do not speak Welsh as a home language might have the tendency to over-produce formal variants, in line with previous research from other immersion settings. Educational emphasis on correct and written usage might restrict speakers to more literary language, particularly in formal classroom settings. Where Welsh is not a dominant community language, pupils might be exposed to fewer examples of local dialect features, thus restricting their use of casual variation further (M. C. Jones 1998). Although little is documented about English home language (EHL) speakers' use of Welsh, the following section summarises main findings from previous research.

3.1.8 The stylistic variation of English home language pupils

As previously discussed (see section 1.2), the situation in Wales is comparable to that of other minoritized languages, in the sense that the majority of Welsh-English bilinguals, particularly in more anglicised areas (Cardiff Council 2024) acquire Welsh through immersion education as opposed to acquiring the language through parental transmission (Jones 2012), making the majority of pupils EHL pupils. The extent to which an EHL variety of Welsh exists, is yet to be found. Recall that EHL pupils in Welsh immersion education are taught in the same classrooms as those who acquire the language from their caregivers, which sets the Welsh context apart from that of, for instance, French immersion in Canada.

Having said that, previous work on L2 speakers of Welsh has reported similar findings to other work on sociolinguistic competence in French immersion settings. As discussed previously, Hatton (1988: 251) compared pupils' ability to shift between a formal and an

informal setting by measuring their use of mutations. The EHL group exhibited minimal style variation, which was attributed to the exclusive influence of formal school patterns on their speech, rendering them “monostylistic” (Hatton 1988: 251). This is further supported by a qualitative account in unpublished doctoral work by Hughes (2013: 276). Here Hughes argues that many pupils who do not speak Welsh outside of class do not have the informal styles of Welsh to the same extent to be able to express their feelings, experiences, and opinions and so on with their peers in less formal situations.

Further to this, previous work on the mutation system which considered how children integrate English borrowings into the Welsh mutation system found that primary aged pupils who do not speak Welsh at home were less likely to do this than those who do use the language at home. Bellin (1998) argues that this is not acquired until after children reach 5 years old, whereas Gathercole and Thomas (2005) found evidence of incomplete mutation systems when pupils were aged eleven. The slow acquisition could be due to a limited critical mass of exposure, which is required in order to generalize mutations across English borrowings.

Not all research, however, has found limited variation in EHL speech. In her study of ‘new speakers’ in south-east Wales, nowadays an English-dominant area, Robert (2009) found that some speakers in the study were seen to mix varieties, using combinations of northern and southern features, as well as using literary forms in their vernacular. New speakers therefore use regional varieties not traditionally associated with their localities (potentially in order to access prestige), and arguably do not systematically acquire more informal features in their speech. Therefore, we can surmise that though variation has been found in EHL speech, the extent to which it is systematic is unknown. Furthermore, we know very little about whether the patterns of variation by EHL speakers match their WHL counterparts.

In terms of phonetics, some work has found that EHL speakers of Welsh differ from WHL speakers. Morris (2013) compared young adults from English home language and Welsh

home language backgrounds and their (r) variation and found that ‘new speaker’ participants (those from English dominant communities and English home language backgrounds) did not produce the typically Welsh alveolar trill nor the post-vocalic /r/, which was common in the speech of ‘traditional’ Welsh speakers. This distinction, however has not been found in much other work comparing home language backgrounds (Mayr *et al.* 2017; Mennen *et al.* 2020; Gruffydd 2022).

To sum up, some research has found EHL to be over-reliant on formal variants (Hatton 1988), others have found their repertoire of variation to be incomplete (Gathercole and Thomas 2005), or unsystematic (Robert 2009), or to display different patterns than WHL speakers (Morris 2013). Other work discussing home language backgrounds, however, suggests that EHL pattern similarly to WHL speakers (see Mayr *et al.* 2017; Mennen *et al.* 2020; Gruffydd 2022). From their point of view, the classroom that both groups of pupils share, and close-knit peer networks level out any difference between the two types of speaker. The aim of this research is to find to what extent morphosyntactic variation can be predicted based on home language backgrounds, in order to address some of the current gaps in understanding. The following section summarises the current chapter.

3.2 Summary and implications

Style in Welsh is arguably affected by the diglossic interplay between the standard and vernacular codes of the language (Prys 2016: 42). B. M. Jones (1993) argues that the two main influences on the standardizing of spoken Welsh are the people and the task involved in the situation at hand (the register), and yet very little work exists in this field (B. M. Jones 1993: 191). Clayton’s (1978) work points to the effect of standard language ideology on people’s efforts to speak their ‘best’ Welsh. However, beyond notions of what is traditionally regarded as ‘correct’ usage of the language, little research has been conducted on the stylistic variation present in the speech of Welsh speakers.

Recent revitalisation efforts have increased the domains of Welsh language use, and newer varieties of Welsh (such as *Cymraeg Byw*) have emerged to suit this evolving landscape. I have shown that a number of different categories of Welsh style exist, and in subsequent chapters on the features analysed for the current study, I will show how different variants of those features are associated with the styles presented here. The focus of the current study will be to look at specific features of Welsh that are found in different speech styles, in order to understand the stylistic repertoires of Welsh immersion pupils (from different home language backgrounds). From there, we can assess the extent to which there are pronounced differences between non-Welsh home language pupils and those who do speak the language at home, as has been found by some authors, who have demonstrated that EHL speakers are not acquiring L1 norms (Hatton 1988, Robert 2009, Morris 2013).

Using a variationist sociolinguistic analysis of Type 2 (horizontal) variation, I will examine the use of three stylistically marked features of Welsh in order to ascertain to what extent patterns of variation in WHL speakers are replicated by EHL speakers. The precise methodology will be presented in the next chapter.

4 Methodology

4.1 Overview of the methodology

Chapters 2 and 3 presented the importance of analysing speech in different contexts of use in order to examine patterns of stylistic variation. This chapter will discuss those contexts used to elicit varying styles from participants. As will be discussed, approximately 34 hours of speech data in different contexts were collected from participating students. The precise research design and method are covered in the current chapter. In it, further information is given on the research communities (section 4.2) and the participating educational institutions. I outline the implications of COVID-19 and ethics on data collection in section 4.3, and then I present my participants in section 4.4. An overview of the research design is then given in section 4.5, and the chapter ends (see section 4.7) by describing the comparative analysis used to determine whether Welsh speakers from different areas and home language backgrounds acquire the same patterns of variation.

4.2 The research communities

Gwynedd and Cardiff were chosen as research sites based on recent census and school data indicating the extent to which Welsh is spoken and used in the community. According to previous studies in other contexts on the acquisition of sociolinguistic competence, less exposure to the target language in the community can limit pupils' language variation (Mougeon *et al.* 2010), and thus it was hypothesised that the area with more Welsh spoken might affect the acquisition of sociolinguistic competence. Whereas this section introduces some participant numbers, broader details about participants and recruitment come in section 4.3.3. The following sections will present the two areas under study: Gwynedd and Cardiff.

4.2.1 Gwynedd

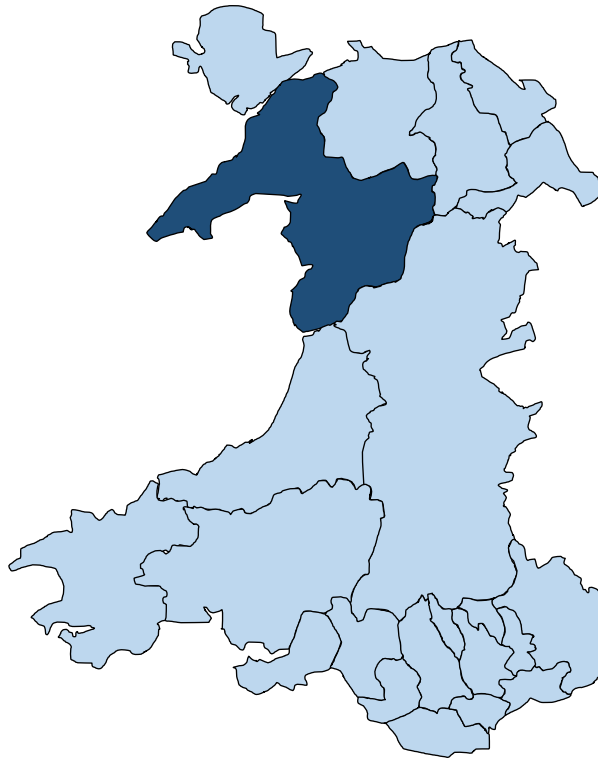


Figure 4.1: Gwynedd in the north-west of Wales

Gwynedd (location presented in Figure 4.1) is a local authority in the north-west of Wales. Gwynedd is traditionally perceived to be part of the Welsh-speaking heartlands (Coupland 2012); that is, an area where the Welsh language is prevalent in the community. According to the 2021 Census (Welsh Government 2022), 64% of people in Gwynedd reported being able to speak Welsh, but this percentage does vary across the area. The latest Annual Population Survey (APS) (2023), which collects data by a different method, found that 68% of Welsh-speakers in Gwynedd use the language on a daily basis. Sparsely populated, Gwynedd's density (persons per square kilometre) is 49.4, making it a predominantly rural area; however, there are some urban centres, such as the city of Bangor and smaller towns of Caernarfon, Blaenau Ffestiniog and Dolgellau in the local authority.

Intergenerational transmission of the language is more likely in Gwynedd (and Ceredigion) than any other area of Wales (Welsh Government 2023b). According to Pupil Level Annual School Census (PLASC) data at the time of data collection, of the total 4,185 secondary school pupils in Gwynedd who were fluent in Welsh, 3,400 pupils spoke Welsh at home, whereas 785 (18%) did not speak Welsh at home (StatsWales 2021).

Across Gwynedd, the demographic trend in recent years has been that “young people were leaving and the only people who were likely to take their place were non-Welsh speaking older people.” (Joint Planning Policy Unit 2015: 27). This is seen to affect the proportion of Welsh speakers in the local authority. The local authority sees bilingual education as playing a key role in language maintenance. In schools in Gwynedd, Welsh is the sole or main medium of instruction for the majority of pupils (where Welsh is the medium of instruction 80%+ of the time) (StatsWales 2021). According to the school census results in 2019, Gwynedd had the highest number of Welsh medium schools in Wales, (97 out of a total 99 schools) (Welsh Government 2019). More detail on the education institution that took part in Gwynedd is provided below.

4.2.1.1 *Coleg y Fro*

Coleg y Fro (pseudonymised for this study) is an educational institution which caters to a wide geographical area in Meirionnydd. It caters to post-16 learners from four different feeder secondary schools, travelling within a radius of approximately 25 miles of the campus, spanning between Tywyn and Blaenau Ffestiniog. All four feeder secondary schools are categorised as Welsh-medium schools.

Although not all pupils follow their A Level courses through the medium of Welsh, teachers at *Coleg y Fro* use both Welsh and English in the classroom to cater to all students, and the campus was found to have “a very strong Welsh identity” (Estyn 2017a¹³: 26). Many

¹³ Reference anonymised in full bibliography.

students at *Coleg y Fro* have strong reading, writing and communication skills in Welsh and use it “naturally” around campus (Estyn 2017a: 31).

A total of 10 pupils participated in the study from *Coleg y Fro*. They were from a class of 12 following a bilingual A Level course.

4.2.2 Cardiff

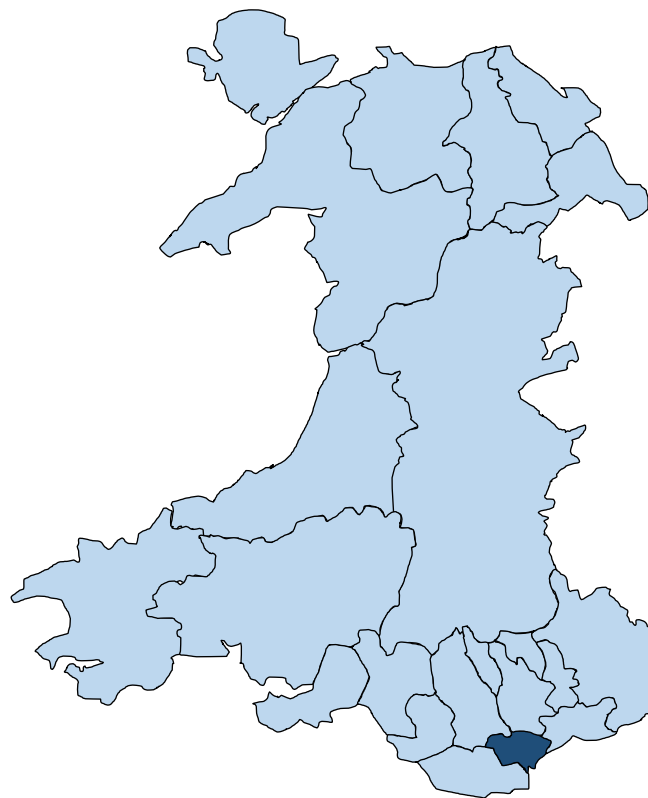


Figure 4.2: Cardiff in the south-east of Wales

Cardiff (location shown in Figure 4.2) is a local authority located in the south east of Wales. It is home to the capital city of Wales which bears the same name, and is a densely populated metropolitan area, with 2,620 persons per square kilometre. More than 335,000 people live in the local authority. It has been described as an urban English-dominant area by previous studies on Welsh-medium education (Hickey *et al.* 2014), however it should be noted that Cardiff is a highly multicultural area, and a number of languages other than English and Welsh are spoken

(a consideration, too, for the whole of Wales, see Durham and Morris 2016: 5). During the Industrial Revolution, Cardiff became the main port for exporting coal from the surrounding valleys. Soon thereafter, the population of the city grew at a fast rate following mass migration from within and outside of Wales.

The Welsh language was prevalent in Cardiff until the Victorian era when the percentage of speakers fell to 10.7%, due to the influx of non-Welsh-speaking and English-born immigrants to the city (Parry and Williams 1999: 55). The first Cardiff-based Welsh-language school was established in the 1950s, and since then the language has steadily gained ground, in large part due to more children attending Welsh-medium education and the establishment of Cardiff as the capital city of Wales.

According to recent figures, of the total 56,943 pupils in Cardiff, 8,419 pupils (14.7%) are in Welsh-medium education in the local authority (StatsWales 2021). This shows that whilst Welsh-medium provision in Cardiff is growing (see Cardiff Council's Welsh in Education Strategic Plan 2022-31), pupils studying through the medium of Welsh are still a minority in the capital.

It should also be noted that figures from 2000 show that 15% of Welsh speakers from "heartland" areas moved to Cardiff (Hodges 2009: 29) as a result of the establishment of Welsh language media services and public administration. Housing the Senedd (Welsh Parliament) and the Welsh Government, Cardiff has been described as the administrative centre of the Welsh language (May 2000) and the use of the language in middle-class professional circles is growing (see Aitchison & Carter, 2004: 72).

According to PLASC data, 2,585 pupils in Cardiff secondary schools (ages 11-15) are fluent in Welsh, and 1,140 (44%) of those speak Welsh at home, whereas 1,445 (55%) do not speak Welsh at home. Given our understanding of the situation, many of those attending

Welsh-medium schools will be migrants from other areas of Wales. Indeed, the home language variety is likely to reflect the geographic origin of parents and grandparents.

4.2.2.1 *Ysgol y Ddinas*

Ysgol y Ddinas (pseudonymised for this study) is a Welsh-medium secondary school in Cardiff. In 2017, 36% of pupils came from homes where Welsh is spoken (Estyn 2017b)¹⁴; however, all pupils study Welsh as a first language and all courses at the 6th form are provided through the medium of Welsh. During its inspection in 2017, it was noted that an exceptional feature of the school was that many pupils speak Welsh “naturally socially, in extra-curricular activities and around the school” (Estyn 2017b: 5), due to school arrangements to strengthen and promote the use of Welsh, including a ‘Byw yn y Gymraeg’ [Live in Welsh] working group. The expectation is that ‘day-to-day’ language (in the corridors, classrooms, on the playground etc.) of Welsh-medium schools such as *Ysgol y Ddinas* is Welsh.

A total of eight pupils participated in the study from *Ysgol y Ddinas*. They were from across year 12, studying various A Level courses.

4.2.3 *Summary*

The number of Welsh speakers in Cardiff is lower than the national average, whereas the number for Gwynedd is higher. According to the 2021 census, 12.2% of people in Cardiff reported that they could speak Welsh. Of the self-reported Welsh speakers who partook in the Annual Population Survey (APS) (2023) 6.6% reported that they used the language on a daily basis¹⁵. This emphasises that in Cardiff there are far fewer opportunities to use the language in the community than in Gwynedd. The main considerations about the two geographic areas and their linguistic contrast (based on APS data in 2023) is shown in Table 4.1.

¹⁴ Reference anonymised in full bibliography.

¹⁵ Readers should note that there are fundamental differences between APS and Census data sources, including differences in how the data are collected.

Table 4.1: Linguistic contrast between Gwynedd and Cardiff

	Gwynedd	Cardiff	National
Proportion of population able to speak Welsh	79.9%	23.1%	29.2%
Proportion of Welsh-speakers who use Welsh on a daily basis	68%	6.6%	14.9%
Proportion of year 12 pupils taught in Welsh as a First Language ¹⁶	81.4%	19.0%	25.4%

4.3 Data collection and analysis

In this section I discuss the data collection process, including the recruitment of participants and ethical considerations. A major amendment was required to the original research design, (and subsequent ethics application) because of limitations surrounding the outbreak of COVID-19 in 2020. Original sampling frames were adjusted to reflect the availability of eligible participants and consenting institutions. Firstly, I outline the disruption experienced during COVID-19, then I present ethical considerations in working with young people, before outlining how I recruited participants, in light of the above.

4.3.1 COVID-19

Following the COVID-19 outbreak, schools were forced to close from March 2020, and remained closed, or at limited capacity following further lockdowns in Wales to tackle the spread of the virus. In December 2021, further disruption to the school year occurred, and all year groups were taught online until March 2021.¹⁷

Originally, the research design consisted of visiting with schools in the two sociolinguistically contrasting areas and recording participating students in two naturally-

¹⁶ Data showing pupils who follow a Welsh as a First Language pathway indicate numbers of fluent speakers in the language. This data is based on 2022/23 data (Welsh Government 2023a)

¹⁷ For a full breakdown of Wales' lockdown measures and their impact on schools, see Senedd Research (2022).

occurring settings during their typical school day (a peer-to-peer interaction over lunch, a classroom presentation), as well as a one-to-one sociolinguistic interview with the researcher.

Following closures of schools during the COVID-19 pandemic in 2020, contact between the researcher and schools diminished and it became increasingly difficult to arrange a time to gain access to the pupils to carry out the research. During the pandemic, news outlets reported on the job market becoming saturated (e.g. ‘*Recruiters inundated as coronavirus takes toll on UK labour market*’, in The Guardian at <https://www.theguardian.com/society/2020/jul/28/recruiters-inundated-coronavirus-toll-job-market-employment>), and the Welsh Audit Office (2019) found that schools are not offering enough employability skills to pupils (Welsh Audit Office 2019: 11). Estyn (2017c) had also been calling for stronger partnerships between schools and external stakeholders to improve the delivery of “impartial advice and guidance” (Estyn 2017c: 6), following cuts to Careers Wales funding since 2015. In order to provide some incentive for schools and pupils to take part in the project, whilst not infringing on the pupils’ free decision to participate (BERA 2018: 19), an alternative methodology was pitched to participating schools.

This involved delivering a module of Careers workshops on a range of topics including CV writing and interview preparation, using a range of Cardiff University outreach resources for schools and students (Cardiff University 2020). Participating pupils would then be contacted to take part in an online mock job interview with the researcher to practice using the skills they had acquired (i.e. using the STAR method¹⁸ to answer competency-based interview questions). Pupils would finally be offered a chance to reflect on their interview with the researcher, highlighting aspects of their performance which they considered to have gone well, and what they might do differently next time. Each of these contexts is covered in more detail in section 4.5.

¹⁸ e.g. National Careers Service (2020)

The participating institutions agreed to and helped co-construct the new proposed methodology in light of the restrictions with conducting remote research, and a revised ethics application was submitted to the University Ethics Board and was accepted in January 2021 (SREC reference: ENCAP/Young/23-07-2020).

4.3.2 Ethics

The present study involved interviewing and recording child (under 18) participants, therefore ethics was a serious consideration. Ethical approval was granted from the Ethics Committee at the School of English, Communication and Philosophy at Cardiff University (<https://www.cardiff.ac.uk/research/our-research-environment/integrity-and-ethics/research-ethics>). The confirmation of approval is attached in Appendix 1 (11.1)¹⁹.

Schools were approached by email (11.2), for the attention of heads of sixth form and heads of school. Gatekeeper information letters (11.3) were attached to the email containing full information about the study (as appropriate to the Gatekeeper's role), in order to ensure their full and informed permission to access the participants. The Gatekeeper was given two weeks to consider the request for permission, and to consult with others about it as necessary. Phone conversations, face-to-face discussions and email exchanges ensued to clarify the arrangements for data collection before they agreed to take part. Gatekeepers then signed a consent form agreeing that their institution would participate (11.4).

In line with the ethical guidelines set out by the University's Ethics Committee, before I recorded the participants, I first gained their informed consent. As the participants were under the age of 18, information sheets were provided to gatekeepers (heads of sixth form at each participating institution) and participants. Parental consent was not required under British Educational Research Guidelines (BERA 2018), but because participants were under the age

¹⁹ All participant-facing documents were written and supplied bilingually. Only the English has been included in this thesis.

of 18, parents were provided with information sheets, too. Participants were encouraged to discuss participation with their parents before agreeing to take part. The information sheets provided to parents and gatekeepers detailed the aims of the project – that it was an investigation into participants’ use of Welsh in a school context, and that they would be recorded in three different registers. Participants were informed that they were able to withdraw from the study at any time, during or after their participation.

The gatekeepers then recruited a class of 20 participants (to allow students to potentially drop out, or if any were not appropriate for the research design) who matched the home language requirements. Home language data was held by gatekeepers. Parent information letters (11.5) and student information letters (11.6) were distributed via gatekeepers. The participant consent form can be seen in 11.7. Due to social distancing measures in place, and restrictions on physical access to the schools at the time of the COVID-19 pandemic, all consent forms were received as scanned copies via email, and original documents were posted to the researcher’s office at Cardiff University where they were kept securely.

4.3.3 Participant recruitment

Recall that ten participants were recruited from Gwynedd, and eight from Cardiff. Participants were recruited in May and June 2021, and data collection occurred between June and July 2021. Data collection occurred during the school term, and pupils were required to negotiate the data collection with their other school commitments. This period was chosen following discussions with prospective participating institutions; it was deemed best to use the time after year 12 summer assessments, which is typically when sixth formers have more free time before their year 13 courses start. The data collection period consisted of two sessions for each participant with the researcher (the interview and conversation contexts), as well as six workshop sessions with their peers, which participants completed in their own time over the course of two to three weeks.

Participant recruitment was complex in the fast-moving and ever-changing context of the COVID-19 pandemic. A total of 25 prospective participants from across the 6th form at the Cardiff school requested information sheets after expressing an interest to their head of sixth form in participating. Consent forms were collected from 17 of those. Due to an outbreak of COVID-19 immediately before data collection was due to begin, all pupils in that year group were forced to self-isolate. A total of eight pupils from that institution took part in the study, as a result. The sample was therefore relatively small and unbalanced, as is discussed below.

4.4 The participants

Data were collected from 18 pupils in Welsh-medium education, aged 16-18 years old. Sixth form students were recruited as it is generally agreed that the process of language acquisition is largely complete by the age of 16. Moreover, the upper bound for vernacular reorganization is about 17 years old; which means that speakers of that age are thought to be at the forefront of language change in any community (Denis *et al.* 2019: 44).

All participants were considered Welsh-English bilinguals on the basis that they had attended Welsh-medium secondary education (age 11-16) prior to attending their sixth form, where classes and examinations were held in Welsh. Some participants opted to follow English-medium courses in sixth form, but all were still deemed bilinguals. Pupils at the participating sixth forms have attended Welsh-medium secondary education consistently, and have lived in their community, and attended their school (or a feeder school in their sixth form catchment area) for at least 5 years.

18 pupils in Welsh-medium education, aged 16-18 years old were recruited using judgment sampling (Milroy and Gordon 2003: 30). 10 pupils attended *Coleg y Fro* in Gwynedd, and eight attended *Ysgol y Ddinas* in Cardiff. Recall that the primary aim of this study is to examine the stylistic variation of speakers who come from two home language backgrounds; those who speak Welsh at home and those who do not. The following section

provides further information about the stratification of the sample by home language and following that presents the difficulties with stratifying the sample by gender, and thus the decision to exclude it as a factor in the analysis. A summary of the participant sample is then provided in section 4.4.2.

4.4.1 Home Language

Attempts were made to stratify the sample equally in terms of participants who spoke Welsh at home and those who do not speak Welsh at home. The purpose of this was to distinguish between pupils who have acquired Welsh from a primary caregiver and those who acquired Welsh through Welsh-medium education. In some cases, where pupils spoke English and Welsh at home, the questionnaire data (further information on the questionnaires is provided in section 4.5) and sociolinguistic interview provide clarity on the amount of Welsh spoken at home. Home language also indicates how much Welsh the participants are exposed to outside of the classroom context. Table 4.2 presents the breakdown of participants by home language and area.

Table 4.2: Total number of participants by home language and area

Coleg y Fro (Gwynedd) <i>n</i> = 10		Ysgol y Ddinas (Cardiff) <i>n</i> = 8		Total
EHL	WHL	EHL	WHL	
4	6	4	4	18

Table 4.2 presents four distinct speaker groups: Gwynedd Welsh home language (GWHL), Gwynedd English home language (GEHL), Cardiff Welsh home language (CWHL), and Cardiff English home language (CEHL) which will form the basis of the comparative methodology, discussed in section 4.7.4 below.

When determining the home language of participants, the following caveat should be noted: it is unlikely that either home language category fully represents the diverse and

complex linguistic background of each participant. Although a participant may describe their home as being primarily English-speaking, that it not to say that they are not exposed to Welsh from extended family members. For example, Lowri speaks English with her Welsh-speaking Mum and Dad, and she lives next door to her grandparents who always speak Welsh with her. Lowri's younger sister attends a Welsh-medium primary school and sometimes asks Lowri for help with homework, during which time they are likely to speak Welsh. However, Lowri has determined that her home language is English. Equally, a participant who claims Welsh as their primary home language is likely to come across some English in the home, too. The categories of home language are useful in order to create a control to establish whether those who mainly grew up using Welsh predominantly at school produce the same type of variation as those who grew up speaking Welsh at home as well as at school.

Where participants stated that they speak a combination of languages at home (e.g. 100% Welsh with Dad, 50% Welsh and 50% English with Mum) I calculated the *primary* home language of the speaker. That is, from the example above, 75%+ Welsh would mean the primary home language of the speaker was Welsh. No home languages other than English or Welsh were reported by participants and based on the primary home language framework presented here, participants were either deemed to be Welsh home language (WHL) or English home language (EHL).

Interestingly, some participants' reports of home language were inconsistent between their interviews and questionnaires. For example, Mathew and Naomi wrote on their questionnaires that they used a combination of English and Welsh at home. In their sociolinguistic interviews, however, both elaborated and stated that only English was spoken at home. It is possible that participants felt pressured to overstate their use of Welsh in the questionnaire; being self-report data, it could reflect how the participants were choosing to be seen by the researcher, rather than actual use. This could be particularly significant in the case

of a minoritized language like Welsh, and participants knew that this was a study looking at the use of the Welsh language (see participant information sheets for further detail). In cases such as these, the sociolinguistic interview data was used as a stronger indication of home language background than the questionnaire. The gender of participants as an independent variable is considered in the following section.

4.4.2 Self-reported gender

Much research exists on the differences between the sociolinguistic variation of male and female speakers. The ‘gender pattern’ identified in much of the existing sociolinguistic work highlights women using more prestigious and innovative patterns than men (e.g. Labov 1972; Trudgill 1974). Gender has not been found to be significant in recent work on Welsh variation in Cardiff (Gruffydd 2022) but in North Wales, gendered differences have been found in other phonetic work (Morris 2017; Morris and Hejrná 2017; Morris 2021). The current study had aimed to explore any correlations between speaker gender and sociolinguistic variation. Attempts were made to stratify the sample equally between female and male participants in order to examine whether gender is a significant factor in the acquisition of sociolinguistic competence. It should be noted that the categorisation of participant gender is based on self-report data in a questionnaire.

A summary of the sample can be seen in Table 4.3. A total of 13 participants were female, and 5 were male. Because of difficulties in recruiting participants during COVID-19 there was a gender imbalance in the dataset, with too few boys in each speaker group to observe any discernible patterns of variation according to gender. Due to the uneven distribution of gender across the various groups, I will not consider gender in the analysis of features.

Table 4.3: Summary of the sample

	Coleg y Fro (Gwynedd) $n = 10$	Ysgol y Ddinas (Cardiff) $n = 8$	Total
--	--	--	--------------

	EHL	WHL	EHL	WHL	
Male	1	1	2	1	5
Female	3	5	2	3	13
Total	4	6	4	4	18

Next, I present the questionnaires and the way in which the three registers (or speech contexts) were devised and defined for the purpose of eliciting different speech styles.

4.5 The questionnaire

A questionnaire (11.10) was provided to all participants at the beginning of the data collection period. Because the research was carried out remotely, paper copies were sent to participating schools. Gatekeepers administered the questionnaires and instructed participants on how to complete them. Completed questionnaires were returned to the researcher via post at the end of the data collection period. The questionnaire was designed to collect data on biographical information of each participant. That biographical information is discussed below.

Participants were asked about their home language use to ensure that they were placed in the right speaker group i.e. that Welsh home language pupils in Gwynedd were placed in the GWHL group (the rates of which are presented in Table 4.2). The sociogram (question 12 of the questionnaire) was intended to be analysed to determine participants' social use of Welsh (as in Ryan 2018). However, limited, inconsistent or unclear completion of this question (likely owing to the researcher not being present to administer the task) meant that this data could not be used in the final analysis. A question about national identity was asked in case this proved an explanatory variable for the study, but 94% identified as having a Welsh-only national identity²⁰, and thus it was not examined further.

²⁰ One participant identified as being both Welsh and British.

The participants also noted how often they used the language outside of school settings. The more frequent use of Welsh outside of school in Gwynedd could be explained by the prevalence of the language in the community, which is in keeping with previous descriptions of the daily use of Welsh (see section 4.2). This language use variable is related to the language used with parents and siblings in the home, and it was therefore not included in the analysis.

Participants were also asked about their parents' professions. There was a clear imbalance when comparing the type of parents' jobs in Gwynedd and Cardiff (see appendix 11.11). This imbalance and its possible link to socioeconomic background is a limitation of the sample and reflects the tendency among Welsh-speaking migrants in Cardiff to be upwardly mobile professionals and members of the "middle class revolution" (Aitchinson and Carter 1987: 492). Although it was not possible to control for socio-economic background in my analysis, this will need to be considered in the discussion of the findings.

The participants were asked where their parents were born as well as where they themselves were born. In the case of all of the Gwynedd participants, their parents came from within a 20-mile radius of the 6th form institution attended by the students. On the other hand, parents of the Cardiff participants hail from towns and cities up to 150 miles from Cardiff, including six parents who come from England. Although it was not possible to control for parents' mobility, there is evidence in my data of a high influx of migrants in Cardiff as has been found in other studies (Hodges 2009: 29).

Most noteworthy of all, where participants were asked in the questionnaire to report their L1, answers from Cardiff matched the language/languages participants reported to speak with their parents (e.g. if they spoke English with their parents, they noted that English was their L1). On the other hand, all pupils in Gwynedd considered Welsh to be their L1, even where they did not report speaking the language with their parents. This, again, may relate to

the prevalence of the language in the community, which provides further justification for examining the acquisition of sociolinguistic competence in relation to community use of the language.

4.6 The registers (speech contexts)

Students were recorded in a number of different situations in order to compare the stylistic variation occurring in their speech. Combining interview data with other ‘real’ speech data in naturally-occurring registers helps us to understand the way in which young Welsh speakers stylistically use morphosyntactic variables in a range of contexts, synonymous with understanding the extent of their sociolinguistic competence.

The current study, in line with previous work on acquiring sociolinguistic competence, examines intra-speaker variation by eliciting spoken data from informal and formal speech contexts. As stated previously, the primary aim of the methodology for this study was to simulate categories of speech situations which were as replicable and consistent as possible for all speakers, in a limited space of time. The reader will recall that these speech contexts were devised in light of limitations with accessing the school sites during COVID-19.

The actual registers post-COVID-19 included a peer-group context in which participants recorded semi-spontaneous conversations with friends in the peer-to-peer context (intended to elicit the most informal varieties), an interview context in which participants take part individually in a mock job interview (intended to elicit the most formal varieties), and a conversation context in which participants take part in a sociolinguistic interview (intended to elicit semi-informal varieties). Although to an extent these registers were simulated, rather than being naturally-occurring, categorical distinctions based on the intended formality were still made between each of the contexts in order to discern to what extent there is variation in the participants’ stylistic repertoires. The contents of each speech context will be discussed in further detail in the subsequent sub-sections.

4.6.1 The peer-to-peer context

A module on the topic of careers was devised using a combination of Cardiff University outreach materials (2020) and other available resources online. The module was aimed at sixth form students applying to University and those looking for work straight from sixth form.²¹

The workshops in Table 4.4 were designed and pre-recorded using Zoom. Links to each pre-recorded workshop and instructions were sent to all participating students. Students completed each workshop in pairs or groups of three. My pre-recorded workshops featuring careers-related content were around 20-30 minutes long, and each included a discussion session, where the pupils were instructed to pause my presentation and talk around semi-structured topics to elicit peer-to-peer speech. The discussion sessions in workshops 1-5 were largely intended to be informal. The discussion session in Workshop 6 was designed as a role play in which participating students took turns to ask and answer interview questions.

²¹ I am a fully qualified secondary school teacher, and thus the design of the workshops was based on my experience and training in that job.

Table 4.4: Careers module design

	Title	Aims	Resources	Discussion Session
1	Getting to know ourselves and each other	Introduce the researcher; the personality test and how the result can help us find a career suited to us.	16personalities.com	Participants introduce themselves, what they like about their courses, their ideal job as children, their part time work etc.
2	University experience	The transition between school and university; clubs and societies; student life	Cardiff University outreach: https://prezi.com/view/FTFTZdHA3z0usQU9XIwp/ https://prezi.com/view/Lbt9JIHx9orctzGzjCMs/	Participants discuss various societies at Cardiff University that might coincide with their own hobbies, what clubs they would join/start.
3	Transferable skills	Identifying transferable skills; how can we improve our skills	Online courses (e.g. FutureLearn and Open University)	Participants discuss their skills and give examples of when they've used these skills.
4	The CV	Understanding the purpose of a CV; designing a CV; comparing two CVs	Researcher's own	Participants discuss the strengths and weaknesses of two mock CVs and how they might do things differently.
5	Preparing for the mock interview 1	How to behave in an interview	Cardiff University outreach: https://prezi.com/view/9RbIxt4rSFAopLv637qf/	Participants describe a number of images depicting interview settings and list what might be going well or badly in them
6	Preparing for the mock interview 2	Learning the STAR technique; examples of interview questions	Cardiff University outreach: https://prezi.com/view/9RbIxt4rSFAopLv637qf/	Participants practice answering interview questions using the STAR technique

Pupils were instructed on how to record and save each 'Discussion Session' on their mobile phones. Recordings for each group discussion were collected through Cardiff University's FastFile service, where pupils followed a secure link to attach their phone recordings. Figure 4.3 shows an example of a 'Discussion Session' from Workshop 2 entitled 'University Life'.

Undebau a chymdeithasau

1

Pwyswch 'record' ar eich ffôn symudol

2

Ar ôl 20 munud (max) pwyswch stop

3

Cadwch y ffeil dan yr enw 'Gweithdy 2'

- Oes rhai o'r cymdeithasau yma yn cyd-fynd â'ch hobiau chi?
- Pe taech chi'n dod i brifysgol Caerdydd – pa gymdeithas fydddech chi'n ymuno? Pam?
- Oes yna gymdeithasau yma sy'n eich synnu chi?
- Pe taech chi am ddechrau cymdeithas newydd, beth fyddai hi?

Unions and societies

1

Press 'record' on your mobile phone

2

After 20 mins (max) press stop

3

Save the file as 'Workshop 2'

- Do any of these societies fit with your hobbies?
- If you were to come to Cardiff University – which society would you join? Why?
- Are any of these societies surprising to you?
- If you were to start a new society, what would it be?

Figure 4.3: Screenshot of prompt slide for Discussion Session in Workshop 2: University Life (also translated into English)

The nature of the workshops allowed for discussions to be kept face-to-face in the school building in a quiet common area, or if pupils were working from home or self-isolating during the course of the module (which many were) they could complete the workshops online using a platform such as Zoom or Teams (whichever their institution gave them access to) to record their discussions. Each ‘Discussion Session’ lasted approximately 10-15 minutes, providing around 5 minutes of talk per speaker for each of the 6 workshops. The peer-to-peer context is

comprised of approximately 10 hours, making up around 27% of the entire corpus. On average, 31 minutes of peer-to-peer conversation data was collected from each participant.

As can be seen from the excerpts below, the discussion sessions in the peer-to-peer register succeeded in prompting natural speech which often veered off the loosely defined topics provided, and led to some pupils sharing tangential anecdotes, indicative of naturally-occurring speech. In the below example, participants discuss whether they would join a political society at university:

Heledd: ti'n gwybod baswn i'n methu *joinio Plaid Cymru achos dw i ddim yn gwybod dim byd am –

Scarlett: na na fi chwaith.

Catrin: oedd fel - wnaethoch chi wneud y - oeddech chi'n gorfod fel cofrestru i *fotio - wnaethoch chi wneud?

Scarlett: ie

Heledd: do. Wnes i gofrestru ond wnes i ddim *pleidio (though)

Catrin: na na finnau chwaith

Heledd: Wnes i anghofio dweud gwir

[Heledd: you know I couldn't join Plaid Cymru because I don't know anything about -

Scarlett: no me neither

Catrin: was it like – did you do the – did you have to like register to vote – did you do it?

Scarlett: yeah

Heledd: yes. I registered but I didn't vote though

Catrin: no me neither

Heledd: I forgot to be honest.

Gwynedd participants' peer to peer conversation

Some participants also recounted long narratives in the peer-to-peer context, containing jokes and funny anecdotes, such as in the example below:

Deio: erm wel mae gen i syniad - 'sa i'n siŵr os mae hwn yn barod yn bodoli ond –

Caryl: ww?

Deio: achos fi'n mynd i drama yn <anonymous> ar y wal wastad maen nhw gyda *little advertisement* am yr hummus *society* a fi yn caru hummus

Caryl: wow *oh my gosh*

Deio: hummus yw fel bywyd fi - *caramelised onion* yn anhygoel baswn i'n dechrau hummus *society* erm chi fel yn rhannu hummus chi fel yn siarad am hummus chi'n cael hummus *events* fel - err *deifio mewn i'r hummus a fel *pin the hummus on the donkey* jest lots o hummus

Caryl: fyddech chi'n creu hummus?

Deio: Efallai - ie wnawn ni greu hummus *chickpeas* bach o *caramelised onions*

Caryl: rhaid dweud fi ddim yn *biggest fan* o *caramelised onion* hummus fi'n *really boring*

ond fi'n – fi'n hoffi *plain*

Deio: Oh fi yn hoffi *plain* hefyd

Caryl: pan ro'n i'n - pan ro'n i'n llai oedd fi ddim yn hoffi hummus a wnaeth Mam cael yr *like moroccan* hummus achos roedd hi'n meddwl byddaf fi'n licio fe a roedden nhw wedi rhoi i fi gormod a fi *massively* jest wedi mynd *off* hummus fel fi'n *gradually* cael fy hun yn ôl mewn a *plain is the most I can do*

[Deio: erm well I have an idea – I'm not sure if this exists already but -

Caryl: ooh?

Deio: because I go to drama at <anonymous> on the wall they always have a little advertisement for the hummus society and I love hummus

Caryl: wow oh my gosh

Deio: hummus is like my life – caramelised onion is amazing I would start a hummus society erm so you can share your hummus like talk about hummus you can have hummus events like – err diving into the hummus and like pin the hummus on the donkey just lots of hummus

Caryl: would you make hummus?

Deio: maybe – yeah we'll make hummus chickpeas a bit of caramelised onion

Caryl: I have to say I'm not the biggest fan of caramelised onion hummus I'm really boring but I – I like plain

Deio: oh I like plain too

Caryl: when I was – when I was younger I didn't like hummus and Mam got the like Moroccan hummus because she thought I'd like it and they gave me too much and I massively just went off hummus like I'm gradually getting myself in and plain is the most I can do]

Cardiff participants' peer to peer conversation

As mentioned in section chapter 2 a key consideration of sociolinguistic methodology is the observer's paradox. Labov (1972: 209) states that "the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain these data by systematic observation". The examples above from the peer-to-peer context suggest that participants are engaging in casual speech, despite being systematically observed through their self-recordings. Their use of anecdotes, jokes and narratives indicates that the effect of the observer's paradox has been mitigated, thus minimising self-editing and attention to speech.

4.6.2 The mock interview context

Following workshops 5 and 6 of the careers module, participants took part in a mock job interview for an imagined tutoring position at a Tuition Centre in their local area. The aim of

the mock interview was to practice the skills they had learnt in a setting that simulates the largely formal setting of an interview.

The purpose of the mock interview was to elicit speech in the most formal speech register, thus eliciting more careful speech from the participants. Some questions were identical to those they had prepared previously in workshop 6 (see Table 4.4), but most were questions that they had not prepared. The mock interview lasted approximately 30 minutes. Mock interviews were held on Zoom because access to schools during COVID-19 was restricted. The audio of interviews was recorded using the Zoom record function. The video was not recorded as pupils only consented to their speech data being processed. Because this study focusses on morphosyntactic variation, the quality of the recording on Zoom was deemed sufficient.

The interview was the first live meeting between participants and the researcher (after having watched 6 pre-recorded workshops). This was intended to create a particularly formal setting, as it was thought that participants might pay more attention to their speech with a stranger. Before starting the recording, the researcher read a scripted preamble covering the following points:

- a) Reiterating the purpose of the session (to role-play a job interview),
- b) Reminding participants that the interview was being recorded for research purposes,
- c) Providing general instructions in case connection during the call was lost,
- d) Inviting participants to ask any questions before the recording started.

The tone of the preamble to the mock interview was formal and distant in order to simulate a professional and formal setting. In order to further increase the formality of the interview, I used the formal 2nd person address *chi* instead of *ti*.

The mock interview questions themselves aimed to elicit lengthy responses to maximise the collection of tokens. The mock interview questions (translated here into English) included:

You've applied to be a Maths or English tutor. Which subject would you prefer to tutor? Why?

What is your previous experience in the field?

What achievement has made you proudest of yourself?

Tell me about the last book you read.

Can you tell me about a time when you worked well with others?

What would you like to achieve by the end of your first week at the centre?

(The full interview guide can be found in 11.8)

Overall, over eight hours of mock interview data were collected, comprising 23% of the entire corpus. On average, 27 minutes of mock interview data was recorded for each speaker.

4.6.3 The sociolinguistic interview

The final stage of the research design was the sociolinguistic interview. This was intended to produce semi-informal speech data (more informal than the mock job interview, but less informal than the peer-to-peer context). The sociolinguistic interview was conducted in the form of a feedback session a week after the mock interview. Participants were invited to reflect on what went well and what they would improve on in their next job or university interview. Following 10-15 minutes of this type of feedback, a sociolinguistic interview was held for 30-45 minutes.

Having conducted the mock interview a week prior to the sociolinguistic interview, I was acutely aware that I could still be perceived as an authority figure in the sociolinguistic interview, therefore I made an effort to establish common ground with participants by:

- (a) Using the informal 2nd person pronoun address *ti*,
- (b) Emphasising my own northern variety of Welsh with the north Wales cohort, thus establishing myself as an insider,
- (c) Emphasising my intimate knowledge of Cardiff with the south Wales cohort, thus establishing myself as an insider,
- (d) Making jokes and adopting a friendly and open manner,
- (e) Showing an interest in the experiences and hobbies of the participants.

The sociolinguistic interview was semi-structured, allowing participants to partially guide the direction of the conversation. Pre-determined questions based on modules adapted from

Gruffydd's (2022) study of Cardiff Welsh were used if the conversation on one topic was running dry and provided a launching pad for more narrative. The modules were titled as follows:

- Family
- Community
- Culture
- School
- Friends
- Fear
- Hobbies
- Future
- Work
- Language

The full schedule of questions in the sociolinguistic interview (translated into English) can be found in appendix 11.9. Due to the age of the participants, and the fact that the topic could be sensitive during the global pandemic, the trigger question (traditionally relating to 'danger of death') in the sociolinguistic interview was changed for the 'fear' module in which participants were asked to describe a horror film they've seen or to talk about a phobia they have. One topic which invariably came up during the sociolinguistic interview (as well as the peer-to-peer context and the mock interview) was how life had changed for the pupils during and since COVID-19. Many participants shared lengthy narratives about their lives during the pandemic.

Overall, more than 17 hours of speech data were collected. On average, sociolinguistic interviews lasted 57 minutes per speaker. This made up 49% of the speech corpus.

4.7 Data analysis

4.7.1 Transcription

A relatively large quantity of speech data was collected, approximately 34 hours in total. This comes out to just over a quarter of a million words when transcribed. The number of words per

hour varied across the three different speech contexts and also across the 18 speakers. Transcription took place on completion of data collection, and was carried out by the researcher, using a transcription protocol based on Morris (2014), which focuses on transcribing Welsh speech data. This provided useful advice on orthographic representations of standard and non-standard or dialectal use of Welsh (e.g. spelling the common colloquial ‘isio’ [want] using the standard orthography ‘eisiau’). Morris’ protocol also contains information on best practice on transcribing code switching (e.g. ‘Dw i (definitely) eisiau mynd’ [I definitely want to go]).

Only the speech of participating students was transcribed: any audible voices in the background were ignored. All transcriptions were anonymised (including names of schools, friends, family, pets) to protect the identity of participants, even at a local level.

The data were transcribed using the annotation tool ELAN between August 2021 and March 2022. All spoken data was transcribed using ELAN software. Where multiple speakers were present, each speaker was given their own tier in ELAN, and a separate text file was created for each speaker containing only their own transcribed speech data. This was not the case for the recordings where the researcher was interviewing, as the speech data of the researcher will not be considered in this research. Each speaker had data transcribed from the peer to peer, the sociolinguistic interview and the job interview contexts.

4.7.2 The selection of variables

The dependent variables chosen for this study are the linguistic features which are thought to vary between registers in Welsh. These features were selected on the basis that there was relatively strong evidence in the literature about expected variation, and where that evidence base was weak for Welsh, literature about languages other than Welsh was consulted. The three features analysed in this work are:

1. Possessive pronouns

2. Inflected and periphrastic simple past verbs
3. Intensifiers

Recent, previous variationist literature on the variable nature of possessive pronouns in Welsh (Davies 2016) was a strong basis for analysing that feature. B. M. Jones (1993) argues that inflected and periphrastic verbs also vary according to context, and further evidence from studies of French (see 6.2.3) supported this hypothesis. Variation in intensifier use was largely based on English work on adolescent speech (Tagliamonte 2016), as there is little reported evidence on the variable use of intensifiers in young people's speech in Welsh. A full literature review for each feature, providing a justification for its analysis, is given in each of the next three chapters.

Because the use of these three features is largely unreported in variationist studies of Welsh, this will be the first work of its kind. Consequently, little is known about the factors predicting variation in young people's speech, making the current work largely exploratory in nature. The proposed predicting factors may not be predicting factors for Welsh at all. It is therefore vital to first present the WHL patterns before examining whether they have been acquired by EHL speakers. The comparative method used to carry this out is presented below.

4.7.3 Comparative sociolinguistics

As stated thus far, the purpose of the current thesis is to determine not only the patterns of variation in the speech of Welsh-medium pupils, but also the extent to which those patterns are shared between Welsh home language and English home language pupils and to what extent patterns are replicated in two distinct areas, where Welsh is used in the community to varying extents. In order to establish this, a comparative sociolinguistic methodology will be used.

The comparative sociolinguistic methodology “enables the many different influences on linguistic features to be disentangled through systematic examination of their behaviour, and [...] situates and explains the linguistic features through comparison with like features in

related varieties” (Tagliamonte 2013: 729). The crucial pieces of evidence used for this purpose are commonly known as the “three lines of evidence,” which are derived from the results of *Goldvarb*’s step-up/step-down variable rule analysis method (see Tagliamonte and Poplack 2001: 92).²² The three lines are as follows:

- (i) statistical significance;
- (ii) relative strength;
- (iii) constraint hierarchy

When patterns of language use are looked at in this way, similarities and differences in the significance, strength, and ordering of constraints provide a microscopic view of the grammar from which we can infer the structure (and possible interaction) of different varieties of language (Tagliamonte 2013: 731). Studying variation guided by these questions helps us to examine the underlying patterns of a feature, in order to uncover similarities and differences beyond the level of general distributions. Where two varieties present the same conditioning and direction of effects, they are said to share the same source variety (Durham 2014: 60).

The comparative sociolinguistic methodology has been used in previous studies on the acquisition of sociolinguistic competence in an English as a lingua franca context (Durham 2014), in English varieties of the African American diaspora (Poplack and Tagliamonte 2001), and in the acquisition of English norms in migrant adolescents (Meyerhoff and Schlee 2012: 409). Meyerhoff and Schlee frame the lines of evidence in slightly different terms, by adding a fourth ((d), below) dimension; they argue that in order to acquire sociolinguistic patterns of variation, non-native speakers must identify and match native speakers’ (a) relative frequencies of key variants, (b) the linguistic and non-linguistic constraints on the variants in native speakers’ speech, (c) the constraint ordering in those factor groups, and (d) stances, acts, activities and styles indexing social categories, such as gender. The current work draws on all

²² The field of variationist sociolinguistics has since adopted mixed-effects modelling over *Goldvarb*.

of this previous work to establish a statistical framework from which to examine variation. This is presented in the statistical analysis section below.

Generational change can also be tracked through comparative sociolinguistics. Take, for example, Tagliamonte and D’Arcy’s (2007) study. They compared rates of quotative *be like* in several generations within the same speech community, and were able to show, through subtle shifts in the constraint ranking and relative strength of factors, that *be like* was doing more than growing in prevalence as a new lexical item. Indeed, there was evidence from this data that the quotative system was being reorganised by younger speakers (Tagliamonte and D’Arcy 2007: 210).

Never before (to my knowledge) has a comparative methodology been employed in the study of Welsh, let alone context of varieties of Welsh as spoken by pupils from different home language backgrounds. This methodology will allow me to examine the constraints and relative strength of factors of four speaker groups, in order to ascertain whether WHL pupils and EHL pupils’ repertoires constitute different varieties of Welsh in two distinct areas. If patterns and ranking of constraints is shared between the two groups, we may be able to point to a shared underlying variety being used in Welsh-medium schools. If patterns and ranking of constraints is not shared between WHL and EHL pupils, this may point to a limited acquisition of sociolinguistic competence among EHL pupils. The methodology will also enable us to assess whether a particular area of Wales, or speakers from a particular home language background might be leading the process of language change.

The comparative analytical method outlined here follows previous work by Poplack (e.g. 2000) and Tagliamonte (e.g. 2006) and later Schleeef and others (Schleeef *et al.* 2011; Schleeef 2013). Three lines of evidence will be presented in each analysis chapter in order to show not only whether each factor is a significant constraint on variation, but also the size of the effect and the hierarchies within each factor (Gardner 2023). Much of the previous work using this

type of framework has studied variation in majority languages, for example migrants acquiring well-known variation patterns in the variables ‘-ing’ and ‘t’ in English (Schleef 2013). The predicting factors on variation for variables such as these have been widely researched, and thus authors such as Schleef are able to demonstrate clear and significant patterns in ‘native’ speech, and thus show with certainty whether ‘non-native’ or ‘learner’ speakers have acquired these patterns. Patterns of variation in languages such as Welsh, however, are under-researched, thus adding a layer of complexity to the current work. The reader should remember throughout this thesis, that I am discussing factors which have not all been found to be significant predictors of variation. Where EHL patterns are discussed in relation to WHL patterns, I present the following caveat: although patterns are reported based on 3 robust lines of evidence, this is introductory, exploratory work and some conjecture about possible patterns in WHL speech prevails.

The statistical analysis which underpins the comparative method is laid out below.

4.7.4 Statistical analysis

Previously in comparative sociolinguistics, a multivariate analysis would have been conducted using the variable rule program (i.e. Goldvarb or Varbrul) (for a full list of programs, see Tagliamonte and Baayen 2012: 136), which allowed authors to obtain the variable rules of a given feature (Durham 2014: 61). In this type of analysis, significance and factor weights would be used as three lines of evidence. As Gardner (2023) states:

the first line of evidence is determined by what factor groups (predictors) are selected in the best step-up/step-down run, the second line of evidence is determined by the range of the factor weights in each of the significant factor groups (predictor), and the third is determined by the ordering of factors (levels) within each factor group (predictor) based on each factor’s (level’s) factor weight.

More recent developments in the field have led to the use of mixed effects models. The new mixed-effects models provide the researcher with a more powerful and principled way of looking at predictors (Tagliamonte and Baayen 2012: 143), which includes treating individual

variation as a random effect, rather than a fixed effect, which makes it possible to generalise beyond the sample and to a wider population. Gardner (2023) proposes that creating a mixed-effects logistic regression model is a good way to perform multivariate statistical analysis in variationist sociolinguistic research. The following paragraphs outline how the comparative statistical methods used provide the three lines of evidence required to determine a) the patterns in WHL speech and b) whether those patterns have been acquired by EHL pupils, in two distinct areas. This method is used in each analysis chapter.

Recall that Meyerhoff & Schlee (2012: 409) argue that the first stage of uncovering patterns of use is to display the relative frequencies of key variants. In order to do this, I provide descriptive statistical methods to identify, describe and visualise patterns of variability in the collected data. Following this, statistical modelling is carried out using *lme4* (Bates Douglas *et al.* 2015) and *lmerTest* (Kuznetsova Alexandra and Christensen 2017) packages for R (R Core Team 2020). The mixed effects model in each ensuing analysis chapter contains all external and internal factors thought to influence the patterns of use for the feature, based on the literature (see the following section). All significant factors in the model (including interactions for all external factors) provide the first line of evidence in the analysis; significant factors in the mixed effects models point to effective predictors on the use of the variants under study. A parsimonious model is created using only significant predictors from the first overall model. If a parameter is found to be significant in that first model, it is interpreted as adding explanatory value to the patterning of the data. If a parameter does not emerge as significant in the model, it does not meaningfully contribute to the understanding of the data and can be set aside. Following the more parsimonious mixed effects model, I use a Wald X^2 (chi square) test, which iteratively adds and removes each predictor and compares how well each iteration fits the distribution of the data, following Gardner (2023). This second stage provides a chi square statistic which can be used to understand the size of the effect, which is the second line of

evidence. The third line of evidence (the hierarchy of factor levels) was then displayed in the form of log likelihood figures, also following Gardner (2023). This process is then repeated for all subset speaker groups, in order to show where patterns converge and diverge for WHL and EHL speakers in both areas.

Where patterns converge between groups, the variation (and thus sociolinguistic competence) will be deemed to have been acquired. Where patterns diverge between groups, I provide an explanation for why groups may be patterning differently. This will be the basis of the discussion (see chapter 8).

4.8 Summary

Chapter 4 has described the methodology employed to carry out the current research, providing an outline of the main decisions relating to the sample design, data collection and data analysis. The speech of 18 speakers is analysed in this thesis, stratified by geographical area and home language. Data were elicited in three contrasting speech contexts, in order to examine the patterns of stylistic variation in Welsh immersion schools. A comparative sociolinguistic method was then applied to the data, in order to ascertain whether patterns observed in WHL speech have been acquired by EHL speakers, to assess the extent to which Welsh-medium education enables the acquisition of sociolinguistic competence. The chapters which follow present the findings from the analyses of the data collected using these methods.

5 Possessive pronouns²³

5.1 Introduction

This chapter presents an analysis of variation in the production of the possessive pronoun feature in Welsh. The possessive in Welsh has three variants; the first variant uses a pre-nominal pronoun with the noun, the second uses a post-nominal pronoun, and the third uses both pre-nominal and post-nominal pronouns. Following Davies (2016), these variants are named the literary variant, the sandwich variant, and the colloquial variant respectively. Table 5.1 presents an example of each variant in context.

Table 5.1: Possessive pronoun variants summary

Variant	Pre-nominal pronoun	noun	Post-nominal pronoun	English gloss
Literary	<i>fy</i>	<i>mam</i>		[my mum]
Sandwich	<i>fy</i>	<i>mam</i>	<i>i</i>	[my mum]
Colloquial		<i>mam</i>	<i>fi</i>	[my mum]

The variants are associated with different registers. The literary variant is expected in more formal varieties of language. The sandwich variant is often taught in second language materials but is also a bridging variant between the more formal literary variant and the colloquial variant. The status of the colloquial variant is less certain; it has been viewed as non-standard, but has been found to be used increasingly frequently among younger speakers of Welsh, and new speakers of Welsh, outside of traditional heartland areas (M. C. Jones 1998: 74). Other research has also found that the feature varies according to age (Thomas 1988b; Davies 2016), class (Thomas 1988b), and region (King 2016: 93).

²³ A modified version of this chapter appears in the journal *Language Variation and Change* (Young *et al.* 2024).

This chapter will measure the effect of different linguistic and social factors on the use of possessive variants. Despite some previous work (Thomas 1988b; M. C. Jones 1998; Davies 2016), there has been little research which examines the production of the colloquial variant and no work on young peoples' variation according to sociolinguistic factors such as home language and context of use.

The following section will introduce previous research on possessive constructions in Welsh. The chapter will then provide the methodology; outlining token extraction, the coding procedure and overview of the statistical analysis undertaken. The results will then follow which include both distributional and statistical analyses. Lastly, I will present a discussion of the findings.

5.2 The feature

Previous discussions have used different terms to describe the syntactic elements of the possessive construction, including the pronouns themselves (called “adjectives” by King 2016: 93), their placement (as “preposed clitics” (Davies 2016: 40), “echo pronouns” (Awbery 1994), and “postnominal pronouns” (Borsley *et al.* 2007: 158 etc), and the noun being possessed (Davies (2016) uses ‘possessum’ interchangeably with ‘noun’). For the sake of clarity, the current work will use the terms ‘pre-nominal pronoun’, ‘post-nominal pronoun’ and ‘noun’ to represent these three syntactic elements.

5.2.1 The literary variant

The literary variant is formed by attaching a clitic to the front of the noun phrase, e.g. *fy mam* [my mum]. Table 5.2 shows the prefixed pronoun for each grammatical person and provides an example of it in use as a possessive in a noun phrase. As can be seen in the ‘Noun’ column

in the table, nouns following genitive prefixed pronouns are subject to aspirate²⁴, nasal²⁵ and soft²⁶ mutation constraints.

Table 5.2: Prefixed pronouns for each grammatical person

Grammatical person	Mutation constraint	Pre-nominal pronoun	Noun	English gloss
1st person singular	Nasal	<i>fy</i>	<i>nghar</i>	my car
2nd person singular	Soft	<i>dy</i>	<i>gar</i>	your car
3rd person singular masculine	Soft	<i>ei</i>	<i>gar</i>	his car
3rd person singular feminine	Aspirate	<i>ei</i>	<i>char</i>	her car
1st person plural	h-prothesis before vowels	<i>ein</i>	<i>hafal</i>	our apple
2nd person plural	None	<i>eich</i>	<i>car</i>	your car
3rd person plural	h-prothesis before vowels	<i>eu</i>	<i>hafal</i>	their apple

Infixed pronouns can be found in written texts (Borsley *et al.* 2007: 157-158) where previous words end in a vowel e.g. *gyda* [with], *i* [to], and *a* [and]. Table 5.3 shows how these infixed pronouns are formed. In the case of 1st person plural, 2nd person plural, and 3rd person plural, the enclitics (called this because they are enclitic to the previous word) are contractions of the fuller preposed clitics, seen in Table 5.2.

²⁴ Where these preceding consonants change: p > ph, t > th, c > ch

²⁵ Where these preceding consonants change: p > mh, t > nh, c > ngh, b > m, d > n, g > ng

²⁶ Where these preceding consonants change: p > b, t > d, c > g, b > f, d > dd, g > Ø, ll > l, m > f, rh > r

Table 5.3: Infixed pronouns for each grammatical person

Grammatical person	Infixed pronoun	English gloss
1st person singular	‘m	my
2nd person singular	‘th	your
3rd person singular	‘i/’w	his/her
1st person plural	‘n	our
2nd person plural	‘ch	your
3rd person plural	‘u	their

Borsley *et al.* (2007: 158) argue that infixed forms for 1st and 2nd person singular are not common in colloquial speech. King (2016: 94), however, notes that some exclusively colloquial genitive infixed pronouns exist, which represent the pronunciation of the clitics more closely. An example of this is in the first person, where **fy siop** [my shop] might auditorily resemble ‘(y)**n siop**.

5.2.2 The sandwich variant

Another way of forming the possessive construction is by using the clitic prefixed pronoun (as discussed in 5.2.1) with a dependent personal pronoun following the possessed noun (Davies 2016: 40). Because the noun is placed between the pre-nominal pronoun and the post-nominal pronoun, Davies (2016 :41) refers to this variant as the sandwich construction. Table 5.4 shows the prenominal pronoun and their corresponding postnominal pronoun in all grammatical person constructions.

Table 5.4: Pre-nominal pronouns and their corresponding post-nominal pronouns

Grammatical person	Mutation constraint	Pre-nominal pronoun	Noun	Post-nominal pronoun	English gloss
1st person singular	Nasal	<i>fy</i>	<i>nghar</i>	<i>i</i>	my car
2nd person singular	Soft	<i>dy</i>	<i>gar</i>	<i>di</i>	your car
3rd person singular masculine	Soft	<i>ei</i>	<i>gar</i>	<i>(f)o/(f)e</i>	his car
3rd person feminine	Aspirate	<i>ei</i>	<i>char</i>	<i>hi</i>	her car
1st person plural	h-prothesis before vowels	<i>ein</i>	<i>hafal</i>	<i>ni</i>	our apple
2nd person plural	None	<i>eich</i>	<i>car</i>	<i>chi</i>	your car
3rd person plural	h-prothesis before vowels	<i>eu</i>	<i>hafal</i>	<i>nhw</i>	their apple

The same mutation constraints apply to the noun following the prefixed pronoun, as can be seen in Table 5.4. Variation exists in the type of echo pronoun used in the 3rd person singular masculine, depending on regional patterns; *o* and *fo* [his] are more common in north Wales, and *fe* and *e* [his] are more common in south Wales.

The practice of ‘echoing’ the pronoun of the possessor after the noun may have emerged as a way of eliminating the phonological ambiguity surrounding *ei* [his or her] and *eu* [their] (King 2016: 94). This construction can also serve an emphatic purpose, with the echo pronoun emphasising the possessor²⁷.

²⁷ The emphatic construction can be formed with the echo post-nominal pronouns presented in Table 5.4, or with contrastive emphatic pronouns (King 2016: 112) e.g. ‘*fy nghar innau*’

5.2.3 The colloquial variant

The third way of constructing the possessive in Welsh is by using a suffixed pronoun (identical to the echo pronoun in all but the 1st and 2nd person singular) following the noun. Table 5.5 presents the construction of this possessive variant in each grammatical person.

Table 5.5: Suffixed pronouns for each grammatical person

Grammatical person	Noun	Suffixed pronoun	English gloss
1st person singular	Car	fi	my car
2nd person singular	Car	ti/chdi	your car
3rd person singular masculine	Car	(f)o/(f)e	his car
3rd person feminine	Car	hi	her car
1st person plural	Car	ni	our car
2nd person plural	Car	chi	your car
3rd person plural	Car	nhw	their car

This variant has been found to be used increasingly frequently among younger speakers of Welsh (M. C. Jones 1998). Indeed, descriptive work often claims that this construction is not grammatical; Awbery (1976: 16) argues that it is not permissible for the possessive pronoun to follow the head noun e.g. *ci e* [(the) dog (of) him] as is the case where the possessive is not a pronoun but a noun e.g. *ci plentyn* [(a) dog (of a) child']. However, although it has been found to be used increasingly frequently in newer literature (e.g. Davies 2016), the variant is so heavily stigmatised that some grammar books still make no reference to it. King's (2016:93–94) *Modern Welsh*, which is considered a comprehensive reference to colloquial and literary Welsh grammar, notes that “while it is common ... [it] is widely regarded as sub-standard.”. Borsley *et al.* (2007: 159) also state that though this construction is possible, it is non-standard.

Previous authors have considered the development of the suffixed pronoun to be an extension of nonpronominal noun phrases (e.g. car Megan [Megan's car]) where the possessor follows the noun (Borsley *et al.* 2007). The literature on the variable use of this construction will be discussed in section 5.3. This is relevant because, as will be discussed further below, the colloquial form has at some points been ascribed to language acquisition processes (in children, but particularly in non-L1 speakers of Welsh). Indeed, D. Willis (2016) analysed various Welsh translations of Uncle Tom's Cabin between 1852 and 1854 and found that the colloquial variant was used as a representation of second-language, non-native Welsh speech.

5.3 Previous work

Having presented the feature, we now turn to examine previous work on the stylistic, social and linguistic factors which influence the production of the possessive in Welsh.

5.3.1 Social factors

5.3.1.1 Style differences

Previous sections (e.g. see 3.1.2) in the current thesis have explained how spoken and written registers of Welsh are different from one another. Literary Welsh has been reported to contain mostly possessive constructions using the pre-nominal pronoun (Borsley *et al.* 2007: 158), however Watkins remarked upon the “conscious effort” (1977: 153) of some authors to include the pre-nominal as well as the post-nominal (i.e. the sandwich form) in the modern literary language. The colloquial construction, using the non-standard post-nominal pronoun only, is reported to occur more often in speech than in written registers, however, informal written registers haven't been studied. Variation in the use of the three variants exists within the spoken register (the only register considered in this research).

The extent to which there is variation in use across styles has not been studied quantitatively thus far. However, Young (2019) found that Welsh teachers reported significant

variation in their own use of the possessive pronoun feature between in-classroom and out-of-classroom contexts. The participating teachers also showed a significantly increased likelihood to correct their students' use of the colloquial possessive variant as the context of use became more formal. This study demonstrated teacher perceptions of their students' use of the variant in different contexts, but no previous work has recorded the actual use of the variant in different registers.

5.3.1.2 *Home language differences*

The stigma attached to the non-standard colloquial variant is also evident in second language teaching. Not only does the colloquial variant not appear in some modern grammar books (see Thomas 1996) or is considered “sub-standard” (see King 2016: 94), but learning materials for adult learners (see Learn Welsh 2023) tend to exclusively teach the sandwich variant. Robert (2009: 104), however, identifies the colloquial variant as being a potential indicator of new speaker speech, as well as frequently occurring in the speech of young L1 speakers of Welsh. No work has analysed the use of the possessive pronoun according to the home language background of students in Welsh-medium education, and the current work seeks to understand whether new speaker patterns found in previous research are replicated by pupils from Welsh and non-Welsh home language backgrounds.

5.3.1.3 *Regional differences*

Post-nominal pronouns (in sandwich and colloquial variants) are known to vary according to region; in the north, 3rd person singular masculine is frequently *fo/o* and in the south it is more likely to appear as *fe/e*. In colloquial constructions, the 2nd person singular post-nominal pronoun also varies, where *chdi* is a widespread colloquial northern variant of *ti*.

King (2016: 93) claims that although the literary and sandwich variants represent a standard system for the spoken language, variation exists from region to region, particularly in relation to the realisation of mutations in each construction. Comparative work exists between

Rhymney (south-east Wales) and Rhosllanerchrugog (north-east Wales) in work by M. C. Jones (1998), who found that the use of the literary variant was more common in the south-east Wales cohort than in north-east Wales where the colloquial variant was more prominent. Similar patterns could be expected in the current data, which compares patterns in the north-west to the south-east.

5.3.1.4 Age differences

The replacement of the literary variant with the colloquial variant is identified as a significant age-related variant in Welsh (Watkins 1977; Thomas 1988b). Hatton (1988: 247) found that amongst young pupils at junior school, the younger cohort favoured the use of the colloquial variant, whereas the older cohort used the literary variant more often. The author hypothesises that an increase in awareness around formality constraints occurs in children sometime between the ages of the two cohorts.

M. C. Jones (1998) noted the prevalence of the colloquial among younger speakers in her work in Rhymney, pointing to the obsolescence of the more formal pre-nominal forms in this age group. More recent work by Davies (2016) also found that the colloquial variant is heavily favoured over the two other types by speakers under the age of 40. Indeed, this work concluded that the older the speaker, the more sandwich and literary variants they use and the less the colloquial variant (Davies 2016: 49).

5.3.2 Linguistic factors

Two linguistic factors have also been found to affect the rates of the possessive pronoun variants. In recent work on possessive pronoun construction in Welsh, certain frequently used possessed nouns (such as *tŷ* [house]) were only used with the colloquial variant, and Davies (2016: 44) hypothesised that high frequency nouns which show limited variation could point

to conventionalised “set phrases”; e.g. *ty ni* as [(our) home]. However, very little is documented about high frequency nouns in Welsh and conventionalised “set phrases”.

Davies (2016) also compared rates of the variants in third person singular constructions and first person plural constructions. Davies found that the colloquial variant is becoming more common for both grammatical person examples, whereas the colloquial variant is less frequent in the third person singular for older speakers (2016: 55). It is thought that the sandwich variant may be preferred in the 3rd person singular in order to distinguish between the pre-nominal pronoun *ei* (masculine) and *ei* (feminine) e.g. *ei gar o* [his car] and *ei char hi* [her car], which, without the mutation and post-nominal pronoun would be identical (King 2016: 94). No work has examined all grammatical person constructions in relation to possessive pronoun use in Welsh thus far.

Issues relating to mutations and the possessive pronoun variants are also discussed in wider literature. Research has indicated that mutation is not fully acquired (and even then not consistently produced), even by children learning Welsh as a first language until they reach 10 years old (Thomas and Gathercole 2007), and little is known about the effect of pedagogical intervention in education settings on this acquisition. Mutation is often described as a complex system and has been found to be a marker of style, age and gender (see Ball 1984; Ball *et al.* 1988). Previous literature (such as M. C. Jones 1998) has intimated that the requirement for mutation in the possessed noun might prompt the use of colloquial variant (so as to avoid the mutation). For example, in the first person singular, *teulu* [family] would mutate as ‘*fy nheulu*’, whereas using the colloquial construction instead would eliminate the need for mutation: *teulu fi*. The analysis presented in this chapter will examine whether the use of colloquial constructions is influenced by the mutation constraints relating to grammatical person.

5.4 Method

Having outlined previous literature that has examined possessive pronoun constructions in Welsh, I will now present how the feature appears in my data, and how it was coded for analysis.

5.4.1 *The feature in the data*

Three variants of the possessive feature have been discussed in this chapter so far, the literary, sandwich and colloquial variants. Though this categorisation represents a general rule regarding the stylistic meaning of variants, the labels above are not intended to imply that the literary variant only belongs in literary settings and colloquial variant in markedly informal settings. My dataset (comprised of only spoken Welsh) contains all three types.

Examples below from the data show the literary, sandwich and colloquial variants with varying 1st person singular pronouns and the noun *ffrindiau* [friends]. The literary variant is presented below:

fi wedi cadw cysylltiad gyda	fy	ffrindiau
	pre-nominal pronoun	noun
<i>[...I have kept in contact with</i>	<i>my friends]</i>	

The sandwich variant is presented below:

rhywle rhwng sut fi'n siarad gyda	fy	ffrindiau	i
	pre-nominal pronoun	noun	post-nominal pronoun
<i>[somewhere between how I talk to</i>	<i>my friends]</i>		

The colloquial variant is presented below:

Oedd yna sinema dw i'n cofio mynd efo	ffrindiau	fi
	noun	post-nominal pronoun
<i>[there was a cinema I remember going with</i>	<i>my friends]</i>	

The presence of the three variants in the data, as well as the previously mentioned stylistic markedness associated with each, and the emergence of the colloquial variant amongst younger speakers, provide a strong justification for conducting variationist research to explore the patterns of their variable use.

5.4.2 Data extraction

The transcribed text files were extracted from ELAN and AntConc was used to search for and extract pronouns. All pronoun tokens were then coded in Excel as literary, sandwich or colloquial variants. All reflexive pronouns in the fixed phrase *fy hun* [myself] ($n = 303$) were excluded from the analysis. Where fixed phrases were used, which are unlikely to take any other form than the fixed construction (e.g. *yn fy marn i* [in my opinion]) ($n = 16$) they were excluded from the statistical analysis. Two possessed nouns in the data could not be identified in the transcription process, and were thus coded <aneglur> [unclear]. These were excluded from the analyses.

Section 5.2.1 noted that the literary variant can contain clitic and enclitic pronouns; because examples of these were rare in the dataset, all enclitic pronouns were transcribed as prefixed pronouns, for example *'n chwaer* was transcribed as *fy chwaer* (the same is applied to enclitic pronouns in the sandwich variant e.g. *'n chwaer i* was coded as *fy chwaer i*). Section 5.2.2 demonstrated that the post-nominal pronoun in the sandwich variant can also serve an emphatic purpose, however no emphatic cases were found in the current dataset.

A total of 1968 possessive pronoun tokens were coded in the data. The breakdown of use by each speaker group (Gwynedd Welsh home language (GWHL), Gwynedd English home

language (GEHL), Cardiff Welsh home language (CWHL) and Cardiff English home language (CEHL)) is provided in Table 5.6.

Table 5.6: Token count for all speaker groups

	Possessive tokens N
GWHL	618
GEHL	253
CWHL	608
CEHL	489
Total	1968

Next, I turn to show what previous research has found about how the possessive pronoun varies, which has informed my analysis.

5.4.3 Social factors

The use of these variants has been widely reported to be stylistically marked (hence the variant labels used in this research based on Davies 2016), however little is known about how they vary across different speech contexts. The current work builds on previous work by comparing the use of the three variants in three stylistically-contrasting speech contexts; one formal (job interview context), one semi-formal (sociolinguistic interview) and one informal context (peer-to-peer context).

Previous research has also pointed to regional differences, with only very few large-scale quantitative research studies reporting these findings in the literature (M. C. Jones 1998). The current work will compare two sociolinguistically-contrasting areas in north-west and south-east Wales. Language exposure differences (i.e. learner speech *vs* native speech) has also been alluded to in the literature in relation to the use of the possessive pronoun variants. My research questions focus on home language differences, and whether those who speak English at home use different rates of the variants than their peers who speak Welsh at home. The speech data analysed here is attributed to four speaker groups to represent the regions and home language

backgrounds of speakers; Gwynedd Welsh home language (GWHL), Gwynedd English home language (GEHL), Cardiff Welsh home language (CWLH) and Cardiff English home language (CEHL).

5.4.4 Linguistic factors

5.4.4.1 Grammatical person

Possessive pronoun use has been found to be subject to constraints such as grammatical person, although only 3rd person singular and 1st person plural constructions have previously been considered. This research will present findings for all grammatical person constructions. Examples of each grammatical person with the three possessive variants are presented in Table 5.7 to demonstrate how this factor was coded in the data.

Table 5.7: Coding grammatical person

	Colloquial	Sandwich	Literary	English gloss
1st person singular	<i>mam fi</i>	<i>fy mam i</i>	<i>fy mam</i>	My mum
2nd person singular	<i>mam ti; chdi</i>	<i>dy fam di</i>	<i>dy fam</i>	Your mum
3rd person singular masculine	<i>mam fo/o; fe/ e</i>	<i>ei fam o/e</i>	<i>ei fam</i>	His mum
3rd person singular feminine	<i>mam hi</i>	<i>ei mam hi</i>	<i>ei mam</i>	Her mum
1st person plural	<i>mam ni</i>	<i>ein mam ni</i>	<i>ein mam</i>	Our mum
2nd person plural	<i>mam chi</i>	<i>eich mam chi</i>	<i>eich mam</i>	Your mum
3rd person plural	<i>mam nhw</i>	<i>eu mam nhw</i>	<i>eu mam</i>	Their mum

5.4.4.2 Lexical frequency

As mentioned previously, frequency could play a part in the use of possessive pronoun variants. Variationist sociolinguistic research has increasingly contained analyses of lexical frequency in order to examine frequency effects on morphosyntactic variation (Erker and Guy 2012; Bayley *et al.* 2013; Linford and Shin 2013; Linford *et al.* 2016), particularly in the field of L1

and L2 acquisition. Previous research has done this by determining lexical frequency based on the corpus under investigation (e.g. Ecker and Guy 2012).

Following this previous work, I determined lexical frequency within my own corpus (cf. Erker and Guy 2012). Frequent nouns were those occurring at a rate higher than two instances per 10,000 words of the corpus, which was 0.0288%. These nouns were extracted from the entire corpus (260, 658 words) using AntConc. These were then compared to the nouns that occurred in our set of possessive pronoun tokens. 91 nouns were found to be frequent in the corpus. These were “surface forms” rather than lemmas (Erker and Guy 2012: 530). That is to say that the frequent nouns appeared as singular and plural (e.g. *ffrindiau* and *ffrind* [friend]), in masculine and feminine (e.g. *athro* and *athrawes* [teacher]) and mutated and unmutated forms (e.g. *gwaith* and *waith* [work]). For ease, instances of plural, singular, masculine, feminine, mutated and unmutated forms were counted as a single noun. For example, *athro* [male teacher], *athrawes* [female teacher], and *athrawon* [teachers] were all coded as *athro*. The numerals 2, 3, 4, 5, 6 and 7, which appeared as frequent in the corpus, were checked again in AntConc for their collocates, and it was determined that they were mostly adnominal. For this reason they were removed. The same can’t be said for the numeral *un*, which was retained as it often acted as a pronoun.

After lemmatising the frequent nouns by combining singular and plural forms, feminine and masculine forms, and mutated and unmutated forms, the total number of frequent words was 68. English and Welsh nouns were kept separate even if they shared a meaning (e.g. job and *swydd*). 51 of the categorised frequent nouns occurred in possessive constructions, and the full list can be found in

Table 5.8.

Table 5.8: Frequent nouns in the corpus

Frequent noun	English gloss	Raw frequency in possessive constructions	Percentage of possessive constructions
ffrind	<i>friend</i>	197	10.01%
sgil	<i>skill</i>	91	4.62%
mam	<i>mum</i>	72	3.66%
enw	<i>name</i>	63	3.20%
dad	<i>dad</i>	59	3.00%
ffôn	<i>phone</i>	58	2.95%
chwaer	<i>sister</i>	57	2.90%
rhieni	<i>parents</i>	55	2.79%
gwaith	<i>work</i>	47	2.39%
bywyd	<i>life</i>	41	2.08%
teulu	<i>family</i>	39	1.98%
amser	<i>time</i>	32	1.63%
swydd	<i>job</i>	27	1.37%
diddordeb	<i>interest</i>	22	1.12%
tŷ	<i>house</i>	21	1.07%
Cymraeg	<i>Welsh</i>	20	1.02%
nain	<i>grandmother</i>	19	0.97%
iaith	<i>language</i>	17	0.86%
profiad	<i>experience</i>	17	0.86%
hyder	<i>confidence</i>	16	0.81%
syniad	<i>idea</i>	14	0.71%
athro	<i>teacher</i>	13	0.66%
pwnc	<i>subject</i>	12	0.61%
Saesneg	<i>English</i>	12	0.61%
TGAU	<i>GCSE</i>	12	0.61%
grŵp	<i>group</i>	11	0.56%
blwyddyn	<i>year</i>	9	0.46%
ysgol	<i>school</i>	9	0.46%
ffordd	<i>way</i>	8	0.41%
<i>job</i>	<i>job</i>	8	0.41%
dosbarth	<i>class</i>	7	0.36%
mathemateg	<i>mathematics</i>	7	0.36%
tîm	<i>team</i>	7	0.36%
acen	<i>accent</i>	5	0.25%
tiwtor	<i>tutor</i>	5	0.25%
celf	<i>art</i>	4	0.20%
gwers	<i>lesson</i>	4	0.20%
lle	<i>place</i>	4	0.20%
plentyn	<i>child</i>	4	0.20%
problem	<i>problem</i>	4	0.20%
wythnos	<i>week</i>	4	0.20%
cyfweliad	<i>interview</i>	3	0.15%
diwrnod	<i>day</i>	3	0.15%
llun	<i>picture/photo</i>	3	0.15%
bwyd	<i>food</i>	2	0.10%
coleg	<i>college</i>	1	0.05%
cwrs	<i>course</i>	1	0.05%
dydd	<i>day</i>	1	0.05%
help	<i>help</i>	1	0.05%
llyfr	<i>book</i>	1	0.05%
stwyff	<i>stuff</i>	1	0.05%
Total		1150	

All other possessed nouns ($n = 299$) were coded as infrequent. The mean raw frequency of these two groups is presented in Table 5.9, and show distinct results. Among frequent forms, the mean raw frequency is 22.5 in possessive constructions, whereas for infrequent forms it is 2.7.

Table 5.9: Mean raw frequency

Discrete frequency	Mean raw frequency	N tokens
Infrequent (nouns that are each in <0.02% of the data)	2.7	818
Frequent (nouns that are each in >0.02% of the data)	22.5	1150

The current work will analyse the frequency of the noun in the possessive construction under study in order to determine whether nouns with higher frequencies are influencing factors in the possessive variant produced by speakers.

5.4.4.3 *Mutation constraints*

Mutations are a morphophonological change determined by specific grammatical conditions (Thomas 1996:2). Such a grammatical condition is the pre-nominal possessive pronoun, where grammatical person constructions require the possessed noun to mutate. Section 5.2 highlighted that the mutation constraints relate to literary and sandwich variants.

Table 5.10 provides an overview of the mutations required with each grammatical person in possessive pronoun constructions (using the literary variant), with examples. The 2nd person plural construction does not require a mutation.

Table 5.10: Mutation constraints on possessive pronouns for each grammatical person

Grammatical person	Mutation required	Example
1 st person singular	nasal	<i>teulu</i> [family] > <i>fy dheulu</i> [My family]
2 nd person singular	soft	<i>mam</i> [mum] > <i>dy fam</i> [Your mum]
3 rd person singular feminine	aspirate	<i>car</i> [car] > <i>ei char</i> [her car]
3 rd person singular masculine	soft	<i>gwaith</i> [work] > <i>ei waith</i> [his work]
1 st person plural	h-prothesis before vowels	<i>afal</i> [apple] > <i>ein hafal</i> [our apple]
3 rd person plural	h-prothesis before vowel	<i>ewythr</i> [uncle] > <i>eu hewythr</i> [their uncle]

In order to explore whether the requirement for mutation is an influencing factor on the use of the colloquial variant, the data were coded to assess the following: whether the possessed noun mutates; which possessive pronoun variant is used (if the noun does mutate); and whether the mutation is realised (in the case of literary or sandwich variants).

5.4.4.4 *Language of the possessed noun*

Welsh contains many loan words and calques from English and many speakers are also known to code-switch and introduce English words when speaking Welsh (Deuchar, Donnelly and Piercy 2016). In order to see if this may have an effect on variant choice, I coded for recent loanwords using the following criteria: where an English loanword appeared in Geiriadur Prifysgol Cymru (i.e. historical Welsh language dictionary) as a Welsh word, it was coded as

Welsh. Other English words were coded as English. This is a commonly used criteria in research on codeswitching in Welsh (see Prys 2016).

5.4.4.5 *Possessed noun category*

The possessed noun can be categorised according to its alienability or inalienability. The alienability of possession is determined by “whether the object can exist apart from its possessor” (Nichols 1988; 575), and thus nouns such as ‘arm’ and ‘smile’ are examples of inalienable possession. On the other hand, if the possessed object can exist apart from its possessor it can be classed as alienable, as in the case of nouns such as ‘phone’ or ‘dog’. Oceanic languages, spoken in Papua New Guinea, Melanesia, Polynesia and Micronesia, distinguish grammatically between alienable and inalienable possessive constructions (Lichtenberk *et al.* 2011). English does not have such a grammaticalized distinction, and it is unknown whether the stylistic variation thought to exist in Welsh is influenced by the alienability of the possessed noun. The categories and subcategories of alienability have been adapted from Lichtenberk *et al.* (2011) and comprise the following (in Table 5.11), with examples from my own corpus:

Table 5.11: Possession type and category

Possession type	Category (adapted from Lichtenberk <i>et al.</i> 2011)	Example from the corpus
Inalienable	Companion and/or kin	<i>ffrind</i> [friend] <i>brawd</i> [brother] <i>teulu</i> [family]
	Part of body and/or body display	<i>calon</i> [heart] <i>bywyd</i> [life] <i>trwyn</i> [nose]

	Excretion and/or secretion	‘sweat patches’
	Behavioural and/or emotional state or manner	<i>barn</i> [opinion] <i>meddwl</i> [mind] <i>hyder</i> [confidence]
	Personal descriptors and/or attributes	<i>enw</i> [name] <i>sgil</i> [skill] <i>gwendid</i> [weakness]
	Membership	‘clique’ <i>tîm</i> [team] <i>dosbarth</i> [class]
Alienable	Ownership/ possession	<i>ffôn</i> [phone] <i>ceffyl</i> [horse] ‘pyjamas’
	Everyday use/ functioning	<i>iaith</i> [language] <i>amser</i> [time] <i>wythnos</i> [week]
	Control/command	‘boss’ <i>athro</i> [teacher] ‘supervisor’
	Activities related to manipulation/handling	<i>cyfrifoldeb</i> [responsibility] <i>pwnc</i> [subject] <i>hobi</i> [hobby]
	Activity leading to material or aesthetic product	<i>gwaith</i> [work] <i>swydd</i> [job]

		<i>coginio</i> [cooking]
	Experience on one's own (solitary activity)	-

No examples were found in the corpus representing the category 'experience on one's own'.

5.5 Results (distributional analysis)

In this sub-section, the results of the distributional analysis and statistical analysis will be presented. In the distributional analysis, each linguistic and social factor will be analysed in order to explore the general patterning of use in the data.

Overall, as Table 5.12 shows, a total of 1968 possessive pronoun variants were coded in the data. The colloquial construction is the most frequently used construction by all participants ($n = 1142$, 58%). Literary constructions feature second-most prominently ($n = 555$, 27%), and sandwich constructions were used least often ($n = 311$, 15%).

Table 5.12: Overall distribution of possessive pronoun variants

Pronoun type	%	n
Colloquial	58%	1142
Literary	27%	524
Sandwich	15%	302
Total		1968

5.5.1 Region

As introduced in section 4.2 in the Methodology chapter, the two sites under study are Gwynedd, where Welsh is commonly heard in the community, and Cardiff, where it is rarely heard in the community. Figure 5.1 shows the pattern of pronoun variant use in both regions.

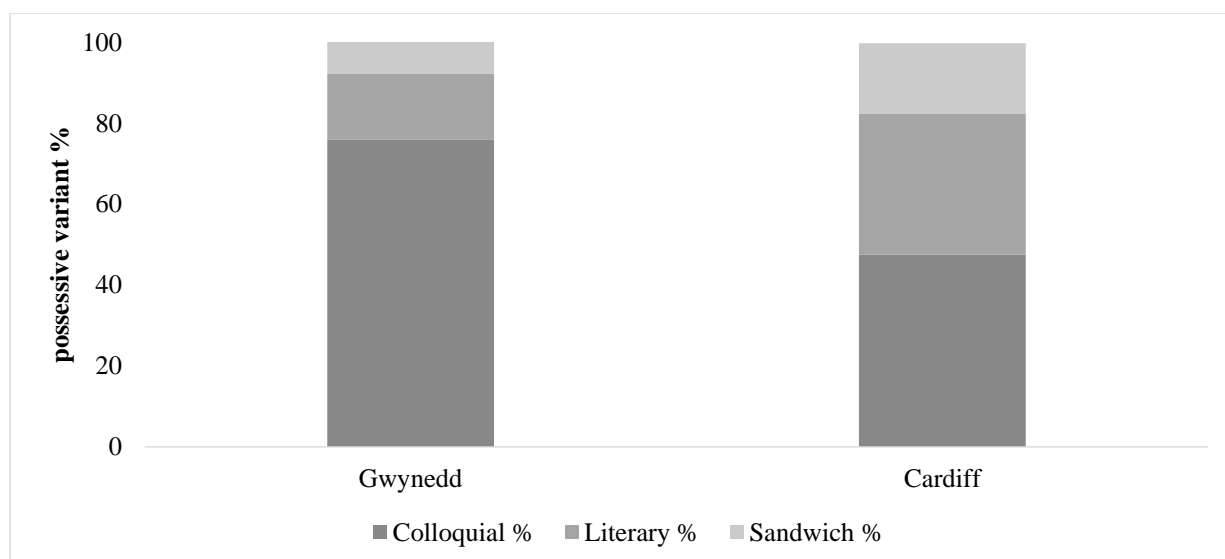


Figure 5.1: Rates of possessive pronoun variants in Gwynedd and Cardiff

In Gwynedd, the colloquial pronoun type was used most frequently (76%), followed by the literary type (16%). The sandwich type is used least frequently (13%) by the Gwynedd participants. This ascending frequency pattern (colloquial, literary, sandwich) is mirrored in the Cardiff data, although the rates of use differ. Overall, in Cardiff, the colloquial pronoun type is also used most frequently (48%), but is followed closely, however, by the literary type (35%). The sandwich type is used least frequently (18%) by the Cardiff participants.

5.5.2 Home language

As discussed in the methodology six students from Gwynedd came from predominantly Welsh-speaking homes, and four came from predominantly English-speaking homes. Table 5.13 presents the distributions of possessive pronoun variant according to area and home language.

Table 5.13: Possessive pronouns variants by area and home language

Area	Home language	Colloquial %	Literary %	Sandwich %	Total (N)
Gwynedd	WHL	63	21	17	618

	EHL	91	6	3	253
Cardiff	WHL	42	41	17	608
	EHL	55	27	18	489

Similar patterns of use can be seen in the data according to home language, with colloquial constructions appearing most frequently, followed by literary and then sandwich constructions. In all four groups, the patterns of possessive pronoun use are the same; the most frequently occurring is colloquial, followed by literary, followed by sandwich. The rate of use between the four groups are distinct and are shown further in Figure 5.2. Cardiff WHL pupils, use very similar rates of colloquial and literary variants, whereas Gwynedd EHL are near-categorical in their use of the colloquial variant.

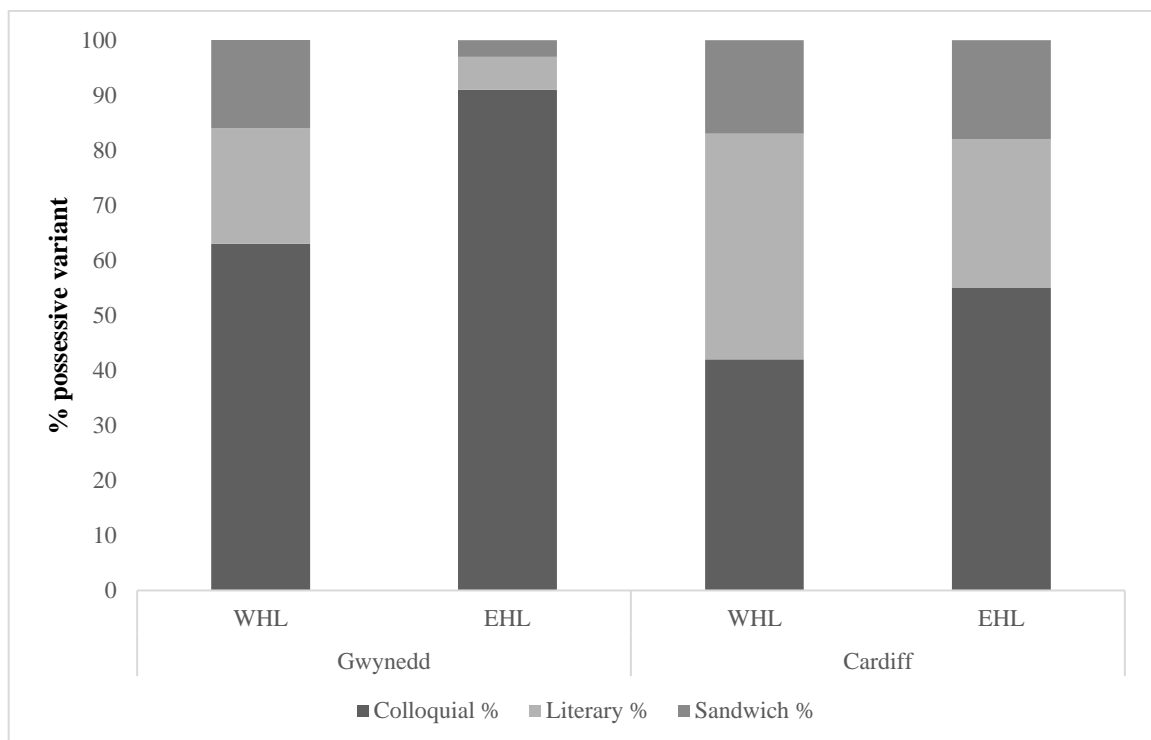


Figure 5.2: Rates of possessive pronoun variants by area and home language

Considering the differences in the use of the variants between the two sites and between the different home language groups, the remainder of the distributional analysis will present the data for each area and home language group separately.

5.5.3 *Speech context*

The next relevant factor is speech context; recall that the data were gathered from peer-to-peer conversations, sociolinguistic interviews and mock job interviews. These contexts were designed to elicit increasing levels of formality, as mentioned in the methodology (4.5). Table 5.14 shows the distribution of possessive pronoun variants according to speech context for each of the speaker groups identified in the previous section.

Table 5.14: Possessive pronouns variants by speech context, by speaker group

Speaker group	Speech context	Colloquial %	Literary %	Sandwich %	Total (N)
GWHL	Job interview	45	34	21	123
	Sociolinguistic interview	69	15	16	269
	Peer to peer	66	20	14	226
GEHL	Job interview	84	11	4	44
	Sociolinguistic interview	98	1	1	148
	Peer to peer	80	13	7	61
CWHL	Job interview	6	66	28	155
	Sociolinguistic interview	56	32	13	215
	Peer to peer	53	33	14	238
CEHL	Job interview	17	49	34	86
	Sociolinguistic interview	75	11	15	206
	Peer to peer	50	35	15	197

In the table, strong shifting patterns can be observed. Overall, the sociolinguistic interview has the highest rates of colloquial variant use, and the lowest rates of colloquial variants can be observed in the job interview context (apart from GEHL who have lower rates of colloquial in the peer-to-peer context). Indeed, students in Gwynedd who spoke English at home favoured the use of the colloquial possessive pronoun type, across the three speech

contexts. Even in the context designed to elicit the most formal speech from participants (the mock interview) the colloquial possessive variant was used most frequently.

The distribution of pronoun types is different in the dataset from participating students in Gwynedd who spoke Welsh at home. Their use of this variant was much higher in the sociolinguistic interview and peer-to-peer contexts, pointing to possible stylistic variation between the three contexts.

The patterns of pronoun types used by students in Cardiff who spoke English at home also suggests stylistic variation. Participants from Cardiff who spoke English at home used the colloquial variant least frequently (17%) in the most formal context. On the other hand, the colloquial type was the most frequent variant in the other contexts. This pattern was also found for the students in Cardiff who spoke Welsh at home. Figure 5.3 presents the distribution of pronoun variants across the three contexts.

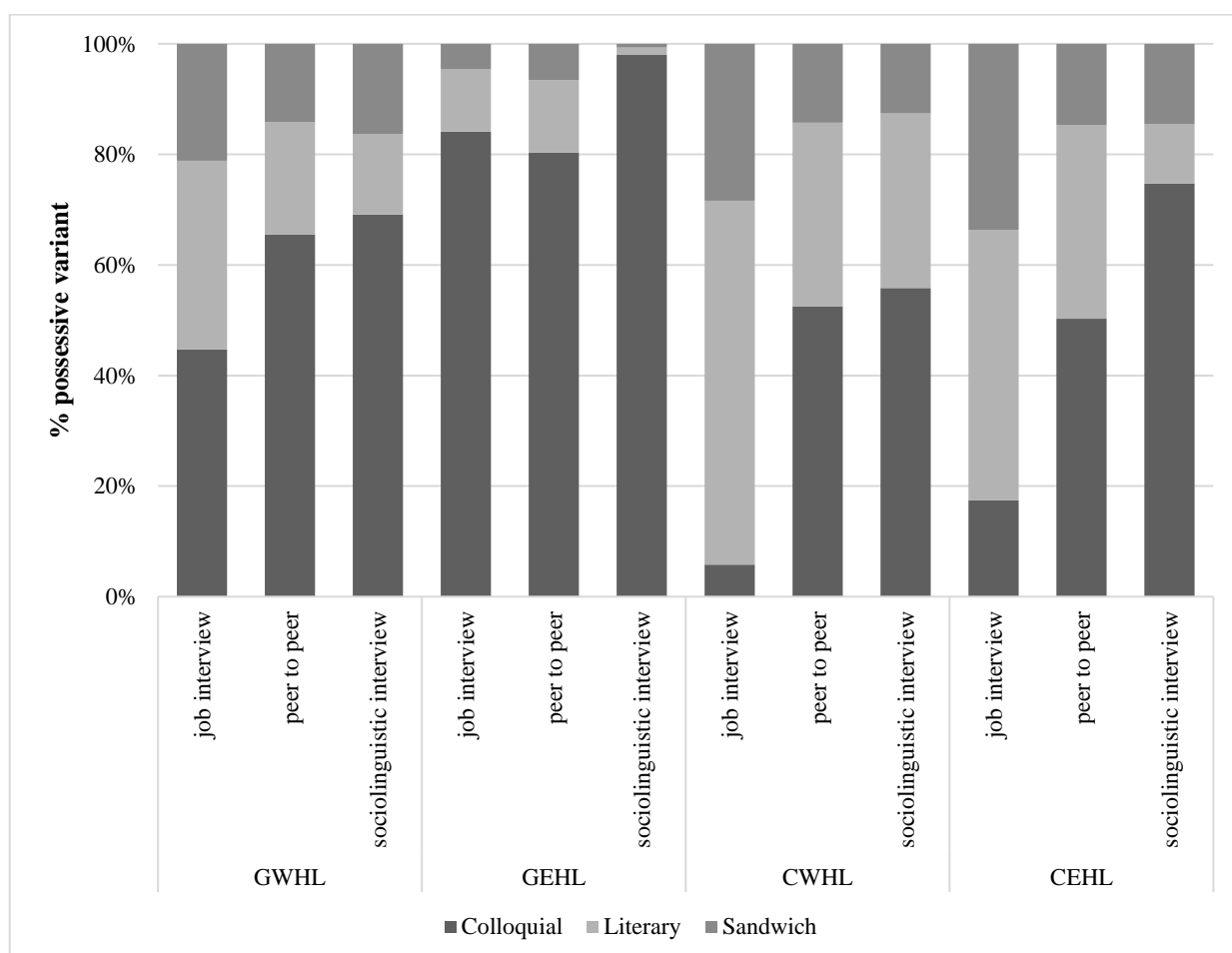


Figure 5.3: Rates of possessive pronoun variants in each speech context by speaker group

The figure shows that all speaker groups produced more of the colloquial variant in the sociolinguistic interview. This differs from what was expected insofar as the peer-to-peer context was devised as being the most informal context. This will be discussed further later as participants could perceive similar levels of familiarity between the sociolinguistic interview and peer-to-peer context as they might not be used to speaking Welsh with peers. The evidence of stylistic variation and shifting between contexts is less evident in the Gwynedd EHL data but does appear in other groups.

5.5.4 Possessed noun category

The first linguistic factor examined as part of the analysis of this feature is the category of the possessed noun as alienable or inalienable. This analysis was conducted to determine whether certain types of noun are used more commonly with colloquial, sandwich or literary possessive constructions by young speakers of Welsh.

A total of 590 (30%) alienable constructions were identified in the corpus, and 1382 (70%) inalienable constructions were identified. The sub-categories are identified in Table 5.15, along with the rates of use of each possessive construction variant for each category.

Table 5.15: Possessive pronoun variants by alienability sub-category

	Colloquial %	Literary %	Sandwich %	Total N
Alienable	54	31	15	590
Command	63	30	7	30
Handling	44	36	20	95
Ownership	64	23	13	219
Product	45	40	15	158
Use	55	27	18	88
Inalienable	60	25	16	1378
Attributes	43	33	24	324
Behavioural	43	36	21	198
Body	61	27	12	154
Excretion	100	0	0	1
Kin	73	18	9	640
Membership	52	15	33	61

The general tendency for this factor mirrors previously mentioned patterns in the social factor analysis; the pronoun constructions used for all categories are mostly colloquial, followed by literary and rates of the sandwich construction are lowest of all.

Inalienable possessed nouns took relatively more colloquial constructions (60%) than alienable (54%), and the differences between colloquial and literary rates is greater in inalienable constructions (25% literary in inalienable vs 31% in alienable). This is not the case, however, with the alienable sub-categories of command (e.g. *athrawes fi* [my teacher]) and ownership (e.g. *ffôn fi* [my phone]), which take far more of the colloquial construction than any other. Kin and body sub-categories of inalienable possessed nouns (such as *braich fo* [his arm] and *mam fi* [my mum]) were also far more likely to appear with the colloquial construction (60%+). The inalienable subcategory membership type is an exception to previously mentioned patterns, with more use in sandwich constructions (33%), such as *fy nosbarth i* [my class] than literary constructions (15%). It is possible that the echoed pronoun emphasised group membership, thus influencing the use of this variant.

Table 5.16 shows the proportions of alienable and inalienable nouns in the three possessive pronoun constructions, by area and home language.

Table 5.16: Possessive pronoun variants by alienability and speaker group

Speaker group	Alienability	Colloquial %	Literary %	Sandwich %	Total N
GWHL	Alienable	61	24	15	208
	Inalienable	65	19	17	410
GEHL	Alienable	84	14	2	50
	Inalienable	93	4	3	203
CWHL	Alienable	45	41	14	184
	Inalienable	40	40	19	424
CEHL	Alienable	47	32	21	148
	Inalienable	58	26	17	341

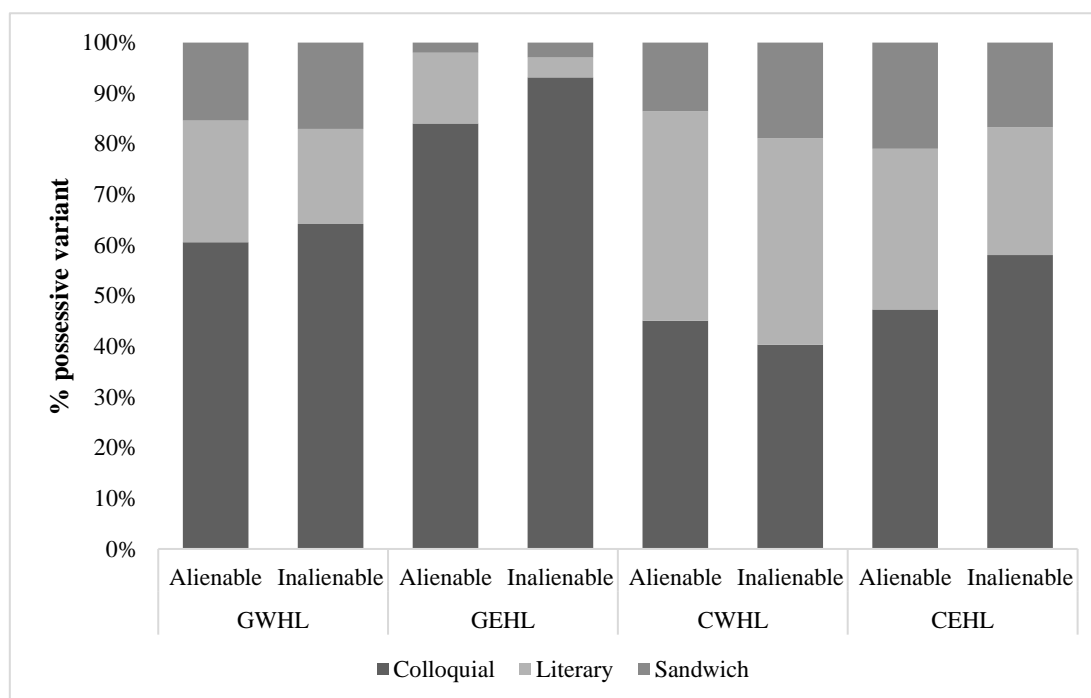


Figure 5.4: Rates of possessive pronoun variants by alienability and speaker group

In the case of all speaker groups but Cardiff WHL, inalienable possessed nouns elicit slightly higher rates of the colloquial variant and lower rates of the literary variant than alienable possessed nouns, which are on the whole marginally more formal. In the Cardiff WHL data, however, inalienable possessed nouns elicit lower rates of colloquial than alienable possessed nouns. Differences between EHL groups in Cardiff and Gwynedd compared with differences between WHL groups in Cardiff and Gwynedd will be explored in greater detail in the mixed effects model presented later in this chapter.

5.5.5 Lexical frequency

Table 5.17 shows the proportion of each variant used with frequent and infrequent nouns, broken down by speaker group. The aim here is to determine whether nouns considered frequent are more likely to occur with the colloquial variants which has been alluded to in previous literature (see Davies (2016: 44) *tŷ ni* [our house] as a set phrase example).

Table 5.17: Possessive pronouns variants by frequency, by speaker group

Speaker group	Frequency	Colloquial %	Literary %	Sandwich %	Total N
GWHL	Frequent	62	20	17	318
	Infrequent	63	21	16	300
GEHL	Frequent	92	6	2	165
	Infrequent	90	6	3	88
CWHL	Frequent	40	43	16	375
	Infrequent	43	38	18	233
CEHL	Frequent	58	26	16	292
	Infrequent	50	29	21	197

The table above and Figure 5.5 below show that in all speaker groups, frequent nouns are more likely to elicit the colloquial possessive variant than infrequent nouns.

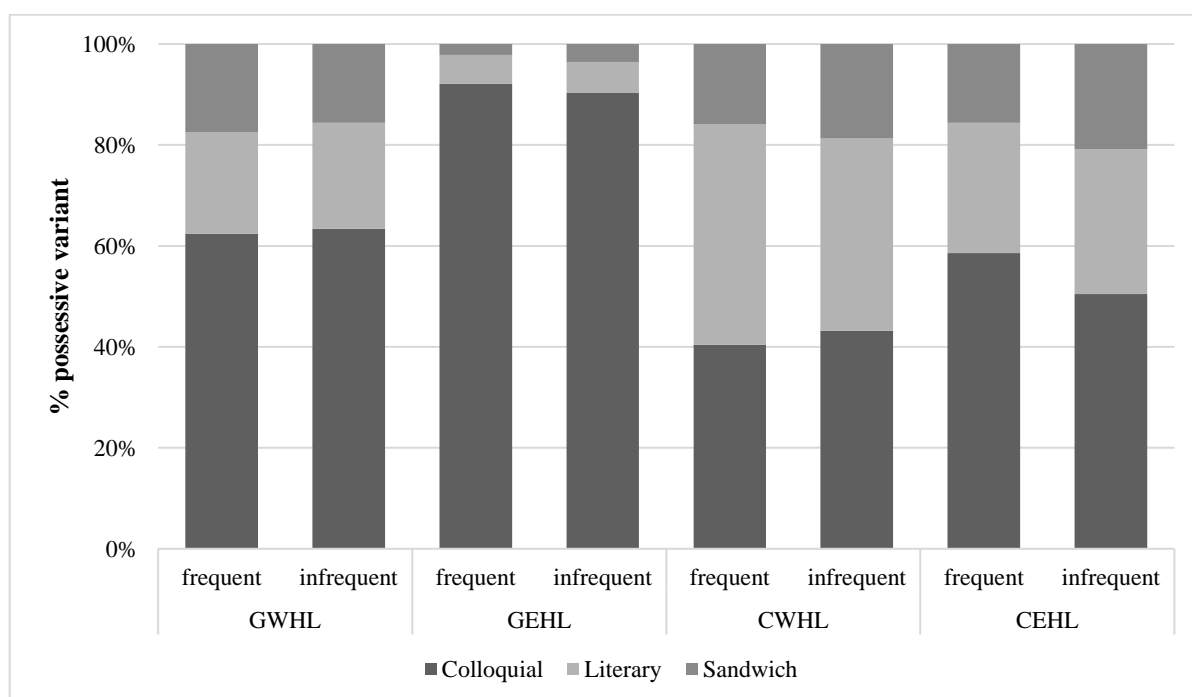


Figure 5.5: Rates of possessive pronoun variants by frequency and speaker group

5.5.6 Grammatical person

The next factor considered in the analysis of the possessive pronoun construction in Welsh is grammatical person²⁸. First person singular was the most frequent type of possessor in the dataset ($n = 1201$), and the first-person singular generally has higher rates of the colloquial variant compared with more the formal variants than the remaining pronouns for all four groups. Recall that Davies (2016) had found that rates of the colloquial third-person singular and the first-person plural forms increased in younger speakers. Following Davies, I have highlighted these rates in the WHL subset in the tables below. For the purpose of analysing whether this may be the case in my own data, I combined third-person singular masculine and feminine constructions.

Table 5.18: Possessive variants used by Gwynedd speakers by grammatical person with highlighted 3rd person singular and 1st person plural

Grammatical person	Gwynedd WHL				Gwynedd EHL			
	Coll %	Lit %	Sand %	<i>N</i>	Coll %	Lit %	Sand %	<i>N</i>
First-person singular	67	19	14	349	96	3	1	162
Second-person singular	55	25	20	49	86	0	14	7
Third-person singular	69	15	17	89	98	0	2	41
First-person plural	79	13	8	24	71	19	9	21
Second-person plural	30	40	30	40	20	40	40	5
Third-person plural	57	24	19	67	77	24	0	17

²⁸ Although Welsh distinguishes between T/V 2nd person informal and formal (*ti* and *chi*), only one example of the latter was observed in the data (where a participant asked a direct question of the researcher). For this reason, it has not been included in the analysis.

Table 5.19: Possessive variants used by Cardiff speakers by grammatical person with highlighted 3rd person singular and 1st person plural as per Davies (2016)

	Cardiff WHL				Cardiff EHL			
Grammatical person	Coll %	Lit %	Sand %	N	Coll %	Lit %	Sand %	N
First-person singular	42	47	11	403	55	30	15	287
Second-person singular	33	54	13	24	41	41	18	17
Third-person singular	45	24	28	71	66	14	21	73
First-person plural	45	16	39	31	45	14	41	22
Second-person plural	35	45	20	20	54	35	10	48
Third-person plural	39	29	34	59	45	21	33	42

Low numbers in some categories (especially for the GEHL group) mean that some rates may not be fully reliable, but for the two WHL groups, the two categories that Davies (2016) studied (third-person singular and first-person plural) have higher rates of the colloquial form compared with all other grammatical person constructions. With respect to the third-person singular this is also the case for the EHL groups.

When considering the sandwich and literary variants, the patterns seem to be tied to community, more than home language. For the Gwynedd speakers, second-person plural has the highest rates of both the literary and the sandwich forms. Instead, in Cardiff, the rates of the literary form are highest in second-person singular, and for the sandwich form they are highest in first-person plural. Based on this, the EHL students share some, but not all, of the general patterning with their WHL counterparts, but do share the more local patterns of use.

5.5.7 Possessed noun language

Although all the most frequent possessed nouns in the corpus were Welsh (see Lexical frequency subheading), there are 71 English possessed words in the corpus, and 281 Welsh possessed nouns. Most English possessed nouns only appeared once in the corpus, but some

(e.g. cousin $n = 10$, job $n = 7$, boss $n = 7$) were repeated. Overall, where the possessed noun language was English, the colloquial construction was used far more frequently (86%) than in Welsh constructions (56%). These distributions point to Welsh nouns eliciting more formal constructions, and English nouns eliciting more informal constructions.

Table 5.20 presents the proportions of Welsh and English possessed nouns in the three possessive pronoun constructions, by speaker group.

Table 5.20: Possessive variants used by speaker group and possessed noun language

Speaker group	Possessed noun language	Colloquial %	Literary %	Sandwich %	Total (N)
Gwynedd EHL	English	91	4	4	23
	Welsh	91	6	3	230
Gwynedd WHL	English	85	10	6	52
	Welsh	61	22	18	566
Cardiff EHL	English	89	5	5	19
	Welsh	53	28	19	470
Cardiff WHL	English	79	16	5	19
	Welsh	41	42	18	589

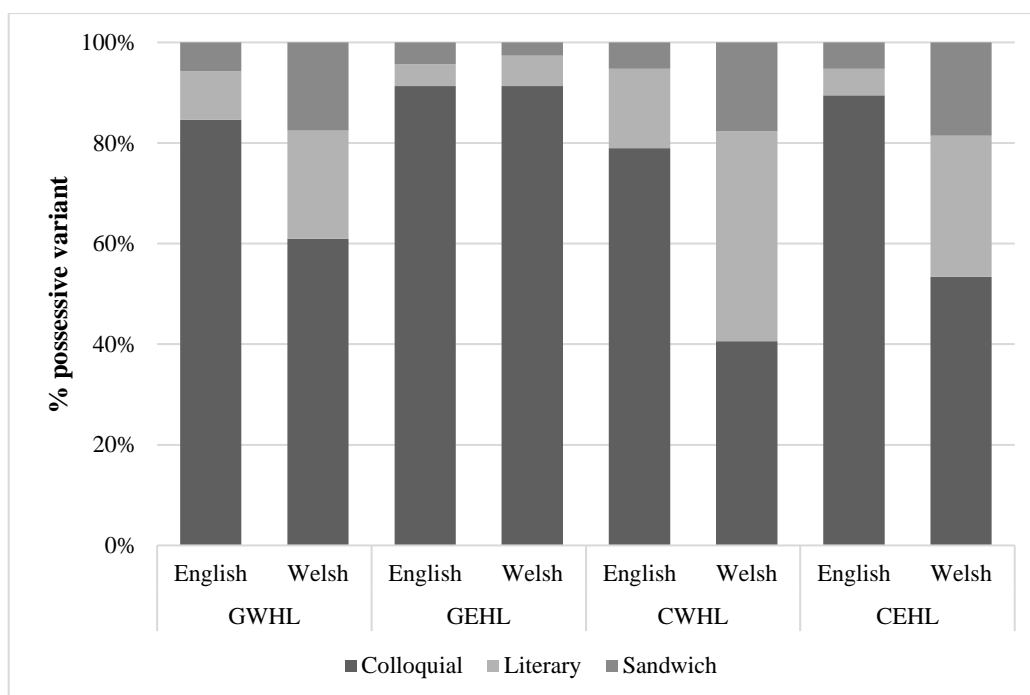


Figure 5.6: Rates of possessive pronoun variants by noun language and speaker group

Overall, patterns are shared between GWHL, CWHL and CEHL groups; the rate of colloquial variants drops with Welsh nouns, compared with English nouns. Indeed, Figure 5.6 demonstrates that in Cardiff, the EHL group seems to be replicating the variation displayed by the WHL group, whereas in Gwynedd, EHL students are not following the pattern of their counterparts.

The data for the language of the possessed noun shows a notable difference between the use of English possessed nouns in Gwynedd and in Cardiff. Gwynedd EHL students used English nouns in 9% of the corpus, and Gwynedd WHL students used English possessed nouns in 8% of the corpus. In Cardiff, these rates are halved, with EHL students using English possessed nouns 4% of the time, and WHL even less frequently at 3% of the time. Students in Cardiff codeswitching less often could point to the enforced educational standard of Welsh being more prevalent in Cardiff, leading young speakers to opt for less non-standard language.

5.5.8 Mutation constraints

Previous studies of possessive pronoun variants have claimed that speakers opt to use the colloquial construction as a form of grammatical simplification, as it removes the requirement to mutate after the pre-nominal pronoun (see M. C. Jones 1998: 62). Most of the possessive constructions in the corpus ($n = 1303$; 66%) contained a possessed noun which did not require mutation, while 34% ($n = 665$) required mutation. Of the ones requiring mutation, 342 tokens (51%) appeared in colloquial constructions, thus eliminating the need for mutation. The remaining 49% ($n = 323$) of constructions were literary or sandwich constructions (where the mutation would be required). This demonstrates that participating students were as likely to use a colloquial construction as they were with a literary or sandwich construction when the possessed noun required a mutation. Although the literature has intimated that the use of the colloquial variant may point to a speaker avoiding a mutation, the current findings cannot support that assertion. For this reason, requirement for mutation will be omitted from the statistical analysis, presented in the next section.

5.6 Results (inferential analysis)

As mentioned in section 4.7.3, variationist comparative sociolinguistics draws on three lines of evidence in order to determine patterns of variation in different populations. Firstly, I use mixed effects logistic regression analysis in order to determine the significance of external and internal factors on the variation of the possessive pronoun variant. The statistical analysis compares the use of colloquial vs non-colloquial variants and thus uses a binomial analysis to ascertain whether use is more informal or formal in line with the factors summarised in 5.6.2 below.

Following the mixed effects model regression analysis, I use a Wald χ^2 (chi square) test, which iteratively adds and removes each predictor and compares how well each iteration fits the distribution of the data (Gardner, 2023), in order to determine the size of the effect, and based on this, a parsimonious model, containing only significant predictors of variation is presented. The Wald χ^2 test results provide evidence for the 2nd line of evidence, and significant factors in the parsimonious model are counted as the 1st line of evidence. The 3rd line of evidence is presented in the form of log likelihood figures, to show the hierarchy of levels within each factor.

This analysis is presented for the whole dataset, and then subset into the four distinct speaker groups to allow comparison between areas and home language backgrounds.

5.6.1 Individual variation

All participants bar one varied in their use of possessives. A GWHL participant used no colloquial variants, opting instead for the relatively uncommon sandwich variant ($n = 17$) and literary variants ($n = 15$) in all cases. Because the statistical analysis below attempts to predict which factors aid the selection of one form over another (colloquial vs not), it was not appropriate to include participants who did not vary in the regression analysis. As such, the

data collected from this participant (a categorical respondent), is not included in the subsequent regression analysis.

5.6.2 Summary of factors

Table 5.21 presents each of the random and fixed factors influencing the possessive construction. The levels presented here were built into the statistical model. Note that mutation is not included in the statistical analysis, as discussed in the descriptive results section (5.4.4).²⁹

Table 5.21: Factors included in the statistical modelling

Factor	Levels
Context	Mock job interview Peer to peer Sociolinguistic interview
Speaker's home language	English Welsh
Region	Cardiff Gwynedd
Grammatical person	1st person plural 1st person singular 2nd person plural 2nd person singular 3rd person plural 3rd person singular
Possessed noun category	Alienable

²⁹ The baseline for each factor level is discussed in subsequent sections and is thus not highlighted in this initial table.

	Inalienable
Possessed noun language	Welsh English
Frequency	Frequent Infrequent

5.6.3 Overall model

Firstly, the model included the maximal random effects structure, and then the structure was simplified until the model converged (Baranowski and Turton 2020; Bates Douglas *et al.* 2015). Participant and noun were controlled for in the model. The maximal model is shown in the below R code:

```
VARIANT ~ CONTEXT * HOME LANGUAGE * AREA + GRAMMATICAL PERSON +  
NOUN LANGUAGE + ALIENABILITY + FREQUENCY + (1|PARTICIPANT) + (1|NOUN)
```

Interactions between social factors were explored in a three-way interaction model including the external factors (context, home language, area), which was found to have a significantly lower deviance ($p < 0.001$) than the model without interactions, and therefore the interactions were kept.

The results tables presented hereon in show the fixed factors which were significant predictors of colloquial variant use. Regression coefficients (β) for each term indicate deviations for the intercept and are included alongside z -values and p -values for the levels associated with each factor. A positive significant coefficient suggests that the named factor level was more likely to influence the production of the colloquial variant than the baseline

factor level. Conversely, a negative significant coefficient indicates that the named factor level was less likely to result in the production of the colloquial variant than the baseline factor level.

Table 5.22 presents the results of generalised mixed-effects modelling on the entire dataset. The baseline is colloquial possessive pronouns. Each factor in this model was reordered based on level estimates so that the reference level, i.e., first level, was also the level that most favoured the application value (i.e. has the highest probability of colloquial pronoun use). The estimate value for the reference level is 0. The reader will notice that all the estimates are negative (and are therefore less likely to be used with the colloquial variants than the reference level). Within each factor, the estimate value of each level is descending; this process shows the constraint hierarchy for the predictor. The overall model presented in Table 5.22 shows which levels of certain predictors are significantly (or not) different from the baseline (the reference level shown in brackets after each fixed effect).

Table 5.22: Final generalised mixed effects model for possessive pronoun variants

Fixed effects:	Estimate	Std. Error	z-value	p-value	Sig ³⁰	Observations	
						n	% coll.
(Intercept)	6.23	1.11	5.64	<0.001	***		
CONTEXT (<i>vs Sociolinguistic interview</i>)						838	72
peer to peer	-2.24	0.73	-3.05	0.002	**	722	58
job interview	-1.67	0.80	-2.08	0.03	*	408	28
HOME LANGUAGE (<i>vs EHL</i>)						742	67
WHL	-2.12	1.16	-1.87	0.07		1226	52
REGION (<i>vs Gwynedd</i>)						871	71
Cardiff	-3.21	1.19	-2.69	0.007	**	1097	48
GRAMMATICAL PERSON (<i>vs 3rd person plural</i>)						185	50
3 rd person singular	-0.38	0.32	-1.21	0.22		97	66
1 st person singular	-0.61	0.26	-2.39	0.02	*	1201	59
2 nd person singular	-0.63	0.40	-1.58	0.11		97	49
1 st person plural	-0.71	0.43	-1.67	0.09		98	59
2 nd person plural	-1.48	0.37	-3.74	<0.001	***	113	40
NOUN LANGUAGE (<i>vs English</i>)						113	86
Welsh	-1.49	0.44	-3.41	<0.001	***	1855	56
ALIENABILITY (<i>vs Alienable</i>)						590	54
Inalienable	-0.03	0.25	-0.12	0.89		1378	60

³⁰ No significance where $p > 0.05$; * where $p \leq 0.05$; ** where $p \leq 0.01$; *** where $p \leq 0.001$

FREQUENCY (<i>vs Frequent</i>)						1150	61
Infrequent	-0.35	0.27	-1.28	0.19		818	54
CONTEXT*HOME LANGUAGE (<i>vs sociolinguistic interview*EHL</i>)							
Peer to peer*WHL	2.24	0.79	2.82	0.004	**		
Job interview*WHL	1.33	0.87	1.52	0.13			
CONTEXT*REGION (<i>vs sociolinguistic interview*Gwynedd</i>)							
Peer to peer*Cardiff	1.76	0.78	2.25	0.02	*		
Job interview*Cardiff	-0.51	0.89	-0.57	0.56			
HOME LANGUAGE*REGION (<i>vs EHL*Gwynedd</i>)							
WHL*Cardiff	0.87	1.54	0.56	0.57			
CONTEXT*HOME LANGUAGE*REGION (<i>vs sociolinguistic interview*EHL*Gwynedd</i>)							
Peer to peer*WHL*Cardiff	-1.63	0.88	-1.85	0.06			
Job interview*WHL*Cardiff	-2.38	1.05	-2.26	0.02	*		
Random effects:						<i>sd</i>	<i>n</i>
SPEAKER						1.38	17
NOUN						1.26	351

Speech context emerged as significant, with the peer to peer and the job interview contexts significantly less likely to contain colloquial possessive types than the sociolinguistic interview. The region of the school was also a strong predictor of possessive variation, with students from Gwynedd being significantly more likely to use the colloquial variant than their counterparts in Cardiff. Where the noun language was Welsh, students were also significantly less likely to use the colloquial variant, compared to where the noun language is English. 2nd person plural constructions and 1st person singular constructions were significantly less likely to be used with colloquial pronouns; that is, *car fi* [my car] and *car chi* [your car] were less likely to appear in the dataset than the baseline 3rd person plural level.

The model also shows a three-way interaction between context, home language, and region. The interaction suggests that there is a difference in the effect of the context within the home language and area groups. The effect is negative, meaning that the effect of the sociolinguistic interview on the production of the colloquial variant is less in the Gwynedd WHL group compared to the Cardiff EHL group. In other words, the effect of speech context

on the production of the colloquial variant differs between the area and home language groups and in some cases these differences are significant. This is going to be examined in further detail by subsetting the speaker groups later in the chapter, to examine more closely the extent to which speakers from different home language backgrounds in different areas are varying according to context, which is crucial to answering the research questions of this thesis.

It is not possible to assess the hierarchy of constraints until a sparser and more parsimonious model is created, using only factors that contribute significantly to predicting the variation. In order to assess which predictors are most explanatory of the variation, I will use the Wald χ^2 (chi square) test, which iteratively adds and removes each predictor and compares how well each iteration fits the distribution of the data (Gardner, 2023). The interactions in the maximal model were removed for the Wald χ^2 test, but any factors in interactions which came out as significant were included. The results of the Wald χ^2 test are presented in Table 5.23, in ascending order of significance (the lowest p -value indicates the highest significance, with $\alpha=0.05$).

Table 5.23: Analysis of deviance, Wald χ^2 test for maximal model

	Wald χ^2	Df	p -value	
Context	72.59	2	<0.001	***
Noun language	13.38	1	<0.001	***
Region	14.94	1	<0.001	***
Grammatical person	18.11	5	0.002	**
Home language	2.27	1	0.13	
Frequency	1.34	1	0.25	
Alienability	0.13	1	0.72	

Context has the largest χ^2 value (72.59) indicating it has the greatest effect on the variation. This is followed by Grammatical person (18.11), Region (14.94), and Noun language (13.38). Neither home language, frequency nor alienability were found to be significant predictors of the variation in possessive pronoun constructions. As stated earlier in this chapter,

all predictors found to be significant can be interpreted as adding explanatory value to the study of variation, and thus I next present a more parsimonious model, containing only significant predictors of variation.

The order of parameters in Table 5.24 is based on the relative ordering in the Wald χ^2 test; i.e. context, the strongest predictor of colloquial pronoun variation, is listed first. The parameter levels are also ordered by their estimates; i.e. the job interview context was the least likely to contain colloquial variants and is thus listed last within the factor (group).

Table 5.24: Parsimonious model with only significant predictors of variation

AIC = 1725.9						Observations	
Fixed effects:	Estimate	Std. Error	z-value	p-value	Sig	n	% coll.
(Intercept)	4.48	0.72	6.64	<0.001	***		
CONTEXT (<i>vs Sociolinguistic interview</i>)						838	72
peer to peer	-0.21	0.19	-1.14	0.26		722	58
job interview	-1.78	0.22	-7.90	<0.001	***	408	28
NOUN LANGUAGE (<i>vs English</i>)						113	86
Welsh	-1.55	0.44	-3.52	<0.001	***	1855	56
GRAMMATICAL PERSON (<i>vs 3rd person plural</i>)						185	50
3 rd person singular	-0.48	0.31	-1.55	0.12		97	66
1 st person singular	-0.64	0.26	-2.44	0.01	*	1201	59
2 nd person singular	-0.66	0.39	-1.67	0.09		97	49
1 st person plural	-0.87	0.43	-2.05	0.04	*	98	59
2 nd person plural	-1.55	0.38	-4.13	<0.001	***	113	40
REGION (<i>vs Gwynedd</i>)						871	71
Cardiff	-2.69	0.76	-3.53	<0.001	***	1097	48
Random effects:						sd	n
SPEAKER						1.39	17
NOUN						1.30	351

The more formal job interview context was significantly less likely to contain the colloquial possessive pronoun variant. Noun language was also a significant predictor; Welsh nouns were significantly less likely to be used with colloquial variants than English ones. 1st person (singular and plural) and 2nd person plural constructions were significantly less likely to be used with the colloquial variant than the reference level. But what about the other

parameter levels? What if there are significant differences within the predictor, beyond the reference level? In order to explore the pattern of variation in greater detail, rather than just comparing each level to the reference level, a contrast matrix of all the comparisons within the grammatical person predictor (e.g. 3rd person singular vs 1st person singular, 3rd person singular vs 2nd person singular, etc.) was run using the *glht* function in R (Hothorn, Bretz, and Westfall 2008) following previous sociolinguistic work by Tagliamonte and Gardner (2020: 244).

Table 5.25: Comparison matrix for possessive by grammatical person

	Estimate	Std. Error	z-value	p-value	Sig
3rd person sing - 3rd person plu == 0	-0.46	0.31	-1.49	0.65	
1st person sing - 3rd person plu == 0	-0.63	0.26	-2.41	0.14	
2nd person sing - 3rd person plu == 0	-0.64	0.39	-1.62	0.56	
1st person plu - 3rd person plu == 0	-0.86	0.43	-2.03	0.31	
2nd person plu - 3rd person plu == 0	-1.54	0.38	-4.10	<0.001	***
1st person sing - 3rd person sing == 0	-0.17	0.22	-0.77	0.96	
2nd person sing - 3rd person sing == 0	-0.18	0.37	-0.50	0.99	
1st person plu - 3rd person sing == 0	-0.41	0.41	-1.00	0.91	
2nd person plu - 3rd person sing == 0	-1.09	0.35	-3.13	0.01	*
2nd person sing - 1st person sing == 0	-0.01	0.34	-0.04	1	
1st person plu - 1st person sing == 0	-0.23	0.36	-0.64	0.98	
2nd person plu - 1st person sing == 0	-0.91	0.31	-2.97	0.03	*
1st person plu - 2nd person sing == 0	-0.22	0.48	-0.46	0.99	
2nd person plu - 2nd person sing == 0	-0.89	0.42	-2.16	0.24	
2nd person plu - 1st person plu == 0	-0.68	0.46	-1.49	0.65	

The results here indicate that the biggest contrast for this factor group is 2nd person plural possessors vs 3rd person (singular and plural) and 1st person singular possessors. 2nd person plural disfavour colloquial variants overall. All other results indicate that the difference in likelihood of any paired comparison of different factor levels on the intercept are not significantly different from zero.

Next, in order to present the hierarchies within each significant predictor, I present a log-odds graph (Figure 5.7) where zero on the x-axis represents the likelihood of the colloquial variant occurring, when all predictors are set to their reference values. Recall that all reference

values from the models presented so far were reordered to the most favouring values, so all values appearing in the plot are below zero; that is, they all disfavour the colloquial compared to the reference value. Each factor has been grouped by colour in the log-odds graph, and p values are presented along with confidence intervals (in horizontal error bars) in order to display with what certainty any significant factors can predict the use of the colloquial variant.

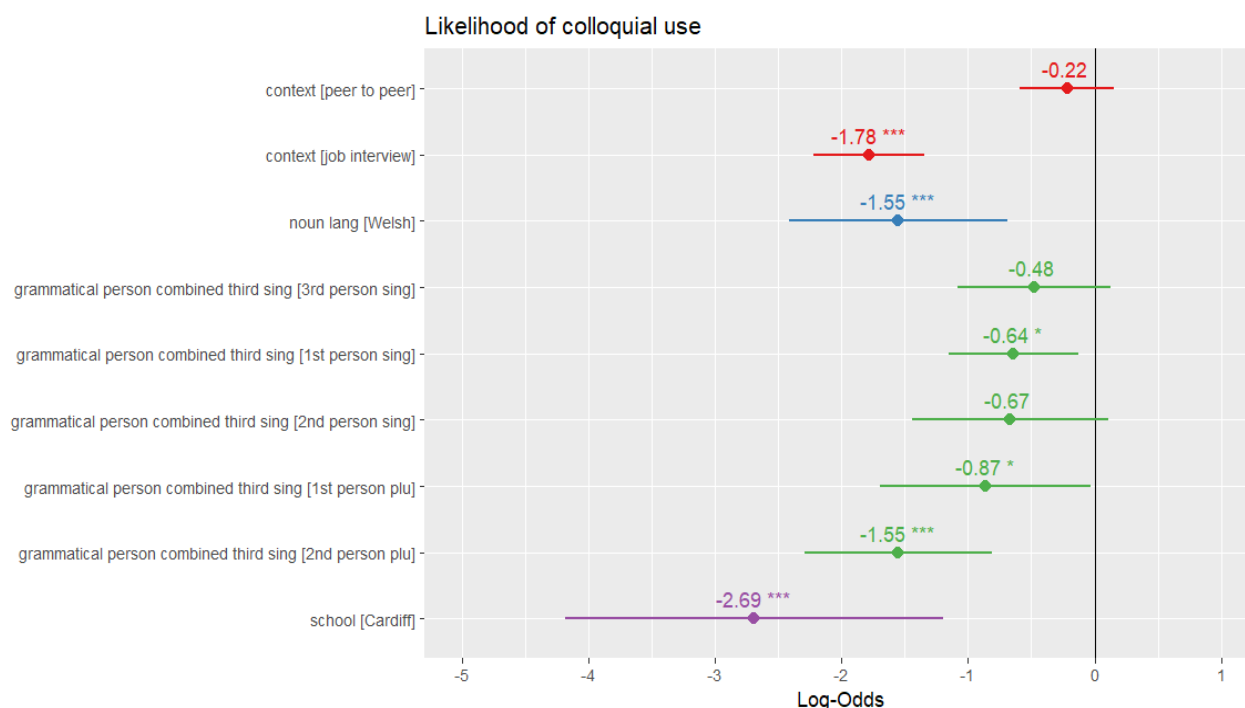


Figure 5.7: Log-odds likelihood of colloquial use

Any predictor levels with error bars overlapping the zero line (for example peer to peer and 3rd person singular), are not significantly different from the reference level of that predictor. Cardiff students are 2.69 times less likely to use the colloquial variant than Gwynedd students. Similar strong effects can also be seen for the job interview context, English nouns and some grammatical person constructions.

5.6.4 Summary

We have seen, so far, that context is the strongest predictor of colloquial pronoun use. Statistically significant differences were found between the formal job interview (which

contained the fewest colloquial variants) and the informal sociolinguistic interview (which contained the most colloquial variants). The language of the nouns had the second strongest effect, with English nouns being used significantly more with colloquial variants than Welsh nouns. Grammatical person was the third strongest predictor, with 2nd person plural possessors least likely to be used with the colloquial variant, compared to 3rd person and 1st person singular possessors, which were more likely to be colloquial. Recall that rates of 2nd person plural were low, particularly for the GEHL group ($n = 5$); this is likely to have an effect on these significant findings. Finally, Cardiff speakers were significantly less likely to use the colloquial variant than Gwynedd speakers. Although home language was not found to be a significant predictor of colloquial variant use, recall that interactions containing home language in the maximal model were found to be significant, showing that home language and region effects are likely to be linked (see page 129). Analyses of each of the four speaker groups will confirm these findings.

Next I will compare the four speaker groups, by running a statistical analysis for each group. This will allow me to examine the differences between the groups, in order to ascertain a) the patterns of variation in WHL speakers, and b) the extent to which the EHL speakers replicate these patterns. The subset models include the predictors outlined in the maximal model. The contrast matrix presented above showed that 2nd person plural was less likely to be used with the colloquial variant than 1st person and 3rd person possessors. When used in the four subset models grammatical person caused convergence problems (likely due to low tokens). The only other variationist work on the possessive pronoun feature (Davies 2016), did not consider person-number combinations other than 3rd person singular and 1st person plural possessors. However, I also found the high rates of colloquial use in 1st person plural and 3rd person singular possessors. In order to make comparisons between the current work and this previous finding, and to examine the patterning of these constructions against all others, 3rd

person singular and 1st person plural possessors were combined and set as the reference level for the grammatical person factor. In this way, it will be possible to see patterns in WHL use, and whether these are replicated by EHL speakers. This also helped account for very low token counts for the numerous levels of grammatical person in each subset model.

5.6.5 Gwynedd Welsh home language subset

Table 5.26 shows the mixed effects analysis for the GWHL subset speaker group.³¹

Table 5.26: Subset mixed effects model for GWHL

AIC: 497.5						Observations	
Fixed effects:	Estimate	Std. Error	z-value	p-value	Sig	n	% coll.
(Intercept)	3.56	1.30	2.73	0.006	**		
CONTEXT (<i>vs sociolinguistic interview</i>)						252	74
Peer to peer	-0.19	0.34	-0.57	0.57		218	68
Job interview	-0.41	0.37	-1.10	0.27		116	47
NOUN LANGUAGE (<i>vs English</i>)						51	86
Welsh	-0.92	0.66	-1.39	0.16		535	64
GRAMMATICAL PERSON (<i>vs 1st person plural and 3rd person singular</i>)						105	76
other	-0.63	0.36	-1.75	0.08		447	69
ALIENABILITY (<i>vs Alienable</i>)						204	62
Inalienable	-0.01	0.35	-0.03	0.97		382	69
FREQUENCY (<i>vs frequent</i>)						276	67
Infrequent	-0.24	0.35	-0.68	0.50		310	65
Random effects:						sd	n
SPEAKER						2.23	5
NOUN						1.09	181

There are no significant results in this model; none of the levels within each factor were significant predictors of the use of possessive pronouns in the current dataset. The estimate result for context, however, shows the direction of the effect; the colloquial form is used least in the job interview context, followed by the peer-to-peer context, followed by the sociolinguistic interview. This speaker group were also likely to use more formal variants

³¹ Factor level ordering were kept the same for all subsequent subsets as in the maximal model in order to allow for an easier comparison across groups.

(sandwich or literary) where the possessed noun was Welsh, and more colloquial variants where the possessed noun was English, though the effect is not significant. Similarly, grammatical person constructions other than 3rd person singular and 1st person plural were mostly used with more formal variants, as was the case for inalienable and infrequent nouns.

The size of the χ^2 statistic in a Wald χ^2 test can reveal the order of magnitude of effect. Such a test was carried out on each speaker group subset model, in order to compare the strength of each predictor between the groups. The results are presented in Table 5.27.

Table 5.27: Wald χ^2 test for GWHL colloquial variant use

	Wald χ^2	Df	<i>p</i> -value
Grammatical person	3.06	1	0.08
Noun language	1.95	1	0.16
Context	1.21	2	0.54
Frequency	0.46	1	0.50
Alienability	0.001	1	0.98

The Wald χ^2 test for GWHL revealed that grammatical person was the strongest factor influencing colloquial variant use ($\chi^2=3.06$, *df*= 1, *p*-value= 0.08), followed by noun language and context. None of these factors can be considered predictors, however, as they cannot significantly predict the use of the colloquial pronoun variant. For this reason, a more parsimonious model cannot be produced. I now turn to analyse the second subset group.

5.6.6 Gwynedd English home language subset

Table 5.28 presents the analysis of the GEHL subset group.

Table 5.28: Subset mixed effects model for GEHL

	Estimate	Std. Error	<i>z</i> -value	<i>p</i> -value	Sig	Observations	
						<i>n</i>	% coll.
AIC: 132.4							
Fixed effects:							
(Intercept)	9.24	3.39	2.72	0.006	**		
CONTEXT (<i>vs sociolinguistic interview</i>)						148	98
Peer-to-peer	-3.04	1.3	-2.34	0.02	*	61	80
Job interview	-2.73	1.49	-1.83	0.06		44	84

NOUN LANGUAGE (<i>vs English</i>)					23	91
Welsh	-1.11	2.26	-0.49	0.62	230	91
GRAMMATICAL PERSON (<i>vs 1st person plural and 3rd person singular</i>)					62	88
other	0.46	1.17	0.39	0.69	191	92
ALIENABILITY (<i>vs Alienable</i>)					204	61
Inalienable	1.67	1.69	0.98	0.32	382	69
FREQUENCY (<i>vs frequent</i>)					165	94
Infrequent	-1.71	1.68	-1.02	0.30	88	86
Random effects:					<i>sd</i>	<i>n</i>
SPEAKER					0.00	4
NOUN					6.93	84

Participants in Gwynedd who spoke English at home used the colloquial variant far more in all contexts than their Welsh home language counterparts. For this group, context was a significant predictor of colloquial pronoun use. However, differently to their WHL counterparts, this group used the fewest colloquial variants in the peer-to-peer setting, which was significantly different to their use in the sociolinguistic interview.

The Wald χ^2 test for GEHL revealed that context was the strongest predictor (though not to a significant extent) of colloquial variant use ($\chi^2 = 5.49$, $df = 2$, p -value = 0.06), followed by frequency and alienability. Noun language and grammatical person were the weakest predictors for this group; likely because of low token counts for this final predictor, as discussed previously. The results are presented in Table 5.29 below. Due to a lack of significance, a more parsimonious model cannot be produced for this group.

Table 5.29: Wald χ^2 test for GEHL colloquial variant use

	Wald χ^2	Df	<i>p</i> -value
Context	5.49	2	0.06
Frequency	1.04	1	0.30
Alienability	0.97	1	0.32
Noun language	0.24	1	0.62
Grammatical person	0.15	1	0.69

5.6.7 Cardiff Welsh home language subset

The subset model for CWHL is presented in Table 5.30 below.

Table 5.30: Subset mixed effects model for CWHL

AIC: 554.2						Observations	
Fixed effects:	Estimate	Std. Error	z-value	p-value	Sig	n	% coll.
(Intercept)	2.23	1.17	1.91	0.05			
CONTEXT (<i>vs sociolinguistic interview</i>)						215	56
Peer to peer	-0.30	0.33	-0.89	0.37		238	53
Job interview	-3.59	0.49	-7.21	<0.001	***	155	6
NOUN LANGUAGE (<i>vs English</i>)						19	79
Welsh	-1.69	0.87	-1.95	0.05		589	41
GRAMMATICAL PERSON (<i>vs 1st person plural and 3rd person singular</i>)						102	45
other	-0.36	0.35	-1.02	0.31		506	41
ALIENABILITY (<i>vs Alienable</i>)						184	45
Inalienable	-0.58	0.39	-1.47	0.14		424	40
FREQUENCY (<i>vs frequent</i>)						375	42
Infrequent	-0.09	0.39	-0.25	0.80		233	41
Random effects:						sd	n
SPEAKER						1.29	4
NOUN						1.32	157

For CWHL (similarly to GEHL), context was a significant predictor of colloquial variant use. However, the direction of the effect is more similar to GWHL; the more formal job interview is less likely to contain the colloquial variant than the peer-to-peer setting. Recall that for GEHL, the peer-to-peer setting contained significantly fewer instances of the variant. On closer inspection of the model output, it can be seen that the pattern of effect is the same for all the factor levels as for the previous WHL group in Gwynedd. Infrequent, inalienable, Welsh nouns are less likely to be used with the colloquial, and constructions other than 1st person plural and 3rd person singular are less likely to be used with the colloquial too.

The Wald χ^2 test for CWHL shows that context is the strongest predictor of colloquial variant use for CWHL, followed by noun language, alienability, grammatical person and frequency (see Table 5.31).

Table 5.31: Wald χ^2 test for CWHL colloquial variant use

	Wald χ^2	Df	p -value	
Context	57.55	2	<0.001	***
Noun language	3.82	1	0.05	
Alienability	2.15	1	0.14	
Grammatical				
person	1.03	1	0.31	
Frequency	0.06	1	0.80	

A parsimonious model for CWHL was run with context (as this was the only significant predictor) and is presented in Table 5.32. This model confirms the results from the full subset model for CWHL; that is, the job interview is significantly less likely to contain the informal pronoun variant.

Table 5.32: Parsimonious model for CWHL subset

AIC: 554.2						Observations	
Fixed effects:	Estimate	Std. Error	z-value	p-value	Sig	<i>n</i>	% coll.
(Intercept)	-0.09	0.68	-0.13	0.89			
CONTEXT (<i>vs sociolinguistic interview</i>)						215	56
Peer to peer	-0.19	0.33	-0.57	0.56		238	53
Job interview	-3.62	0.50	-7.22	<0.001	***	155	6
Random effects:						<i>sd</i>	<i>n</i>
SPEAKER						1.26	4
NOUN						1.41	157

5.6.8 Cardiff English home language subset

The mixed effects model for CEHL is presented in Table 5.33 below.

Table 5.33: Subset mixed effects model for CEHL

AIC: 560.7						Observations	
Fixed effects:	Estimate	Std. Error	z-value	p-value	Sig	<i>n</i>	% coll.
(Intercept)	3.46	0.98	3.50	<0.001	***		
CONTEXT (<i>vs sociolinguistic interview</i>)						206	75
Peer to peer	-0.81	0.28	-2.86	0.004	**	197	50
Job interview	-2.47	0.39	-6.18	<0.001	***	86	17
NOUN LANGUAGE (<i>vs English</i>)						19	89
Welsh	-2.55	0.89	-2.86	0.004	**	470	53

GRAMMATICAL PERSON (<i>vs 1st person plural and 3rd person singular</i>)					95	61
other	0.04	0.32	0.14	0.89	394	53
ALIENABILITY (<i>vs Alienable</i>)					148	47
Inalienable	0.34	0.32	1.06	0.28	341	58
FREQUENCY (<i>vs frequent</i>)					292	60
Infrequent	-0.68	0.31	-2.15	0.03 *	197	47
Random effects:					<i>sd</i>	<i>n</i>
SPEAKER					0.52	4
NOUN					0.88	161

As with the CWHL group, CEHL speakers were significantly more likely to use colloquial variants in the sociolinguistic interview compared to the job interview. This was also the case for the peer-to-peer context, which contained significantly fewer colloquial variant pronouns than the sociolinguistic interview. For this group, noun language reached significance at the $p = 0.05$ level ($p = 0.004$), whereas for the CWHL group, it did not. For both, however, the direction of the effect is the same; Welsh nouns are less likely to be used with the colloquial variant. For CEHL (and this group alone), frequency was also a significant predictor of colloquial use; infrequent nouns were less likely to be used with the colloquial variant. Recall that for the WHL groups, all estimate results were negative. That is, the factor level presented was less likely to occur with the colloquial than it was for the reference level. On the other hand, a pattern emerges with the EHL speakers. Both grammatical person and alienability incur positive estimates, showing that for both groups, 1st person plural and 3rd person singular constructions and inalienable constructions were more likely to occur (though not to a significant extent) with the colloquial pronoun than the reference level.

In order to compare the strength of the effect across factors, the Wald χ^2 test results are presented in Table 5.34. For CEHL, context has the strongest effect, followed by noun language (and both were significant). This was the same as for the CWHL group. Frequency followed, and also came out as significant. Alienability and grammatical person had the weakest effect.

Table 5.34: Wald χ^2 test for CEHL colloquial variant use

	Wald χ^2	Df	p -value	
Context	38.25	2	<0.001	***
Noun language	8.18	1	0.04	**
Frequency	4.61	1	0.03	*
Alienability	1.13	1	0.28	
Grammatical person	0.02	1	0.89	

A parsimonious model with only the significant predictors from the Wald χ^2 test is presented in Table 5.35. Only job interview and Welsh nouns were found to be significant predictors of less colloquial use; we see differences between these findings and the full subset model in Table 5.33. This means that there is insufficient statistical validation in the summary statistics (Gardner 2023) for the previously observed differences between frequent and infrequent nouns and different speech contexts.

Table 5.35: Parsimonious model for CEHL subset

AIC: 553.5						Observations	
Fixed effects:	Estimate	Std. Error	z-value	p -value	Sig	n	% coll.
(Intercept)	1.69	1.12	1.50	0.132			
CONTEXT (<i>vs sociolinguistic interview</i>)						206	75
Peer to peer	-0.24	0.33	-0.72	0.47		197	50
Job interview	-3.56	0.49	-7.17	<0.001	***	86	17
NOUN LANGUAGE (<i>vs English</i>)						19	89
Welsh	-1.81	0.87	-2.06	0.04	*	470	53
FREQUENCY (<i>vs frequent</i>)						292	60
Infrequent	-0.11	0.39	-2.88	0.77		197	47
Random effects:						sd	n
SPEAKER						1.25	4
NOUN						1.36	157

5.6.9 Hierarchies for the subset groups

So, having considered all the analyses of the subset groups, what are the main patterns that emerge? Figure 5.8 below presents the log-odds (likelihood of colloquial variant use) for the

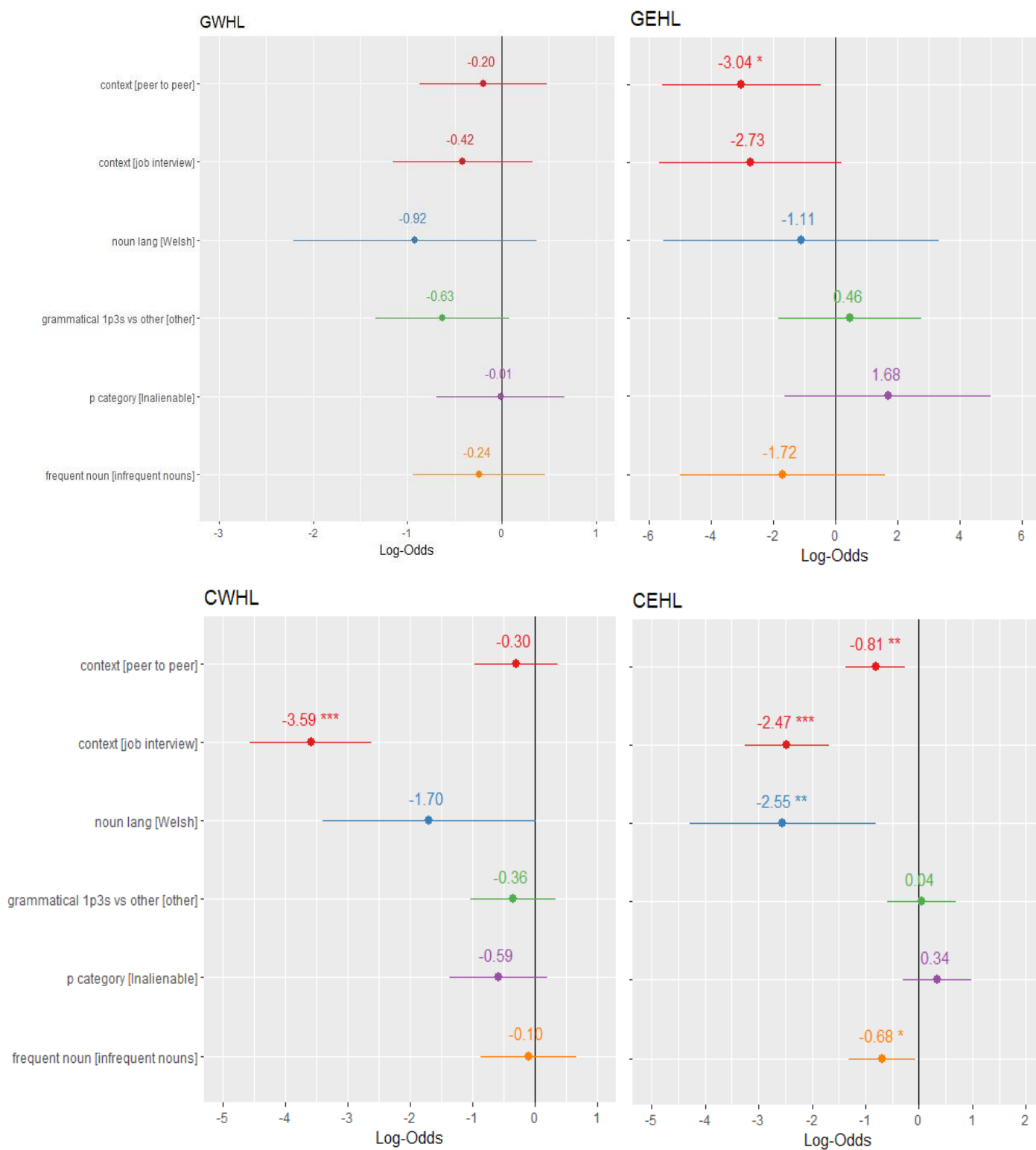


Figure 5.8: Log-odds likelihood of colloquial pronoun use across four speaker groups

subset models for all four speaker groups, in order to compare the hierarchies of each predictor under investigation. Note that the full subset models are included here, rather than the more parsimonious models. As stated in section 4.7.3, this work is exploratory in nature, and although it is useful to know to what extent there is strong predicting power in the factors under analysis, it was deemed more important to review the main patterns, which take into account all the factor levels for the three lines of evidence.

The line indicating zero in Figure 5.8 represents the estimate for the missing predictor level (namely the sociolinguistic interview, Welsh nouns, the 1st person plural and 3rd person singular grammatical person construction, alienable nouns and infrequent nouns). The position of the other predictor levels shows whether they are more or less likely to use colloquial variants than the reference (missing) level. Note that the x-axis for each varies slightly to account for the scaled log likelihood for each group. Any error bars crossing the zero line (the intercept) are not significant, as is the case for all predictors in the GWHL figure.

Some clear patterns can be observed in comparing the four groups. For all groups, the sociolinguistic interview elicited the highest rates of the colloquial variant. In Cardiff, the use of the colloquial was significantly lower in the job interview contexts, whereas for GEHL speakers, the peer-to-peer context elicited significantly fewer cases of colloquial pronouns. Although for all groups Welsh nouns occurred less frequently with colloquial variants, this was only significant in CEHL speakers (though close to significance level for CWHL). Similarly, infrequent nouns elicited fewer colloquial variants than frequent ones, but this was not statistically significant³² for any groups.

For WHL speakers, ‘other’ grammatical person and inalienable nouns were less likely to be used with the colloquial variant, but the opposite can be observed in the EHL speakers.

³² The parsimonious model for the CEHL subgroup showed that the significant finding for frequency was likely to be invalid.

Instead, those who spoke English at home were more likely to use the colloquial pronoun with alienable nouns and in 1st person plural and 3rd person singular constructions. Although there is a suggestion in this analysis that these linguistic constraints have not been acquired by EHL students in the two regions, this should be read with caution, due to the distinct lack of statistical significance in the WHL group.

Next, I present a discussion of the findings presented in this analysis of colloquial pronoun use in Welsh.

5.7 Discussion

In this chapter, I have used regression analyses to provide statistical validation for the central research question of the thesis which is to confirm whether students who do not speak Welsh at home are acquiring the same patterns observed in the speech of those who do speak Welsh at home, in two distinct areas. The “three lines of evidence” (see Poplack and Tagliamonte 2001; Tagliamonte 2002; Tagliamonte 2006) are referred to by Gardner (2023) as ‘the key facts’ required to answer this central question. Each line of evidence will be presented below.

5.7.1 Line of evidence 1: Statistical significance

The first line of evidence tells us which independent variables are significant predictors of the variation and which are not. These results from the most parsimonious subset models are presented in Table 5.36. The first line of evidence needs to be based on the more parsimonious models because they contain the statistical power of each factor (unlike the third line of evidence, below, which assess the hierarchies). Only two factors were significant predictors of colloquial possessive pronoun use: context and noun language. Gwynedd WHL had no significant predictors, which was replicated by GEHL in all cases. In Cardiff, the sociolinguistic interview context elicited significantly more colloquial variants than the job interview for CWHL; a pattern that was replicated by the EHL counterparts. CEHL replicated

all CWHL patterns of non-significance, apart from the noun language factor. The former group was significantly less likely to use the colloquial variant with Welsh nouns. For the CWHL group, this was not found to a significant extent.

On the whole, EHL participants are replicating most of the patterns observed in WHL speakers.

Table 5.36: Line of evidence 1 (possessives)

	Gwynedd		Cardiff	
	WHL	EHL	WHL	EHL
Context	No	No	Yes	Yes
Noun language	No	No	No	Yes
Grammatical person	No	No	No	No
Alienability	No	No	No	No
Frequency	No	No	No	No

5.7.2 Line of evidence 2: Strength of effect

The second line of evidence is the relative ordering of predictors, based on their magnitude of effect on the variation, deduced from the Wald χ^2 tests conducted for each subset model (Table 5.37). The strongest effects are at the top of the table and are shown in decreasing order. The descending WHL pattern for both regions has been highlighted (darker grey for stronger effect, lighter grey for weaker effect), in order to demonstrate whether the EHL speakers have acquired the same patterns as their local counterparts.

Table 5.37: Line of evidence 2 (possessives)

	Gwynedd				Cardiff			
Strength of effect	WHL		EHL		WHL		EHL	
↓	Grammatical person	3.06	Context	5.49	Context	57.55	Context	38.25
	Noun language	1.95	Frequency	1.04	Noun language	3.82	Noun language	8.18
	Context	1.21	Alienability	0.97	Alienability	2.15	Frequency	4.61
	Frequency	0.46	Noun language	0.24	Grammatical person	1.03	Alienability	1.13
	Alienability	0.001	Grammatical person	0.15	Frequency	0.06	Grammatical person	0.02

Starting with the Cardiff groups, context is a clear constraint on the use of colloquial pronouns. A high chi-square statistic shows that this is a similarly strong effect for both WHL and EHL students. Noun language has the second largest effect (and explanatory power for patterns of variation) for both groups, though a larger chi-square statistic for CEHL than CWHL. Patterns diverge from the 3rd strongest effect; though frequency ($\chi^2 = 4.61$) is relatively strong for CEHL, for CWHL, this has the smallest effect ($\chi^2 = 0.06$). Similarly, the magnitude of alienability and grammatical person do not match exactly, but for both groups the effect is weak for the two predictors.

The story in Gwynedd is quite different. Firstly, even the largest effect for WHL is relatively small ($\chi^2 = 3.06$ in GWHL speakers). Although GEHL shares the pattern with Cardiff counterparts in terms of being constrained by context, this is to a much smaller extent ($\chi^2 =$

147

5.49). As a consequence, all the following factors have very little power to explain the variation observed in possessive pronoun use (hence the lack of significance in the first line of evidence, see Table 5.36). It appears that frequency has a stronger effect on EHL speakers than on WHL speakers.

There is insufficient evidence from the second line of evidence to tell whether GEHL students have acquired the same constraints as GWHL. This could relate to the much higher rates of the colloquial form in Gwynedd. In Cardiff, however, EHL students are mostly replicating their WHL counterparts' patterns.

5.7.3 Line of evidence 3: Constraint hierarchy

The third line of evidence is concerned with the constraint hierarchy (or the direction of the effect). Here (in Table 5.38) I present the relative ordering of levels of a predictor, from the level with the greatest positive effect to the level with the greatest negative effect on the overall probability of using the colloquial pronoun variant. The darker shaded cells indicate where the EHL speakers have not acquired the same constraint hierarchy as their WHL counterparts.

Table 5.38: Line of evidence 3 (possessives)

	Gwynedd		Cardiff	
	WHL	EHL	WHL	EHL
Context	Sociolinguistic interview > peer to peer > job interview	Sociolinguistic interview > job interview > peer to peer	Sociolinguistic interview > peer to peer > job interview	Sociolinguistic interview > peer to peer > job interview
Noun language	Welsh nouns less colloquial	Welsh nouns less colloquial	Welsh nouns less colloquial	Welsh nouns less colloquial

Grammatical person	1 st person plural, 3 rd person singular less colloquial	1 st person plural, 3 rd person singular more colloquial	1 st person plural, 3 rd person singular less colloquial	1 st person plural, 3 rd person singular more colloquial
Alienability	Inalienable nouns less colloquial	Inalienable nouns more colloquial	Inalienable nouns less colloquial	Inalienable nouns more colloquial
Frequency	Infrequent nouns less colloquial	Infrequent nouns less colloquial	Infrequent nouns less colloquial	Infrequent nouns less colloquial

Where nouns are Welsh and infrequent, less instances of the colloquial pronoun variant were found; this was the case for WHL and EHL in both areas. However, the other two internal, linguistic constraints (grammatical person and alienability) are not replicated by the EHL groups in either Gwynedd or Cardiff. Although there is a clear trend in all groups that the sociolinguistic interview contained the highest rates of colloquial variants, in Gwynedd, EHL students have not replicated the exact pattern of WHL speakers. For the former group, the lowest rates of colloquial pronoun use are in the peer-to-peer context, not the job interview context. In Cardiff, however, EHL students follow the same patterns as WHL in terms of context.

Therefore, across all three lines of evidence, it appears that EHL students stick to the WHL norms more closely in Cardiff than in Gwynedd, which points to them having acquired more sociolinguistic competence. Noun language and frequency constrained the use of possessive pronouns in the same way in both regions. Grammatical person and alienability

patterns were not replicated by either EHL group, although these patterns were not strong in WHL groups. GEHL used most formal pronoun variants in the peer-to-peer setting, which could reflect the ‘unnatural’ nature of the task for a naturally English-speaking friendship group. The highest rates of colloquial pronouns were found in the sociolinguistic interview, however, as was the case with GWHL. Sociolinguistic competence has been acquired in both areas, but patterns are strongest in Cardiff, as demonstrated by the similarities in the two home language groups.

5.7.4 Insights to the feature beyond sociolinguistic competence

The above analysis revealed the following points about possessive pronoun use by Welsh immersion pupils:

- The colloquial variant is the most frequent possessive pronoun variant, for all speaker groups. The high rates support previous research which shows that possessive constructions are undergoing change towards the colloquial construction (M. C. Jones 1998) which is nearing completion (Davies 2016).
- GwyneddM. C. Jones (1998) found the opposite in her exploration of two sociolinguistically contrasting areas. In her work on the Gwenhwyseg dialect (where the community language was mostly English) and the Rhosllanerchrugog dialect (where the community language was mostly Welsh), the former cohort of participants used a higher proportion of the colloquial variant, than the latter participants, although the token count here is small. Further evidence would be required to assess where the speakers of this peripheral Gwynedd dialect are in relation to the wider process of language change towards the colloquial variant in younger speakers.
- All speaker groups showed elements of style-shifting between speech contexts, and the Cardiff groups showed a statistically significant different use in contexts.

- The colloquial variant is reported to be considered non-standard and has previously been stigmatised as a feature of learner speech. Despite its stigmatised status, it is clearly a part of the Welsh possessive pronoun system, and Learn Welsh (previously called Welsh for Adults) and other second language learning providers could consider introducing the variant to beginner learners of Welsh, to help promote the appropriate use of social Welsh.
- GEHL groups' style-shifting did not pattern onto their local peers' hierarchy. The differences in hierarchy between GWHL and GEHL speakers could point to the perceived formality of the peer-to-peer setting for EHL groups. Some students in the sociolinguistic interviews pointed out that although they undertook the peer-to-peer task in their group through the medium of Welsh (as the task required), they would naturally speak English with their peers. Thus, these speakers could perceive the group task as being 'artificial' or school-based, which could increase the formality of the possessive variant use.
- Where the language of the noun is Welsh, all groups were more likely to use a more formal variant, and where the noun was English higher rates of the colloquial variant were observed. This was consistent with expectations outlined in 5.3.2.

5.8 Conclusion

The analysis of context showed that the use of the possessive variants changed according to the speech context; more literary and sandwich constructions were used in tasks aimed at eliciting more formal language. In keeping with recent phonological work on Welsh in immersion settings (Gruffydd 2022), these results show that the standardization of the colloquial variant does not lead to less stylistic variation in Welsh. Further, these findings clearly demonstrate that pupils in Welsh education do have stylistic repertoires, regardless of the language they speak at home. However, they are not always using their informal styles with

their peers, possibly because they do not usually use Welsh at all with their peers. The data suggests that the GEHL group may have viewed the peer-to-peer setting as rather more formal than their WHL counterparts. The next two features in chapters 6 and 7 will shed further light on this, and it will form part of the discussion chapter in section 8.3.

Some internal WHL patterns were acquired by EHL students; namely noun language constraints, however, the extent to which they were acquired depended on the area. Indeed, regional variation has been found to be a significant factor in the production and variation of the possessive pronoun construction in Welsh, highlighting the importance of community-specific studies, as found in Morris (2021; 2022).

6 Inflected and Periphrastic verbs

6.1 Introduction

This chapter presents an analysis of variation in the Simple Past tense in Welsh. Verbs in this tense can be formed in two ways; the first is an inflected construction, achieved by adding a set of endings to the stem of the verb (King 2014: 69) e.g. *cysgais i* [I slept], and the second is a periphrastic construction using an auxiliary with a verb noun e.g. *gwnes i gysgu* [I did sleep]. As shown in the squared bracket translations, English also has a dual system for forming the past tense, which in Standard English serves an emphatic purpose. In Welsh, however, the periphrastic variant serves no emphatic purpose; indeed, in Welsh, the two forms are thought to be “entirely interchangeable” when used in spoken contexts (King 2016: 232). It has been suggested that the inflected form is more appropriate in formal written contexts (B. M. Jones 1993:157) and the standard spoken Welsh heard in the media (King 2016: 226). The periphrastic variant, on the other hand, is considered more informal and is also thought to be “spreading [...] in the modern language” (Borsley *et al.* 2007: 12). Some note that the periphrastic is particularly prevalent in northern dialects (B. M. Jones 1993: 157), however no work has been undertaken to substantiate this claim. Furthermore, little is known about how the two variants pattern in different speech settings and no work has examined the production of the periphrastic variants according to sociolinguistic factors like home language and contexts of use.

In this chapter, I will introduce the limited previous research on the feature, outline the methods used for extracting and analysing the feature, and then present the results of the distributional and inferential analysis, followed by a discussion of the findings.

6.2 The feature

6.2.1 The Simple Past tense

Verbs in the Simple Past tense are used to refer to a condition or an action that happened once or multiple times in the past (Thomas 1996: 105; Borsley *et al.* 2007). More specifically, it refers to:

1. a condition or specific action in the past;
2. an individual action that took place one time in the past;
3. an individual action that took place a number of times in the past;
4. a continuous action that finished in the past.

Thomas (1996: 105) provides examples of the above, and in this work he also makes a distinction between the Simple Past and the Specific Past, which uses *bod* [to be] to express a condition or an action which last for a period in the past and came to an end in the past, e.g. *bu'n poeni am ganlyniadau'r prawf* [he/she was worried about the results of the test], *buom yn hir bendroni cyn penderfynu* [we puzzled over it for a long time before deciding], *buont yn gweithio ar y tir yn ystod y rhyfel* [they were working on the land during the war] (Thomas 1996: 106). The Simple and Specific Tense form part of the preterite tense in Welsh. The continuous nature of the Specific Past makes it closer to an imperfect tense than the Simple Past, and thus it is not included in the analysis. Whereas the Specific Past is always periphrastic, the Simple Past can be used in periphrastic or inflected constructions, explained below.

6.2.2 Inflected verbs

Verbs in the Simple Past can be inflected in Welsh by converting the infinitive verb³³ to a stem and attaching the appropriate ending (King 2016: 168).

³³ Known in Welsh as the verb-noun as it has both the properties of a verb and a noun (Borsley *et al.* 2007: 68)

Table 6.1 shows the inflection of the regular infinitive, or verb-noun, *cysgu* [to sleep] (root = *cysg-*) in the tense under study. The examples in Table 6.1 contain the pronoun relevant to the subject which in spoken Welsh is usually used in conjunction with the inflected verb (Borsley *et al.* 2007: 34). The pronoun appears in all examples from hereon in, though it is recognised that it is optional, and may be omitted, particularly in written contexts.

Table 6.1: Inflected verb constructions by subject

Subject	Simple Past inflection
1 st person sing.	cysgais i
2 nd person sing.	cysgaist ti
3 rd person sing.	cysgodd e/hi
1 st person plural	cysgon ni
2 nd person plural	cysgoch chi
3 rd person plural	cysgon nhw

Inflected verb endings are consistent for all Simple Past tense regular verbs after isolating the stem, e.g. *bwytais i* [I ate], *rhedaist ti* [you ran], *eisteddon ni* [we sat]. Irregular verbs, on the other hand, follow a similar inflection pattern for each subject but the verb root changes. The most common irregular verbs in Welsh are *mynd* [to go], *gwneud* [to do/make], *dod* [to come] and *cael* [to have/get]. Other irregular verbs also exist, such as *adnabod* [to know someone] and *gwybod* [to know something]. Whereas English does not distinguish the past tense use between active verbs and stative verbs (expressing mental states), e.g. she ran (action) *vs* she knew (mental state), in Welsh, non-active verbs (e.g. *adnabod* [to know], *meddwl* [to think], *hoffi* [to like] etc.) are usually expressed using the imperfect tense instead (King 2016: 234), because in their Simple Past tense, the meaning is inceptive. As stated previously, the Specific Past is expressed using the non-active verb *bod* [to be]. The imperfect tense and Specific Past

tense are not covered in this thesis; the verbs discussed herein are Simple Past, active verbs.

Table 6.2 below shows the most common irregular active verbs in all subject constructions.

Table 6.2: Irregular verb constructions (past) in spoken Welsh by subject

	Mynd [to go]	Gwneud [to do]	Dod [to come]	Cael [to have/get] ³⁴
1 st person sing.	es i [I went]	gwnes i [I did]	des i [I came]	ces i [I had]
2 nd person sing.	est ti	gwnest ti	dest ti	cest ti
3 rd person sing.	aeth o/e/hi	gwnaeth o/e/hi	daeth o/e/hi	caeth o/e/hi
1 st person plural	aethon ni	gwnaethon ni	daethon ni	caethon ni
2 nd person plural	aethoch chi	gwnaethon ni	daethon ni	caethon ni
3 rd person plural	aethon nhw	gwnaethon nhw	daethon nhw	caethon nhw

Mutation constraints apply to inflected verbs in some cases. In negative phrases, the inflected verb takes a soft (e.g. *welais i ddim* [I didn't see]) or an aspirate mutation (e.g. *chysgon ni ddim* [we didn't sleep]), depending on the ability of the consonant to take a mutation. King (2016: 227) purports that the soft mutation is most frequent, however, the extent to which this is done in varieties of spoken Welsh is unknown. In formal written (literary) Welsh, where an interrogative particle *a* precedes the inflected verb, a soft mutation would normally occur e.g. *a gysgodd hi?* [did she sleep?]. In many varieties of spoken Welsh, the interrogative particle is mostly lost, but in some cases the mutation remains (Borsley *et al.* 2007: 36). The current work will seek to assess the extent to which the mutation constraints are met in inflected verbs.

³⁴ Cael does not always pattern like *mynd*, *gwneud* and *dod*; in some parts of Wales, 3rd person singular *cafodd*, 1st person plural *cafon* or *cawson*, 2nd person plural *cafoch* or *cawsoch*, 3rd person plural *cafon* or *cawon* are also spoken forms (King 2016: 228).

6.2.3 Periphrastic verbs

Periphrastic constructions are an alternative formation to the inflected verbs (King 2018: 29) presented above. Periphrastic constructions contain an initial auxiliary verb, with the infinitive verb in post-subject position (Borsley *et al.* 2007: 12). Two auxiliaries are possible in past periphrastic constructions, that is *gwneud* [to do] and *darfu* [to happen], which now appears in modern Welsh as *ddaru*³⁵. *Gwneud* as an auxiliary is widely used in all dialects, but *ddaru* (and sometimes ‘*aru*’) is restricted to northern dialects of Welsh (King 2016: 233). Where the auxiliary *gwneud* is used, it is inflected for person and number with a pronominal subject (as per Table 6.2), whereas the auxiliary *ddaru* remains uninflected. Indeed, *ddaru* has no personal suffixes and cannot be inflected in any tense (Thomas 2012: 124). In many learner materials, *gwnes i* [I did] undergoes a soft mutation, becoming *wnes i*. In all examples presented in this chapter, however, the radical form is used. Periphrastic verb constructions with both *gwneud* and *ddaru* auxiliaries and the infinitive verb-noun *cysgu* [to sleep] can be seen in Table 6.3 below.

Table 6.3: Periphrastic verb constructions by subject

	<i>gwneud</i>	<i>ddaru</i>	English gloss
1 st person singular	<i>gwnes i gysgu</i>	<i>ddaru mi gysgu</i>	I slept
2 nd person singular	<i>gwnest ti gysgu</i>	<i>ddaru ti/chdi gysgu</i>	You slept
3 rd person singular	<i>gwnaeth o/e/hi gysgu</i>	<i>ddaru fo/hi gysgu</i>	He/she slept
1 st person plural	<i>gwnaethon ni gysgu</i>	<i>ddaru ni gysgu</i>	We slept
2 nd person plural	<i>gwnaethoch chi gysgu</i>	<i>ddaru chi gysgu</i>	You slept
3 rd person plural	<i>gwnaethon nhw gysgu</i>	<i>ddaru nhw gysgu</i>	They slept

³⁵ The *ddaru* construction originally took a preposition e.g. *darfu i mi (fynd)* [It happened to me (to go)], which is generally no longer part of the construction (Rottet and Morris 2018: 171).

Where the infinitive verb-noun in a periphrastic construction is *gnweud* [to do], the auxiliary can also be *gwneud*, i.e. *gwnes i wneud* [I did]. As mentioned previously, the periphrastic and inflected variants are variants of the same Simple Past variable; although the auxiliary functions as an emphasis in English periphrastic constructions, emphasis in Welsh tends to be created by (but not limited to) using an affirmative particle such as *fe/mi* before the auxiliary (Thomas 2012: 86). Emphatic phrases are not considered in the current discussion of this feature.

6.2.4 *An historical perspective*

The classical tendency of inflecting verbs in Latin is said to have influenced the development of inflected verbs in Welsh (B. M. Jones 1993: 155), as has been found in the development of the inflected future tense in French (Tristram 2021: 26). However, little is documented about the historical development of inflected and periphrastic variants in Welsh.

Some research discusses the grammaticalization of certain periphrastic verbal constructions, where lexical items turn into grammatical morphs over time. Work has been done mapping the development of the periphrastic future *mynd i* [going to] in Welsh (Webb-Davies and Shank 2020). In their paper, the authors found that the process of grammaticalization of this feature began at least 500 years ago (Webb-Davies and Shank 2020: 23), but the structure did not become dominant until the twentieth century. They argue that this is an example of the influence of grammaticalization in English, which has seen rates of BE + *going to* increase (see Fehring and Corrigan 2015 for a variationist perspective on Tyneside English). Other authors have focussed on the grammaticalization of the periphrastic perfect ‘after’ construction, found in Celtic languages, e.g. *dw i wedi gorffen* ‘I’m after finishing/I have finished’ (Ó Sé 2004), which appeared when a gap in the tense and aspect system of Welsh emerged after 1000 AD (Ronan 2012: 240). Periphrastically constructed forms are now so frequent in Modern Welsh that they have all but ousted synthetic, inflected tense forms in

spoken registers (Ronan 2012: 241). For example, the continuous periphrasis using the particle *yn* e.g. *dw i yn* ‘I am’ is now also used as a general present tense.

Studies of languages other than Welsh have also found a rise in the use of the periphrastic over time. Research shows that speakers of Quebec French use the periphrastic future tense more often than the inflected future tense, even in contexts where, according to grammarians, the former should be employed (e.g. in hypothetical contexts, in distant temporal contexts, etc., see Nadasdi *et al.* 2003:202). This preference for the periphrastic has been found to be a trend particularly for younger speakers of French (Wagner and Sankoff 2011). The current work will provide a good evidence base for how often periphrastic and inflected variants are used by young speakers of Welsh, which could help map language change processes (similar to work on French mentioned in this section).

6.3 Previous work

Having presented the two variants of the feature under study, I turn to present previous work on the stylistic, social and linguistic factors which influence the production of inflected and periphrastic verb constructions in the Simple Past tense in Welsh.

6.3.1 Social factors

Overall, grammars and descriptions of Welsh syntax have claimed that written Welsh is more likely to contain the inflected variant than the periphrastic variant (B. M. Jones 1993). King’s description (2018: 29) further claims that the inflected variant is standard in most forms of written Welsh, whereas the periphrastic forms using the auxiliaries *gwneud* and *ddaru* are not. By the same token, it has been claimed that periphrastic constructions are more common in spoken Welsh than written Welsh (P. Willis 1988: 202). In none of these cases, however, have actual rates of use been compared across registers.

Even less is known about how the two variants might pattern across different speech contexts. Intuitively, it may stand to reason that the supposedly more ‘spoken’ periphrastic variant may be used more often in colloquial registers, and the inflected variant could appear more in formal contexts, which may be closer to written language. Both variants have been described as ‘normal’ and ‘interchangeable’ in spoken Welsh (King 2016: 232), but inflected forms have been associated with standard colloquial speech and with the style of Welsh encountered in the media (King 2016: 226). It is recognised, however, that the style of Welsh in the media is not monolithic, and no variationist work, to my knowledge, has considered this feature. This will be the first known work to quantitatively examine actual usage of the two variants across styles.

In King’s (2016) guide to Welsh grammar in spoken Welsh, the periphrastic construction is argued to be widely used by native speakers. It is unclear how King defines a ‘native’ speaker in this context. This claim does not seem to be supported by any quantitative evidence of use by different types of speaker, but implies that non-native speakers (or learners) might not use the periphrastic as often, and may opt instead for the inflected variant. Indeed, learner textbooks tend to favour teaching the inflected forms before teaching the periphrastic Simple Past (see King’s 2014 workbook and the National Centre for Learning Welsh Entry level workbook (Learn Welsh 2023a; 2023b)), and therefore it comes as no surprise that learner use of Simple Past verbs does not match that which King (2016: 30) defines as ‘native’ use. The current work will go some of the way to address the extent to which home language background might affect the use of periphrastic verbs in Welsh.

B. M. Jones (1993: 157), claims that periphrastic constructions are more commonly heard in north Wales. My analysis of rates of use in north west and south east Wales will provide evidence to support or dispute the regional distinction attested by B. M. Jones (1993). Further to this, some authors have noted that the auxiliary *ddaru* is almost exclusively used in northern

regions (e.g. Borsley *et al.* 2007: 42), compared with the auxiliary *gwneud* which is used in all regions. However, little work has considered whether periphrastic and inflected verbs are used to different extents in different regions in Wales. It is not known whether having the choice between two auxiliary variants in north Wales leads to more use of periphrastic constructions, but no quantitative variationist research has explored this.

6.3.2 Linguistic factors

King's (2016) description of spoken Welsh grammar states that the periphrastic may be preferred as it requires no knowledge of the verb-stem; it uses the pattern auxiliary + verb noun (VN). Indeed, part of the suggested 'informality' of the periphrastic construction (versus the inflected construction, as per the written/spoken distinction) is tied to its grammatical simplicity. Whereas the inflected form requires the user to make morphological changes to the verb, the periphrastic form is simplified, using an infinitive as a main verb. Although this is suggested by King (2016: 30) to be a benefit to learners of Welsh, no evidence has been found to suggest that learners opt to use this construction, because of its simpler pattern. Recall, too, that evidence shows that learners tend to be exposed to the inflected form first (see the previous section).

In addition to this, B. M. Jones (1993: 156) writes about the added morphological complexity of some inflected verbs; with some requiring irregular inflection, including changing vowels in some cases (such as in the case of *canu* [to sing] e.g. *cenais i* [I sang]³⁶ vs *gwnes i ganu* [I sang]), and in others, a much greater morphological change (such as in the case of *mynd* [to go] e.g. *es i* [I went] vs *gwnes i fynd* [I went]). The author argues that the morphological complexities of such inflections may prompt the speaker to use the periphrastic construction.

³⁶ *canais* is also widespread (D. G. Lewis 1995: 26)

Section 6.2.2 on the inflected variant demonstrated how interrogative and negative sentences can affect standard varieties of Welsh, by causing a mutation to occur in the initial consonant of the verb being inflected e.g. *welais i ddim* [I didn't see], mutated from *gwelais*. King (2016: 30) suggests that learners may find the periphrastic variant easier to use as it is not governed by the same mutation constraints in different types of sentence, as the auxiliary is always mutated, regardless of sentence type, e.g. *wnes i* [I did], *wnes i ddim* [I didn't], *wnest ti?* [did you?], all mutated from *gwnes*. The assertion that the auxiliary is always mutated is contested by other descriptions of Welsh syntax (Borsley *et al.* 2007: 98).

On the subject of mutation, it should be noted that the auxiliary verb in the periphrastic construction does also prompt a soft mutation in following verbs e.g. *gwnes i fynd* [I went], mutated from *mynd*. Table 6.4 presents the rules of soft mutation in Welsh.

Table 6.4: Soft mutations triggered by the periphrastic variant

Initial consonant	Mutated consonant
c	g
p	b
t	d
g	-
b	f
d	dd
rh	r
m	f
ll	l

To sum up, the authors mentioned above make a number of assumptions about the periphrastic construction being grammatically less complex than the inflected construction. However, as has previously been stated, the authors cited here appear to write from intuition about language use, rather than analysis of actual evidence. The current work will examine whether sentence type (affirmative, interrogative, negative) and regularity are predictors of the periphrastic variant in Welsh. I will also look to see whether periphrastic constructions are favoured over

inflected by assessing whether verbs which can mutate are chosen in periphrastic constructions, so as to avoid the need to mutate.

6.4 Method

I will now present how the feature appears in my data and the method for coding the variables.

6.4.1 The feature in the data

The examples below from the data show how the feature appears in my data in the 1st person singular with the verb *mynd* [to go]. The periphrastic variant is presented below:

wnes i	fynd	i Alton Towers
auxiliary	infinitive	
<i>[I went</i>		<i>to Alton Towers]</i>

The inflected variant is presented below:

es i	ar open day ddoe i Exeter
Inflected verb	
<i>[I went</i>	<i>to an open day yesterday in Exeter ...]</i>

6.4.2 Data extraction

As for the previous possessive pronouns analysis, transcribed text files were extracted from ELAN and AntConc was used to search for and extract verbs, by searching for inflected endings (e.g. *-ais*, *-aist*, *-odd*, *-on* etc.) and auxiliary verbs (*gwnes*, *gwnest*, *ddaru* etc.). All verb tokens were then coded in Excel as inflected or periphrastic. Section 6.2.3 noted that the periphrastic variant can contain the affirmative particle *fe/mi* before the auxiliary, however no emphatic cases were found in the current dataset, thus this does not apply to the analysis. A search was conducted for the Specific Past (Thomas 1996: 106), where the verb *bod* can be used to express a condition or an action which lasts for a period in the past and came to an end in the past. Only

6 tokens were found for this tense, ($n = 5$ GWHL participant, $n = 1$ CEHL participant). Due to its infrequency, and because this type of construction can only be periphrastic (there is no inflected equivalent), it was removed from the analysis. Interruptions and repetitions ($n = 47$) of any tokens e.g. “gwnes i – gwnes i fynd adref” [I – I went home] were addressed by only analysing the latter token. All forms of the irregular preterite verb *cael* [to have/get/receive] e.g. *cawson ni /cafon ni* [we had] were transcribed as they were heard, but the current work does not focus on these regional or dialectal differences. All variants were coded as inflected.

A total of 1492 Simple Past tense verbs were coded in the data.

6.4.3 Social factors

As discussed in section 6.3.1, it has been argued that region and style are influencing factors on the use of periphrastic and inflected verbs in Welsh. Namely, claims about distinctions between spoken and written registers (P. Willis 1988; B. M. Jones 1993; King 2016) might suggest that periphrastic variants are more common in more informal speech contexts, and north Wales speakers are thought to use the periphrastic more often than other regions (B. M. Jones 1993: 157). The literature does demonstrate clearly that there are auxiliary differences between south and north Wales dialects, with *ddaru* belonging exclusively to the latter (B. M. Jones 1993; King 2016; Rottet and Morris 2018: 171).

6.4.4 Internal factors

Internal factors were also coded based on the literature relating to the verb feature, as follows.

6.4.4.1 Polarity

Previous work (such as King 2016: 232) has described the periphrastic and inflected past tense as entirely interchangeable in affirmative and negative sentences in spoken Welsh (e.g. *wnes i fynd* vs *es i* [I went] and *wnes i ddim mynd* vs *es i ddim* [I didn’t go]). However, no known quantitative work has analysed this use in spoken Welsh.

Studies on the periphrastic future construction have found a contrast between negative and affirmative, with the English construction ‘be going to’ favoured in negative constructions (e.g. Torres-Cacoullos & Walker 2009). Tagliamonte, Durham and Smith (2014) looked at a number of different communities and found the effect of polarity depends on how advanced the feature is in each place, with younger speakers of York disavouring the periphrastic in negative sentences. Fehringer and Corrigan (2015: 217) compared the two English variants in Tyneside and did not find sentence type (affirmative and negative) to have an effect on the production of ‘will/won’t’ or ‘not going to/going to’. Studies on verb constructions of French have also examined periphrastic and inflected future, such as *je vais mettre* vs *je mettrai* [I will put]. Tristram (2021) found polarity to be a significant factor in a quantitative analysis of a corpus of spoken French. In Tristram’s work, where used in negative contexts, the inflected future was strongly favoured. In positive contexts the inflected was less strongly favoured, though it was still the most common variant. In order to assess whether similar patterns can be seen in Welsh in the current work, inflected and periphrastic verbs were coded by assessing whether the verb was in negative or affirmative syntax, by assessing the presence/absence of *ddim* e.g. *aeth o ddim/gwnaeth o ddim mynd* [he didn’t go]³⁷.

6.4.4.2 *Interrogative*

Similarly to the case with polarity, little or no research exists in Welsh on the use of periphrastic and inflected variants in interrogative phrases. King (2016: 232) states that interrogative and non-interrogative phrases can be comprised of either variant (e.g. *est ti?* vs *wnest ti fynd?* [did you go?]) without there being any difference in meaning. In the case of English future, some synchronic studies indicate a higher occurrence of ‘going to’ compared to ‘will’ in interrogative phrases. Tagliamonte (2013) observes this in York, and Torres-Cacoullos and Walker (2009)

³⁷ Although the more formal *nid* was also searched e.g. *nid aeth* [he did not go], no results were found of this hyperformal negative item.

report a consistent preference for ‘going to’ in Canadian interrogative contexts, too, especially when involving second-person subjects.

In order to examine this as a potentially influencing linguistic factor, I coded all verb tokens as either interrogative or not.

6.4.4.3 Regularity

B. M. Jones (1993: 156) states that the morphological complexity of some inflected verbs may discourage some speakers from using them and instead elicit periphrastic constructions in their place. According to B. M. Jones’ (1993) theory, highly morphologically-irregular verbs would appear most often in periphrastic constructions in order to avoid this complexity. However, studies of French (Poplack and Turpin 1999: 155-156) observe that “highly frequent and morphologically-irregular verbs” including verbs such as *vouloir*, *pouvoir*, and *savoir*, rather, show an association with inflected variants. As noted in much of the literature, high frequency forms, such as the regularization of irregulars, resist analogical change “because their frequency makes them easy to access whole and there is no need to re-form them by regular rule” (Bybee and Hopper 2001: 17).

Thomas (1996: 33) defines irregular verbs as verbs that have more than one root. In Welsh, there are 7 irregular verbs which fall into 3 groups:

1. *bod* group: *bod* [to be], *adnabod* [to know somebody], *gwybod* [to know something]
2. *mynd* group: *mynd* [to go], *gwneud* [to do], *dod* [to come]
3. *Cael* [to get]

As previously discussed in 1.2.1, stative verbs (the first group) do not tend to take the Simple Past tense in spoken Welsh, instead they use the Imperfect or Specific Past tense. For this reason, the current analysis will focus on groups 2 and 3; the irregular verbs covered are *mynd* [to go], *gwneud* [to do], *dod* [to come] and *cael* [to get].

6.4.4.4 Subject

Another linguistic factor which is yet to be explored in Welsh Simple Past verb constructions, is the role of the subject of the verb. Put simply, little is known about whether particular subjects favour the use of the periphrastic over the inflected verb variant, i.e. are speakers as likely to use the inflected variant with 3rd person singular feminine, *aeth hi* [she went] as with the periphrastic variant in the 1st person singular *gwnes i gysgu* [I slept]? Although no known work has examined this in Welsh, studies on languages other than Welsh have considered this. Many studies on the future temporal reference system in English report that *be going to* is favoured for second- and third-person subjects (Torres-Cacoullos & Walker 2009), due, at least in part, to the fact that first-person subjects are more likely to show volition than other subjects (Tagliamonte *et al.* 2014: 92). Denis and Tagliamonte (2018: 423) came to a similar conclusion, with first-person subjects disfavoured *be going to* overall. However, these authors found a significant interaction with age, showing that younger people have an increasingly levelled use of the grammaticalized periphrastic structure across all subjects. Work on French (Sundell 1991) shows that in a corpus of French novels, third person singular subjects are generally found with much higher rates of inflected future in French. The subject was not found to be a significant factor in the production of the periphrastic future in French immersion students in Canada, however (Nadasdi, Mougeon and Rehner 2003).

In order to explore this factor in the current data, the subject of the verb was coded for as follows: each subject was coded (1st person singular, 2nd person singular, 3rd person singular masculine, 3rd person singular feminine, 1st person plural, 2nd person plural, 3rd person plural). Where conjoined subjects were referenced for the verb (e.g. *gwnaeth fi a Dad dreulio blwyddyn* [me and Dad spent a year]), the tokens ($n = 4$) were coded as compound. Where the subject was a demonstrative pronoun (e.g. *gwnaeth hwnna helpu* [that helped]) tokens ($n = 18$) were coded as demonstrative. Where the subject was an indefinite pronoun, (e.g. *gwnaeth pawb fynd*

[everyone went]), tokens ($n = 29$) were coded as indefinite. Interrogatives (e.g. *beth ddigwyddodd?* [what happened?]) were coded as such ($n = 7$). Lexical noun phrases were also coded (e.g. *wnaeth COVID canslo nhw* [COVID cancelled them], and distinctions were made between feminine ($n = 18$), masculine ($n = 54$) and plural ($n = 11$) noun phrases in the coding.

Table 6.5 below summarises each subject coded for in the data.

Table 6.5: Subject coding for verb variants

Subject	n
1 st person plural	225
1 st person singular	811
2 nd person plural	19
2 nd person singular	32
3 rd person plural	74
3 rd person singular feminine	71
3 rd person singular masculine	119
Compound	4
Demonstrative	18
Indefinite	29
Interrogative	7
Noun phrase feminine	18
Noun phrase masculine	54
Noun phrase plural	11

Due to low token counts, compound, interrogative, indefinite and demonstrative verbs were combined as ‘infrequent’ in order to examine patterns for each speaker group. Noun phrase

tokens (feminine, masculine, plural) were all combined in the analysis presented later in the chapter as patterns appeared largely similar, and tokens were low for all speakers.

6.4.4.5 *Mutation constraints*

All periphrastic auxiliary verbs were mutated in the dataset e.g. *wnes i* (some of these may have also been truncated forms, e.g. *nes i*, but this was not the focus of this analysis) rather than *gwnes i*, supporting King's (2016) observation (see section 6.3.2). As previously stated, periphrastic constructions require the verb following the auxiliary to undergo a soft mutation. The previous chapter showed that possessive use was unlikely to be constrained by mutations, even though some authors had argued that the colloquial possessive variant may be used to avoid the need for mutation. It is unknown whether this might be the case for verb variants in Welsh; if speakers were looking to avoid mutations in affirmative sentences, they might opt to use the inflected variant rather than the periphrastic which triggers a soft mutation. However, inflected verbs also usually mutate in speech too, which might lead to no expectation of this kind. In order to explore whether the requirement for mutation is an influencing factor on the use of the periphrastic variant, the data was coded to assess the following: whether the verb can mutate (according to soft mutation rules); which verb variant is used (if the verb does mutate); and whether the mutation is realised (in the case of periphrastic variants).

6.4.4.6 *Verb language*

As mentioned in the possessive pronoun analysis, Welsh contains many loan words and calques from English, and the use of these has previously been associated with informal speech (Deuchar and Davies 2009). A common suffix for Welsh verbs is *-io* (e.g. *nofio* [to swim], *herio* [to challenge]), and 'established'³⁸ loan verbs from English commonly take the *-io* suffix too, such as *peint-io* [to paint]. This suffix is also heavily used to form Welsh verbs in what

³⁸ 'established' here denotes that an historic English loan now appears as Welsh in Geiriadur Prifysgol Cymru

Borsley *et al.* (2007: 68) describe as “nonce borrowings”; that is, borrowings that are not recurrent or community-accepted loanwords but are “morphologically and syntactically incorporated into the host language” (Poplack and Sankoff 1984: 12). The example below shows a periphrastic verb construction where the main verb is a nonce borrowing from the English verb ‘to panic’.

Ddaru fi (*"panicio") wedyn

[I panicked then]

Scarlett (Coleg y Fro, Gwynedd)

In an inflected construction, the borrowing *panicio* might appear as *paniciais i*. As has been stated, inflected forms are more likely to be found in writing (B. M. Jones 1993), and thus, they may be heard more frequently in more formal speech. Inflected verbs’ association with more formal contexts might also potentially lead to it being used less often with nonce borrowings. No research has considered whether borrowings are used by young speakers to different extents in periphrastic and inflected constructions, and this work will be the first of its kind to view the patterns of use across different speaker groups and in different contexts of use.

6.5 Results (distributional analysis)

Overall, in the past tense coded data, the periphrastic variant (e.g. *wnes i weld* [I saw]/ *ddaru fi weld* [I saw]) was most common, and the inflected variant (e.g. *gwelais i* [I saw]) occurred in 23% of constructions. These rates are presented in Table 6.6.

Table 6.6: Overall rates of periphrastic and inflected verbs

Verb type	%	<i>N</i>
Periphrastic	77	1153
Inflected	23	339
Total		1492

Four (out of 10) of the Gwynedd speakers using *ddaru* and *gwneud* as an auxiliary in periphrastic constructions, and Cardiff participants using only *gwneud*. This confirms that *ddaru* is a northern feature and still exists in this peripheral area of Gwynedd. The periphrastic *ddaru* represented around 5% of periphrastic tokens, but other than being the main choice for one single speaker and an infrequent form for three others there were no visible differences in the use of the two auxiliaries *ddaru* and *wnes* for the other factors, so from hereon in, I will be considering them together

6.5.1 Region

Recall that Gwynedd has much higher rates of daily users of Welsh than Cardiff. Table 6.7 shows that the rates of use of periphrastic and inflected variants in both regions are similar; Cardiff uses marginally more inflected verbs.

Table 6.7: Verb variant by area

Region	Periphrastic (%)	Inflected (%)	Total (N)
Gwynedd	75	25	619
Cardiff	79	21	873

6.5.2 Home language

Next, I break down the use of verb variants by home language. Table 6.8 shows the distribution of verb variants according to the four speaker groups, Gwynedd Welsh home language (GWHL), Gwynedd English home language (GEHL), Cardiff Welsh home language (CWLH) and Cardiff English home language (CEHL).

Table 6.8: Verb variant by home language and area

	Home language	Periphrastic (%)	Inflected (%)	Total (N)

Gwynedd	WHL	69	31	423
	EHL	89	11	196
Cardiff	WHL	70	30	251
	EHL	82	18	622

The group using the most periphrastic constructions (89%) is the Gwynedd EHL group (who also overwhelmingly used the most colloquial construction in the possessive pronoun analysis). This group shows the least amount of variation and seem to rely heavily on the more informal colloquial variant. On the other hand, in both areas, WHL students use similar rates of periphrastic and inflected variants. This points to the variation potentially being more strongly influenced by home language background than by region (see Figure 6.1).

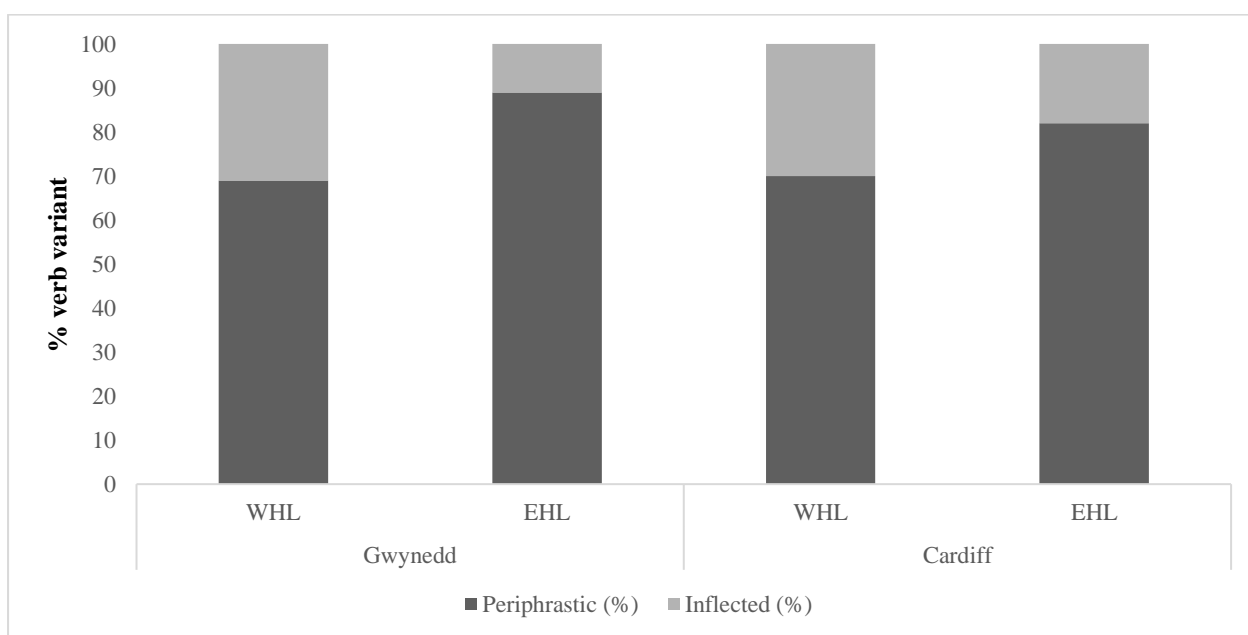


Figure 6.1: Rates of verb variant by home language and region

6.5.3 Speech context

The current section examines the use of these verb constructions in the three contexts of use; peer-to-peer conversation, sociolinguistic interview and mock job interview. It was expected that the use of the inflected variant would increase as the speech context became more formal.

Table 6.9: Verb variant by speaker group and speech context

Speaker group	Speech context	Periphrastic %	Inflected %	Total N
Gwynedd WHL	Job interview	79	21	94
	Sociolinguistic interview	71	29	56
	Peer to peer	66	34	273
Gwynedd EHL	Job interview	90	10	58
	Sociolinguistic interview	88	12	16
	Peer to peer	88	12	122
Cardiff WHL	Job interview	66	34	47
	Sociolinguistic interview	79	21	93
	Peer to peer	66	34	111
Cardiff EHL	Job interview	82	18	155
	Sociolinguistic interview	82	18	191
	Peer to peer	81	19	276

In all groups and in all speech contexts, the periphrastic variant was used more frequently than the inflected variant. In groups where English is spoken at home, very little variation can be observed between different contexts of use. Gwynedd EHL student used the periphrastic slightly more than Cardiff EHL students. On the other hand, more variation can be observed between contexts of use for students with Welsh as a home language. Gwynedd WHL students used more periphrastic variants in the more formal context, followed by the sociolinguistic interview, followed by the peer-to-peer context. For the Cardiff WHL group, the formal job interview elicited relatively lower rates of the periphrastic construction, with

similar rates in the peer-to-peer context. The Cardiff WHL sociolinguistic interview (the least formal context) elicited the highest rates of periphrastic construction.

Whereas EHL students do not vary their use of these variants much between contexts of formality (see the flat distribution across contexts for EHL groups in Figure 6.2), the variation between contexts for Gwynedd WHL and Cardiff WHL students appears inverse. In other words, whereas the data show a higher proportion of periphrastic variants in more formal settings for Gwynedd WHL pupils, Cardiff WHL pupils appear to behave in the opposite way, with higher proportions of the periphrastic variants in informal settings. This will be explored further in the statistical testing. The data show that EHL students do not vary much between settings, whereas WHL students vary, but pattern differently across settings of differing formality.

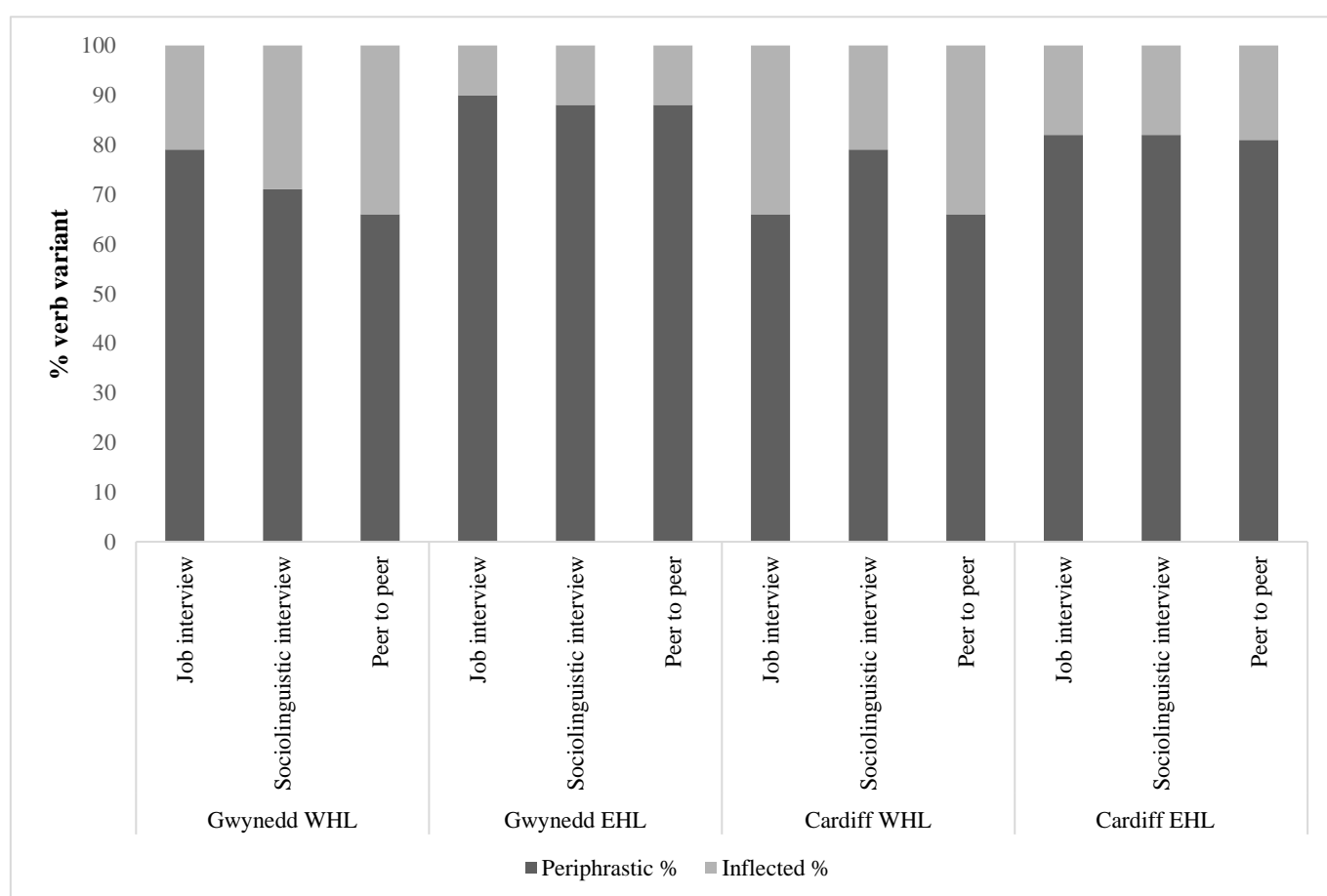


Figure 6.2: Rates of verb variant by speech context and speaker group

As mentioned previously, for the Gwynedd speakers, the formal context elicited the most periphrastic variants, whereas for Cardiff speakers, the informal context elicited the most periphrastic variants. Although similar patterns emerged between schools and home language backgrounds, no clear pattern has been uncovered about the use of periphrastic and inflected verbs in different contexts of formality.

The above suggests that stylistic context may not be a strong predictor of periphrastic use in spoken Welsh, but this will be explored further in the mixed effects modelling later in the chapter. Before I turn to examine the linguistic factors which could influence the use of inflected and periphrastic verbs in Welsh, the main findings so far can be summarised as follows: the periphrastic variant is the most frequently used variant overall and WHL speakers use more inflected variants than their EHL peers, in Gwynedd and in Cardiff. For WHL participants, the Gwynedd speakers use more periphrastic variants in formal settings and Cardiff speakers use more periphrastic variants in informal settings. The EHL speakers do not appear to vary between contexts.

6.5.4 Polarity

According to King (2016: 232) the periphrastic and inflected past tense are entirely interchangeable in affirmative and negative sentences in spoken Welsh (e.g. *wnes i fynd* vs *es i* [I went] and *wnes i ddim mynd* vs *es i ddim* [I didn't go]). The four speaker groups under investigation show different patterns of use, however. These results are presented in Table 6.10.

Table 6.10: Verb variant by polarity and speaker group

Speaker group	Polarity	Periphrastic %	Inflected %	Total (N)
Gwynedd WHL	Affirmative	69	31	393
	Negative	73	27	30
Gwynedd EHL	Affirmative	88	12	188

	Negative	100	0	8
Cardiff WHL	Affirmative	69	31	237
	Negative	85	15	14
Cardiff EHL	Affirmative	82	18	593
	Negative	79	21	29

In Gwynedd, the EHL group used the periphrastic variant in 100% of negative constructions. This categorical use of periphrastic verbs implies that the EHL group is not using the same patterns of periphrastic and inflected verbs in negative sentences as their WHL counterparts. As previously stated, the GEHL group show very little variation in their variant use, but low tokens could account for the little observed variation.

When comparing home language, WHL groups (in Gwynedd and Cardiff) show similar rates of use of periphrastic and inflected variants in affirmative and in negative sentences, but CWHL use slightly more periphrastic constructions in negative sentences than those in Gwynedd. Cardiff EHL is the only group that appears to have an opposite effect, they use relatively higher rates of inflected constructions in negative sentences than in affirmative ones. These proportions are represented in Figure 6.3.

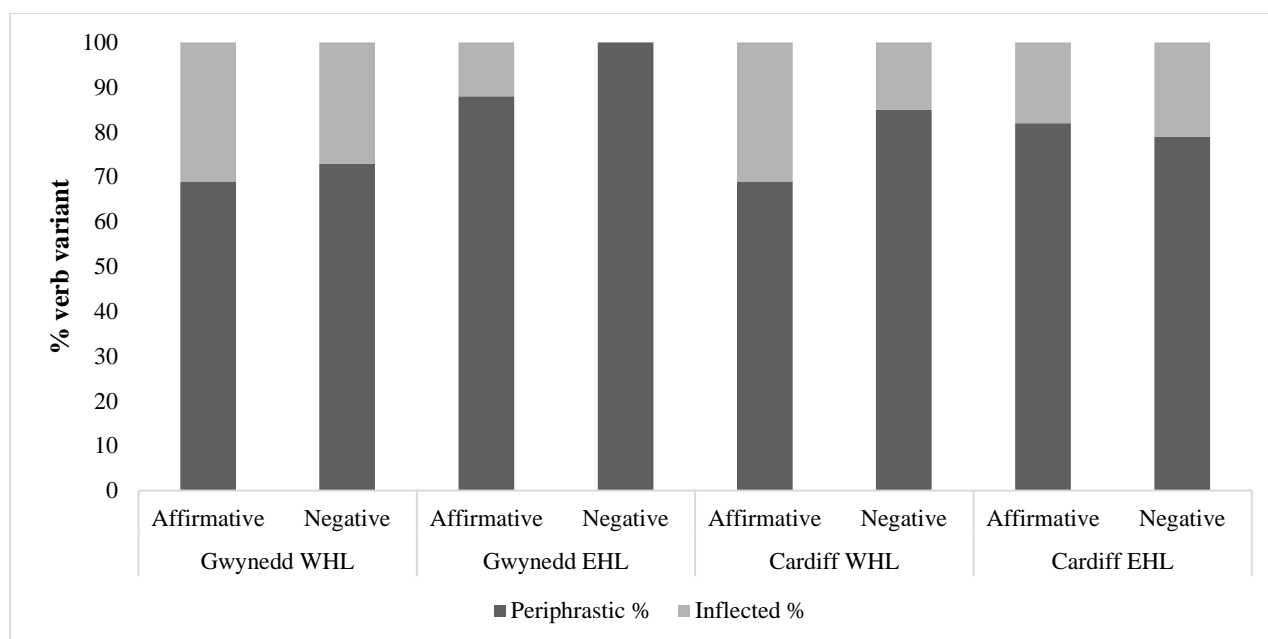


Figure 6.3: Rates of verb variant by polarity and speaker group

6.5.5 Interrogative

King (2016: 232) also states that interrogative and non-interrogative phrases can be comprised of either periphrastic or inflected constructions without a difference in meaning (e.g. *est ti?* vs *wnest ti fynd?* [did you go?]). Table 6.11 shows the rates of use of the two verb variants in interrogative and affirmative sentences. As rates of interrogative sentences are below 10 for all but one speaker group, it is difficult to see the way the feature patterns in the data. Note, however, that Cardiff EHL's patterns of use of periphrastic and inflected verbs is the same across interrogative sentences and affirmative sentences. For the other groups, it is not possible to see whether this is the case because of low tokens.

Table 6.11: Verb variant by interrogative and speaker group

Speaker group	Interrogative	Periphrastic %	Inflected %	Total (N)
Gwynedd WHL	No	69	31	416
	Yes	57	43	7
Gwynedd EHL	No	89	11	195

	Yes	100	0	1
Cardiff WHL	No	70	30	243
	Yes	100	0	8
Cardiff EHL	No	82	18	605
	Yes	82	18	17

6.5.6 Subject

Here I present an analysis of the subject of all periphrastic and inflected past tense verbs, with a view to understanding subject as a grammatical constraint influencing its use. Recall that due to low token counts compound, interrogative, indefinite, and demonstrative categories were combined as ‘infrequent’ in order to examine patterns for each speaker group, presented below. Noun phrases were also combined.

An examination of Gwynedd WHL patterns (Table 6.12) reveals that this speaker group used more inflected verbs than periphrastic with 3rd person singular feminine e.g. *aeth hi* rather than *wnaeth hi fynd* [she went], which is vastly different to their use with 3rd person singular masculine constructions, which were found to be 90% periphrastic e.g. *wnaeth o fynd* rather than *aeth o* [he went]. GWHL participants also used a relatively high rate of inflected variants in 1st person plural verb constructions (40%). As shown previously, GEHL students used the periphrastic variant far more than the inflected variant. This was found to be true in all subject constructions, with 2nd person singular, 3rd person singular masculine and 2nd person plural taking the periphrastic construction 100% of the time, although low cell counts could account for some of this limited variation. GEHL do not seem to vary markedly between use of inflected variants in 3rd person singular feminine (6%) and 3rd person singular masculine (0%) constructions, which sets them apart from the GWHL group. On the other hand, the highest rates of inflected variants for this group are in the 1st person plural construction, the rates of

which are also high for GWHL. This initial exploration may point to different patterns of use between EHL and WHL students, but the observations could be skewed due to the low number of tokens.

Table 6.12: Distribution of Gwynedd inflected and periphrastic variants by subject

	Gwynedd WHL			Gwynedd EHL		
Subject	Periphrastic %	Inflected %	<i>N</i>	Periphrastic %	Inflected %	<i>N</i>
First-person singular	70	30	235	88	12	89
Second-person singular	57	43	7	100	0	2
Third-person singular feminine	46	54	28	94	6	18
Third-person singular masculine	90	10	20	100	0	7
First-person plural	60	40	63	80	20	41
Second-person plural	86	14	7	100	0	1
Third-person plural	85	15	26	90	10	10
infrequent type	82	18	17	92	8	12
Noun phrases	65	35	20	100	0	16

Table 6.13: Distribution of Cardiff inflected and periphrastic variants by subject

	Cardiff WHL			Cardiff EHL		
Subject	Periphrastic %	Inflected %	<i>N</i>	Periphrastic %	Inflected %	<i>N</i>
First-person singular	71	29	153	78	22	334
Second-person singular	100	0	5	83	17	18
Third-person singular feminine	63	38	8	94	6	17
Third-person singular masculine	81	19	16	92	8	76
First-person plural	61	39	38	80	20	83
Second-person plural	100	0	3	88	13	8
Third-person plural	88	13	8	97	3	30
infrequent type	44	56	9	70	30	20
Noun phrases	73	27	11	89	11	36

We now turn to look at the Cardiff data. Similarly to the previous two groups in Gwynedd, CWHL also use many inflected variants in 1st person plural constructions e.g. *aethon ni* rather than *wnaethon ni fynd* [we went]. The marked differences observed between masculine and feminine 3rd person singular constructions seen previously in the GWHL group can also be observed here to an extent, with far more inflected variants occurring in 3rd person singular feminine constructions than in 3rd person singular masculine. There appears to be a WHL pattern in both areas for distinguishing between inflected feminine and periphrastic masculine 3rd person singular phrases. However, low cell counts could contribute to this finding, and further data would be required in order to ascertain whether this is a strong pattern.

Similarly to the CWHL group, in the Cardiff EHL group relatively high proportions of inflected variants were also found in 1st person constructions. However, the CEHL group also follows a similar pattern to GEHL participants, in that little difference can be seen between 3rd person singular feminine and masculine constructions, whereas the WHL groups used more inflected with feminine constructions, and more periphrastic with masculine constructions. This could point to stronger home language influence than region influence on the production of the verb variant.

The mixed effects modelling will explore the significance of the variation observed in the four speaker groups, but Table 6.14 summarises the main patterns observed in the WHL speaker groups, and whether these are replicated by their EHL counterparts.

Table 6.14: Summary of subject patterns observed by WHL and replicated by EHL

Speaker group	Patterns observed	Replicated by EHL counterparts?
GWHL	1 st person plural (relatively) high inflected	Yes

	3 rd person singular feminine high inflected and 3 rd person singular masculine high periphrastic	No
CWHL	1 st person plural (relatively) high inflected	Yes
	3 rd person singular feminine high inflected and 3 rd person singular masculine high periphrastic	No

The mixed effects analysis will keep 3rd person singular feminine and 3rd person singular masculine separate rather than combining them (as was done in the previous analysis) because it looks like for the WHL groups at least, this may be influencing the production of inflected variants. On the other hand, 2nd person plural and 2nd person singular rates are very low for most groups and will therefore be combined for the inferential analysis.

6.5.7 Regularity

There were 208 different verbs in the corpus of periphrastic and inflected constructions in the past tense. Of the six most frequent verbs in the dataset, displayed in Table 6.15, four are the irregular verbs *mynd*, *cael*, *gwneud* and *dod* (highlighted), demonstrating the overall frequency of irregular verbs in Welsh.

Table 6.15: Six most frequent verbs in the dataset

Verb	Gloss	Total (N)
<i>mynd</i>	to go	225
<i>cael</i>	to get/receive/have	173
<i>gwneud</i>	to do	151
<i>dweud</i>	to say	55
<i>rhoi</i>	to give	47
<i>dod</i>	to come	45

Here, I examine the extent to which regularity is an influencing factor on the production of the periphrastic and inflected variants in the four groups. Table 6.16 presents how the four irregular verbs are distributed fairly evenly across periphrastic and inflected constructions, however *gwneud* [to do] is mostly periphrastic (60%), and *mynd* [to go] is mostly inflected (57%). The high rates of inflection for *mynd* [to go] are likely linked to the frequency with which this verb is produced (Table 6.15).

Table 6.16: Periphrastic vs Inflected distribution of irregular verbs

Irregular verb	Periphrastic %	Inflected %	Total (N)
<i>gwneud</i>	60	40	149
<i>dod</i>	55	45	38
<i>cael</i>	45	55	165
<i>mynd</i>	43	57	222

The rates of use for each speaker group are presented in Table 6.17. The table shows that regular verbs are consistently used in periphrastic constructions in the majority of cases, across all four speaker groups. All groups use the periphrastic variant at least 93% of the time when the verb is regular.

Table 6.17: Distribution of verbs by speaker group and verb regularity

Speaker group	Regularity	Periphrastic %	Inflected %	Total (N)
Gwynedd WHL	Irregular	37	60	197
	Regular	95	5	226
Gwynedd EHL	Irregular	78	22	93
	Regular	98	2	103
Cardiff WHL	Irregular	26	74	85
	Regular	93	7	166

Cardiff EHL	Irregular	54	46	219
	Regular	97	3	403

However, a clear distinction can be seen between home language groups in the production of periphrastic and inflected variants with irregular verbs. For both WHL groups, more inflected variants are used with irregular verbs than periphrastic variants; 60% of GWHL verb variants are inflected, and 74% of CWHL verb variants are inflected. Considering the lower rates of inflected verbs overall, this finding points to a clear pattern regarding the regularity of the verb, i.e. irregularity could be a strong predictor of increased inflected verb use. Whereas the CEHL group uses similar rates of periphrastic (54%) and inflected (46%) variants with irregular verbs, this group has acquired the pattern of inflecting irregular verbs to an extent. The GEHL group differs insofar as the same direction of effect is observed, but the strength of the shift is weaker. Figure 6.4 demonstrates the similar direction of pattern for each group.

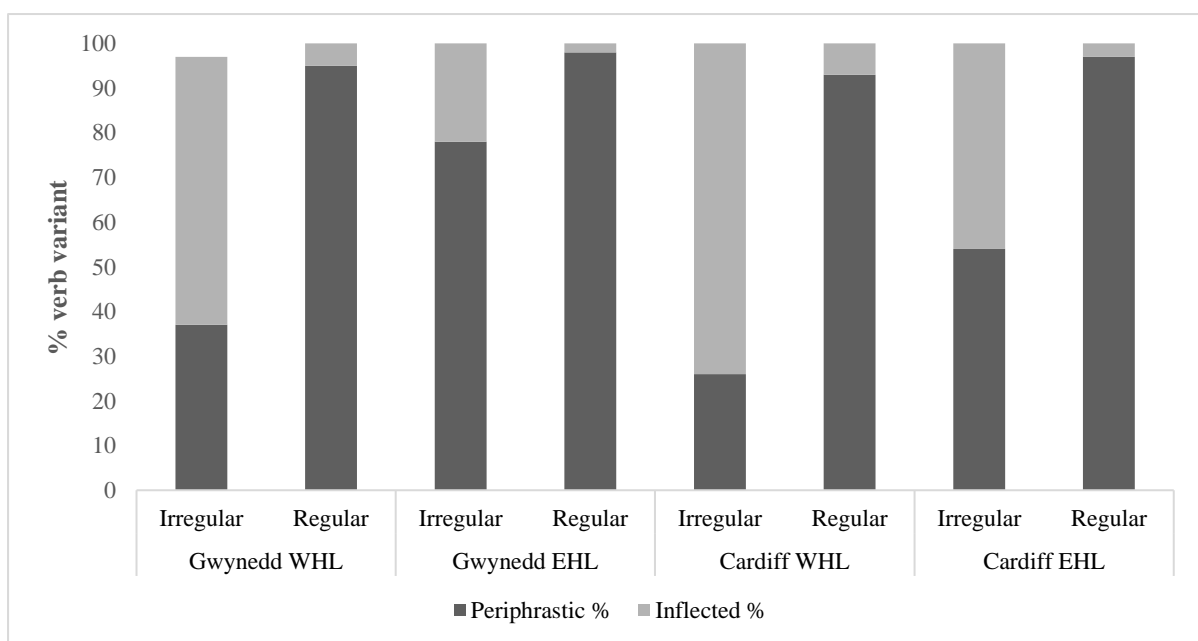


Figure 6.4: Rates of verb variant by regularity and speaker group

6.5.8 Mutation constraints

It has been shown that the periphrastic variant triggers a soft mutation in the following verb (where this applies).³⁹ Most of the items in the data set ($n = 1153$; 77%) would be subject to mutation of the verb-noun if expressed periphrastically.. Of the ones requiring mutation, 332 tokens (29%) appeared in inflected constructions, thus eliminating the need for mutation. The remaining 71% ($n = 821$) of constructions were periphrastic constructions (where the mutation would be required). The frequency of inflection was 3% in the cases where mutation would not be required in the periphrasis. This demonstrates that participating students were more likely to use a periphrastic construction even though the verb required a mutation; pointing to the fact that ‘avoidance of mutation’ is not likely to be an influencing factor on the choice of using verb variants⁴⁰. The requirement for mutation will be omitted from the subsequent inferential analysis.

It has been intimated that speakers may opt to use the periphrastic construction so as to avoid the need for soft mutation in word-initial inflected negative and interrogative constructions. As has been demonstrated in previous sections, the number of tokens for interrogative and negative sentences is low; and thus no conclusive findings can be drawn.

6.5.9 Verb language

Participants in the current study used 34 verbs that were classed as nonce borrowings; that is, they were “morphologically and syntactically incorporated into the host language” (Poplack and Sankoff 1984: 12). This was mostly achieved by adding the *-io* suffix to the English verb. Table 6.18 presents all of the borrowings, along with how they were found to be

³⁹ See 6.3.2 for a full description of how various consonants mutate in this way.

⁴⁰ It is unlikely, but it could be argued that the periphrasis avoids the soft mutation commonly used in the affirmative main clause.

morphologically incorporated into Welsh speech⁴¹. The raw count used in Simple Past tense verb constructions is also included in this table.

Table 6.18: Nonce borrowings in the data and the raw count in Simple Past verb constructions

Borrowed verb	Morphological incorporation in periphrastic constructions	Raw count
enjoy	<i>enjoio</i>	5
tweet	<i>tweetio</i>	4
settle	<i>setlo</i>	3
mention	<i>mentionio</i>	2
order	<i>ordro</i>	2
panic	<i>panicio</i>	2
quit	<i>quitio</i>	2
wrap	<i>wrapio</i>	2
abuse	<i>abusio</i>	1
account	<i>accountio</i>	1
announce	<i>annuncio</i>	1
annoy	<i>annoyio</i>	1
blame	<i>blamio</i>	1
build	<i>buildio</i>	1
collapse	<i>collapsio</i>	1
confess	<i>confessio</i>	1
cut	<i>cytio</i>	1
end	<i>endio</i>	1
fake	<i>fakeio</i>	1
flop	<i>flopio</i>	1
fracture	<i>fracturio</i>	1
freak	<i>freakio</i>	1
Google	<i>Googlo</i>	1
grab	<i>grabio</i>	1
handle	<i>handlo</i>	1
like	<i>like-io</i> ⁴²	1
lock	<i>locio</i>	1
merge	<i>mergio</i>	1
mess	<i>messio</i>	1
put	<i>put</i> ⁴³	1
refuse	<i>refusio</i>	1
root	<i>rootio</i>	1
shake	<i>shakeio</i>	1

⁴¹ Spelling is my own

⁴² Different to the loan verb *licio* [to like], *like-io* refers to the action of ‘liking’ something on-line.

⁴³ There was no syntactic change with the use of the verb put

The table below presents the rates of use of periphrastic and inflected verb variants when used with verbs that are nonce borrowings and Welsh verbs. It was hypothesised that nonce borrowings would elicit lower rates of inflected variants than periphrastic variants, because of the association between borrowings and informal speech. The data presented here show that although tokens were low ($n = 66$) for Simple Past verb variants using nonce borrowings, in 100% of cases, borrowings were used with the more informal periphrastic variants. This was the case for all speaker groups, regardless of area or home language background. This implies that the use of the periphrastic with nonce borrowings is not stylistically constrained (there was no variation between speech contexts), but rather, it is a grammatical constraint in itself. In the current data, no participants were found to inflect borrowings such as *panicio* [to panic] to make the inflected variant *paniciais i*. Table 6.19 shows the rates of use of periphrastic and inflected variants with borrowings and Welsh verb by speaker group. Figure 6.5 shows how each group patterns in the same way.

Table 6.19: Distribution of verbs by speaker group and verb language

Speaker group	Verb language	Periphrastic %	Inflected %	Total (N)
Gwynedd WHL	Borrowing	100	0	27
	Welsh	69	31	396
Gwynedd EHL	Borrowing	100	0	9
	Welsh	88	12	187
Cardiff WHL	Borrowing	100	0	4
	Welsh	70	30	247
Cardiff EHL	Borrowing	100	0	15
	Welsh	81	19	607

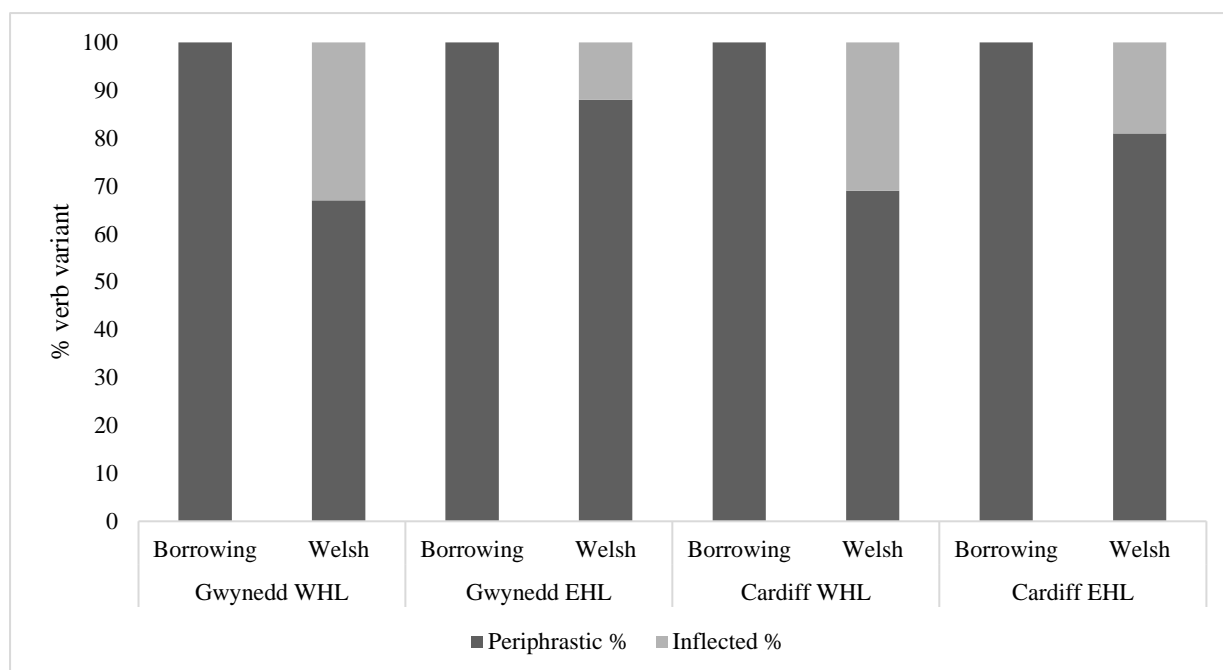


Figure 6.5: Rates of verb variant by verb language and speaker group

6.6 Results (inferential analysis)

A mixed effects model was constructed to examine which linguistic and social factors presented thus far in this chapter have the most significant effects on the production of the more informal periphrastic variant. The model shows the likelihood of the speakers producing the periphrastic variant (the dependent variable) against the inflected variants. Following a maximal model, a Wald χ^2 test was run to determine the magnitude of effect (reported through a χ^2 statistic) for each predictor (Gardner 2023). Significant predictors from the Wald χ^2 test are most likely to accurately explain the variation in the model, and thus a further, more parsimonious, model was run (containing only significant predictors). The hierarchies of the predictor levels was then gauged, this allowing me to report on the three lines of evidence required to compare the constraints (as in previous comparative variationist sociolinguistic studies).

6.6.1 Summary of factors

Table 6.20 presents the random and fixed factors thought to influence inflected and periphrastic verb variation. The levels presented in the table were used to construct the statistical model. Note that the baseline is not indicated in this table, as it was altered in the maximal model to reflect the level hierarchy (as in the analysis of possessive pronouns).

Table 6.20: Factors included in the statistical modelling

Factor	Levels
Context	Mock job interview
	Peer to peer
	Sociolinguistic interview
Speaker's home language	English
	Welsh

Area	Cardiff Gwynedd
Subject	1st person plural 1st person singular 2nd person plural 2nd person singular 3rd person plural 3rd person singular feminine 3rd person singular masculine Noun phrases (masculine, feminine and plural combined) Infrequent (interrogative, demonstrative, compound and indefinite combined)
Polarity	Affirmative Negative
Interrogative	Yes No
Regularity	Regular Irregular
Verb language	English Welsh

6.6.2 Overall model

Here I present the analysis of the periphrastic variant in different speech contexts, in different regions with students from different language backgrounds. As with the possessives analysis, gender was removed from the model because of the imbalance in the dataset. I consider subject

of the verb, polarity, whether it is interrogative and regularity of the verb as predictors. Infrequent subjects ($n = 58$) were removed from the model, due to low tokens, and due to the fact that they combined different types of subject (compound, demonstrative, indefinite). The language of the verb (i.e. whether it is a Welsh verb or a nonce borrowing) was also removed as a predictor as this factor showed no variation in the descriptive analysis; verbs which were nonce borrowing were used in periphrastic constructions 100% of the time. Participant and verb were controlled for in the model. The maximal model is shown in the below R code:

```
VARIANT ~ CONTEXT * HOME LANGUAGE * AREA + SUBJECT + POLARITY +  
INTERROGATIVE + REGULARITY + (1|PARTICIPANT) + (1|VERB)
```

Interactions between social factors were explored in a three-way interaction model including the external factors (demonstrated by an asterisk in the R code above), which was found to have a significantly lower deviance ($p < 0.001$) than the model without interactions, and therefore the interactions were kept. The four speaker groups were analysed as sub-set models in order to compare patterns of use in each.

Each analysis is presented in the subsequent regression tables, which express an intercept corresponding to a baseline combination of levels. Regression coefficients (β) for each factor indicate deviations from the intercept and are included alongside z -values and p -values for the levels associated with each factor. A positive significant coefficient suggests that the named factor level was more likely to influence the production of the periphrastic variant. Conversely, a negative significant coefficient indicates that the named factor level was less likely to result in the production of the periphrastic variant.

Table 6.21 presents the results of generalised mixed-effects modelling on the dataset for the inflected and periphrastic verb variants. The baseline is periphrastic verbs. Each factor in

this model was reordered based on level estimates so that the reference level, i.e., first level, was also the level that most favoured the application value (i.e. has the highest probability of periphrastic verb use). The estimate value for the reference level is 0, and therefore all the estimates are negative and less likely to be used with the periphrastic variants. Within each factor, the estimate value of each level is descending; this step shows the constraint hierarchy for the predictors, although as stated, the hierarchies will not be discussed until the most parsimonious model was created so as to avoid examining an overstuffed model (Gardner 2023). This model shows which levels of certain predictors are significantly (or not) different from the baseline (the reference level shown in brackets after each fixed effect). In Table 6.21 we see that WHL, Cardiff, 1st, 2nd and 3rd person feminine subjects and irregular verbs are significantly different from their respective reference levels. A significant interaction between job interview, WHL and Cardiff was also found in this model.

Table 6.21: Final generalised mixed effects model for verb variants

Fixed effects:	Estimate	Std. Error	z-value	p-value	Sig ⁴⁴	Observations	
						n	% periph.
(Intercept)	13.05	1.76	7.42	<0.001	***		
CONTEXT (<i>vs Sociolinguistic interview</i>)						757	75
job interview	-0.12	0.64	-0.18	0.85		334	80
peer to peer	-0.79	1.02	-0.78	0.43		343	80
HOME LANGUAGE (<i>vs EHL</i>)						786	83
WHL	-2.12	0.76	-2.77	0.005	**	648	69
REGION (<i>vs Gwynedd</i>)						590	74
Cardiff	-1.88	0.82	-2.29	0.02	*	844	79
SUBJECT (<i>vs 3rd person sing. masc.</i>)						119	90
Noun Phrase	-0.55	0.55	-1.00	0.31		83	83
3 rd person plural	-0.89	0.67	-1.33	0.18		74	90
3 rd person singular feminine	-1.38	0.57	-2.42	0.01	*	71	72
1 st person plural	-1.49	0.46	-3.25	0.001	***	225	71
1 st person singular	-1.82	0.43	-4.27	<0.001	***	811	75
2 nd person combined	-2.22	0.78	-2.86	0.004	**	51	84
INTERROGATIVE (<i>vs Yes</i>)						31	83
No	-0.02	0.79	-0.02	0.98		1403	77
POLARITY (<i>vs Negative</i>)						76	80

⁴⁴ No significance where $p > 0.05$; * where $p \leq 0.05$; ** where $p \leq 0.01$; *** where $p \leq 0.001$

Affirmative	-0.33	0.41	-0.81	0.41		1358	77
REGULARITY (<i>vs Regular</i>)						860	96
Irregular	-9.27	3.14	-2.95	0.003	**	574	49
CONTEXT*HOME LANGUAGE (<i>vs sociolinguistic interview*EHL</i>)							
Job interview*WHL	0.72	0.75	0.96	0.33			
Peer to peer*WHL	1.63	1.11	1.46	0.14			
CONTEXT*REGION (<i>vs sociolinguistic interview*Gwynedd</i>)							
Job interview*Cardiff	-0.19	0.74	-0.26	0.79			
Peer to peer*Cardiff	0.44	1.07	0.41	0.67			
HOME LANGUAGE*REGION (<i>vs EHL*Gwynedd</i>)							
WHL*Cardiff	0.62	1.09	0.57	0.56			
CONTEXT*HOME LANGUAGE*REGION (<i>vs sociolinguistic interview*EHL*Gwynedd</i>)							
Job interview*WHL*Cardiff	-2.33	1.14	-2.04	0.04	*		
Peer to peer*WHL*Cardiff	-0.32	1.29	-0.25	0.80			
Random effects:						<i>sd</i>	<i>n</i>
SPEAKER						0.91	18
MAIN VERB						5.75	203

In order to assess the hierarchy of constraints, a sparser and more parsimonious model is needed, using only factors that contribute significantly to predicting the variation (including predictors which had a significant interaction). In order to show which predictors are most explanatory of the variation, results of the Wald χ^2 test are presented in Table 6.22, in ascending order of significance.

Table 6.22: Analysis of deviance, Wald χ^2 test for maximal model

	Wald χ^2	Df	<i>p</i> -value	
Subject	27.47	6	<0.001	***
Region	10.95	1	<0.001	***
Home language	9.62	1	0.001	**
Regularity	8.43	1	0.003	**
Polarity	0.88	1	0.34	
Context	1.71	2	0.42	
Interrogative	0.008	1	0.93	

Subject has the largest χ^2 value (27.47) indicating it has the largest magnitude of effect on the variation. This is followed by Region (10.95), Home language (9.62) and Regularity (8.43). Interestingly (and in line with what was noted in the descriptive analysis) context was not found to be an overall predictor of variation; i.e. inflected variants were not more likely to be used in more formal contexts. Polarity and interrogative constructions were also not found to be significant predictors of the variation. As I have stated in the previous analysis (chapter 5), all predictors found to be significant can be interpreted as adding explanatory value to the study of variation.

The order of parameters in Table 6.23 is based on the relative ordering of the Wald χ^2 test; i.e. Subject, the strongest predictor of periphrastic and inflected verb variation, is listed first. The parameter levels are also ordered by their estimates; i.e. 2nd person constructions were the least likely to be used with periphrastic variants.

Table 6.23: Parsimonious model with only significant predictors of variation

Fixed effects:	Estimate	Std. Error	z-value	p-value	Sig	Observations	
						n	% periph.
AIC = 890.4							
(Intercept)	12.09	1.42	8.52	< 0.001	***		
SUBJECT (<i>vs 3rd person sing. masc.</i>)						119	90
Noun Phrase	-0.64	0.54	-1.20	0.228		83	83
3 rd person plural	-0.93	0.65	-1.43	0.153		74	90
3 rd person singular feminine	-1.44	0.55	-2.64	0.008	**	71	72
1 st person plural	-1.54	0.44	-3.48	<0.001	***	225	71
1 st person singular	-1.81	0.41	-4.42	<0.001	***	811	75
2 nd person combined	-1.96	0.68	-2.88	0.004	**	51	84
REGION (<i>vs Gwynedd</i>)						590	74
Cardiff	-1.59	0.49	-3.22	0.001	**	844	79
HOME LANGUAGE (<i>vs EHL</i>)						786	83
WHL	-1.48	0.49	-3.05	0.002	**	648	69
REGULARITY (<i>vs Regular</i>)						860	96
Irregular	-8.93	3.06	-2.91	<0.001	**	574	49
Random effects:						sd	n
SPEAKER						0.91	18
MAIN VERB						5.75	203

EHL participants were significantly more likely to use the periphrasis than their WHL peers, and the variant was significantly more likely to occur in Gwynedd speech than in Cardiff speech. Regularity was also a significant predictor in the mixed effects model; irregular verbs were significantly more likely to be used with inflected verbs. 3rd person singular masculine constructions were most likely to be used with the periphrastic construction, but 3rd person singular feminine, and all 1st and 2nd person subjects were significantly more likely to be used with the inflected variant than the reference level. However, in order to delve deeper into the nuance between the subject predictor levels, a contrast matrix of all the comparisons (e.g. noun phrase vs 3rd person plural, noun phrase vs 3rd person feminine, noun phrase vs 1st person plural etc.) was devised, the results of which are presented in Table 6.24 below. This was done using the `glht` function in R (Hothorn, Bretz, and Westfall 2008).

Table 6.24: Comparison matrix for verbs by subject

	Estimate	Std. Error	z-value	p-value	Sig
NP - 3PSM == 0	-0.64	0.53	-1.20	0.87	
3PP - 3PSM == 0	-0.92	0.65	-1.42	0.76	
3PSF - 3PSM == 0	-1.44	0.55	-2.63	0.10	
1PP - 3PSM == 0	-1.54	0.44	-3.47	0.008	**
1PS - 3PSM == 0	-1.81	0.41	-4.41	< 0.001	***
2PC - 3PSM == 0	-1.96	0.68	-2.87	0.05	
3PP - NP == 0	-0.28	0.64	-0.43	0.99	
3PSF - NP == 0	-0.79	0.53	-1.48	0.72	
1PP - NP == 0	-0.90	0.44	-2.03	0.36	
1PS - NP == 0	-1.16	0.40	-2.86	0.05	
2PC - NP == 0	-1.32	0.67	-1.95	0.41	
3PSF - 3PP == 0	-0.51	0.64	-0.80	0.98	
1PP - 3PP == 0	-0.61	0.55	-1.11	0.91	
1PS - 3PP == 0	-0.88	0.52	-1.68	0.59	
2PC - 3PP == 0	-1.03	0.74	-1.38	0.78	
1PP - 3PSF == 0	-0.10	0.43	-0.23	0.99	
1PS - 3PSF == 0	-0.37	0.40	-0.91	0.96	
2PC - 3PSF == 0	-0.52	0.68	-0.76	0.98	
1PS - 1PP == 0	-0.26	0.24	-1.09	0.91	
2PC - 1PP == 0	-0.41	0.59	-0.70	0.99	
2PC - 1PS == 0	-0.15	0.56	-0.26	0.99	

The results indicate that the only real contrast for this factor group is 3rd person singular masculine versus 1st person constructions (singular and plural). 1st person constructions appear to favour inflected verbs, this is similar to work on the future temporal reference system in English (Denis and Tagliamonte 2018: 423), which found that first-person constructions disfavoured the periphrastic *be going to* overall. All other results indicate that the difference in likelihood of any paired comparison of different factor levels on the Intercept are not significantly different from zero. As stated by Gardner (2023), by performing a detailed analysis of the contrasts between factors in addition to an analysis of the contrasts between factors and their mean, a nuanced and superior understanding of the three lines of evidence can be achieved, as you can pinpoint exactly where significant contrasts exist.

Next, I show the results of the parsimonious model in a log-odds graph, where zero on the x-axis represents the likelihood of the periphrastic variant occurring, when all predictors are set to their reference values. Remember that all reference values were reordered in the models to the most favouring values, so all the values represented in the plot are below zero, as they have negative estimates (i.e., they all disfavour the use of the periphrastic relative to the reference values). In the graph below, each predicting factor has been grouped by colour. Figure 6.6 plots the *p* value results as well as their confidence intervals (horizontal error bars) in order to display with what certainty any significant factors can predict the use of the periphrastic verb variant.

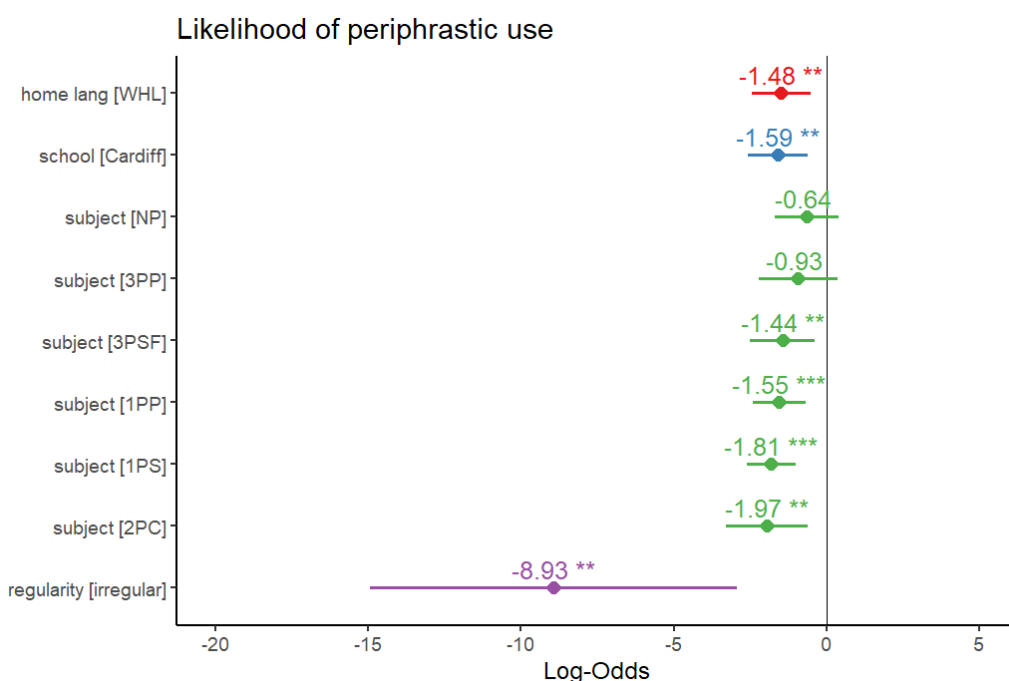


Figure 6.6: Log-odds likelihood of periphrastic use

The predictor levels with error bars overlapping the zero line (for example noun phrase and 3rd person plural constructions), are not significantly different from the reference level of that predictor (3rd person singular masculine). Any error bars that do not overlap between levels of a predictor indicate significant difference from each other (which is what the comparison matrix has already shown).

6.6.3 Summary

We have seen so far that subject is the strongest predictor of periphrastic use. Significant differences were found between 1st person constructions, which tend to be more inflected (e.g. *es i* [I went], *aethon ni* [we went]), and 3rd person singular masculine constructions, which are more likely to be periphrastic (e.g. *gwnaeth e fynd* [he went]). Region had the second largest effect size, with Gwynedd speakers using significantly more periphrastic verbs than Cardiff. Home language was the third strongest predictor, with EHL students using significantly more periphrastic than inflected variants. Finally, irregular verbs (*mynd*, *cael*, *dod* and *gwneud*) were significantly more likely to be inflected.

In order to compare the four speaker groups (considering the strength of effect of the region and home language predictors), a statistical analysis for each group is run below. The subset models include the predictors found to be statistically significant in the maximal model. This includes the context predictor, which, although not significant in the most parsimonious model, was found to be significant in an interaction with home language and region. The subset analysis will allow us to explore this in more detail. The interrogative and polarity predictors were removed from the subset models because no variation was observed in interrogative constructions for two groups (GWHL and CWHL), and no variation was found in polarity for GWHL. This limits the comparison that is able to be made across groups. The contrast matrix presented above pointed to differences particularly around 1st person constructions, therefore, due to the significance of this level, all other subjects were combined into one level ('other'). This also helped account for very low token counts for the numerous levels of subject in each subset model. The levels of each were ordered based on the size of the estimate reported in the maximal model.

6.6.4 Gwynedd Welsh home language subset

Table 6.25 shows the mixed effects analysis for the GWHL subset speaker group.

Table 6.25: Subset mixed effects model for GWHL

	Estimate	Std. Error	z-value	p-value	Sig	Observations	
						n	% periph.
Fixed effects:							
(Intercept)	10.74	2.18	4.90	<0.001	***		
SUBJECT (<i>vs other</i>)						108	70
first	-0.43	0.45	-0.95	0.34		298	68
REGULARITY (<i>vs Regular</i>)						214	95
Irregular	-13.40	6.59	-2.03	0.04	*	192	39
CONTEXT (<i>vs sociolinguistic interview</i>)						262	65
Job interview	0.60	0.52	1.16	0.25		90	78
Peer to peer	0.13	0.57	0.24	0.80		54	72
Random effects:						sd	n
SPEAKER						1.74	6
MAIN VERB						8.35	103

Participants from Gwynedd with Welsh as a home language were significantly more likely to use the periphrastic variant with regular verbs rather than irregular verbs (*mynd* [to go], *gwneud* [to do], *cael* [to get/have receive] and *dod* [to come]). Notice that although context was not found to be a significant predictor of periphrastic variant use, the job interview context elicited the highest rates of the periphrastic variant, followed by the sociolinguistic interview, followed by the peer-to-peer context. This is not what was expected for this variant, as it has been argued that the periphrastic may be more common in informal contexts. The hierarchy of the factors levels will be discussed further in 6.6.8.

Recall that the size of the χ^2 statistic in a Wald χ^2 test can reveal the order of magnitude of effect. A Wald χ^2 test revealed that for GWHL, regularity was the strongest predictor (and only significant predictor) of periphrastic variant use ($\chi^2 = 4.12$, $df = 1$, p -value = 0.04), followed by context and then subject (neither of which were significant). The results are presented in Table 6.26.

Table 6.26: Wald χ^2 test for GWHL periphrastic use

	Wald χ^2	Df	p -value
Regularity	4.12	1	0.04 *
Context	1.36	2	0.50
Subject	0.89	1	0.34

A parsimonious model for GWHL was run with regularity (as this was the only significant predictor) and is presented in Table 6.27. This model confirms the results from the full subset model for GWHL; that is, irregular verbs are significantly more likely to be used in inflected constructions than periphrastic for this group.

Table 6.27: Parsimonious model for GWHL subset

Fixed effects:	Estimate	Std. Error	z-value	p -value	Sig	Observations	
						n	% periph.

(Intercept)	10.59	2.13	4.95	<0.001	***		
REGULARITY (<i>vs Regular</i>)						214	95
Irregular	-13.40	6.39	-2.09	0.04	*	192	39
Random effects:						<i>sd</i>	<i>n</i>
SPEAKER						1.74	6
MAIN VERB						8.35	103

6.6.5 Gwynedd English home language subset

Table 6.28 presents the analysis of the GEHL subset group.

Table 6.28: Subset mixed effects model for GEHL

Fixed effects:	Estimate	Std. Error	z-value	p-value	Sig	Observations	
						<i>n</i>	% periph.
(Intercept)	5.38	1.16	4.64	<0.001	***		
SUBJECT (<i>vs other</i>)						52	96
first	-1.29	0.81	-1.59	0.10		116	85
REGULARITY (<i>vs Regular</i>)						87	78
Irregular	-2.61	0.78	-3.35	<0.001	***	97	98
CONTEXT (<i>vs sociolinguistic interview</i>)						116	87
Job interview	-0.13	0.60	-0.22	0.82		53	89
Peer to peer	-0.82	0.93	-0.88	0.38		15	87
Random effects:						<i>sd</i>	<i>n</i>
SPEAKER						0.6	4
MAIN VERB						0.0	52

Similarly to the GWHL group, GEHL participants were also significantly less likely to use periphrastic variants with irregular verbs. As before, first person constructions were more likely to appear inflected than periphrastic, but not to a significant extent. Rates of periphrastic use patterned very similarly across job interview, sociolinguistic interview and the peer-to-peer context for this group. Recall that it was anticipated that the more informal contexts would elicit the highest rates of periphrastic use (although this was not the case for the GWHL group). The hierarchies will be discussed further in section 6.6.8.

A Wald χ^2 test for GEHL revealed that the magnitude of effect was similar to the GWHL group, in the sense that it was significantly largest for regularity. The second largest effect was

for subject, followed by context. Therefore, not only has the hierarchy of the context predictor not been fully acquired by GEHL speakers; the relative size of effect of context is not the same for the GEHL as for the GWHL group either. This shows that only some constraints of the use of periphrastic variants have been acquired in Gwynedd by non-Welsh home language speakers.

Table 6.29: Wald χ^2 test for GEHL periphrastic use

	Wald χ^2	Df	<i>p</i> -value
Regularity	11.22	1	<0.001 ***
Subject	2.55	1	0.10
Context	0.77	2	0.68

The parsimonious model for GEHL also confirms that irregular verbs are significantly more likely to be used with inflected verbs.

Table 6.30: Parsimonious model for GEHL subset

Fixed effects:	Estimate	Std. Error	<i>z</i> -value	<i>p</i> -value	Sig	Observations	
						<i>n</i>	% periph.
(Intercept)	4.25	0.83	5.12	<0.001	***		
REGULARITY (<i>vs Regular</i>)						87	78
Irregular	-2.62	0.76	-3.46	<0.001	***	97	98
Random effects:						<i>sd</i>	<i>n</i>
SPEAKER						0.6	4
MAIN VERB						0.0	52

6.6.6 Cardiff Welsh home language subset

Table 6.31 shows the analysis of verb variants for the CWHL speaker group.

Table 6.31: Subset mixed effects model for CWHL

Fixed effects:	Estimate	Std. Error	<i>z</i> -value	<i>p</i> -value	Sig	Observations	
						<i>n</i>	% periph.
(Intercept)	5.49	1.36	4.03	<0.001	***		
SUBJECT (<i>vs other</i>)						51	96
first	-0.76	0.61	-1.26	0.21		191	69
REGULARITY (<i>vs Regular</i>)						160	94
Irregular	-6.03	1.56	-3.87	<0.001	***	82	27
CONTEXT (<i>vs sociolinguistic interview</i>)						110	66

Job interview	-2.04	0.89	-2.29	0.02	*	41	68
Peer to peer	0.72	0.59	1.21	0.22		91	79
Random effects:						<i>sd</i>	<i>n</i>
SPEAKER						0.24	4
MAIN VERB						1.85	68

As with the Gwynedd participants, CWHL speakers were also significantly more likely to use the periphrastic variant when the verb is regular. For this group, the context also was a significant predicting factor on the use of periphrastic verb variants.

The magnitude of effect is presented in the Wald χ^2 test in Table 6.32 below. It reveals that regularity is the strongest predictor, followed by context, followed by subject. This is similar to what was found in the GWHL group, implying that constraints on use of periphrastic may be shared across the two regions under study, even though the hierarchy of use across contexts differs between Gwynedd and Cardiff.

Table 6.32: Wald χ^2 test for CWHL periphrastic use

	Wald χ^2	Df	<i>p</i> -value
Regularity	15.01	1	<0.001 ***
Context	8.71	2	0.01 *
Subject	1.59	1	0.21

The parsimonious model for CWHL contains both significant predictors (regularity and context) and the results are presented in Table 6.33 below. Again, the parsimonious model confirms (as with the previous two subset groups), that irregular verbs are significantly more likely to be inflected. The direction of the effect of context shows the least amount of periphrastic variants were used in the job interview context, followed by the sociolinguistic interview, followed by the peer to peer context. The previous analysis of possessive pronouns showed that for the CWHL group, the peer to peer context elicited significantly more informal variants. This analysis provides further evidence that the peer group context was considered the most informal context by this group.

Table 6.33: Parsimonious model for CWHL subset

Fixed effects:	Estimate	Std. Error	z-value	p-value	Sig	Observations	
						n	% periph.
(Intercept)	4.78	1.19	4.02	<0.001	***		
REGULARITY (<i>vs Regular</i>)						160	94
Irregular	-5.93	1.51	-3.93	<0.001	***	82	27
CONTEXT (<i>vs sociolinguistic interview</i>)						110	66
Job interview	-2.07	0.88	-2.34	0.02	*	41	68
Peer to peer	0.71	0.59	1.19	0.23		91	79
Random effects:						sd	n
SPEAKER						0.24	4
MAIN VERB						1.85	68

6.6.7 Cardiff English home language subset

Table 6.34 presents the mixed effects model for the CEHL group subset.

Table 6.34: Subset mixed effects model for CEHL

Fixed effects:	Estimate	Std. Error	z-value	p-value	Sig	Observations	
						n	% periph.
(Intercept)	10.42	1.82	5.72	<0.001	***		
SUBJECT (<i>vs other</i>)						185	91
first	-1.32	0.37	-3.55	<0.001	***	417	78
REGULARITY (<i>vs Regular</i>)						389	97
Irregular	-9.01	3.27	-2.75	0.005	**	213	54
CONTEXT (<i>vs sociolinguistic interview</i>)						269	81
Job interview	-0.42	0.4	-1.01	0.31		150	83
Peer to peer	-0.71	0.40	-1.77	0.07		183	83
Random effects:						sd	n
SPEAKER						0.92	4
MAIN VERB						5.57	119

Similarly to the CWHL group (and both Gwynedd groups), the CEHL subset were significantly less likely to use the periphrastic with irregular verbs. Along with all previously discussed groups, verbs in the first person were more likely to be inflected, and for CEHL this was true to a significant extent. This group, however, do not display the same hierarchy between contexts; that is to say that periphrastic verbs were least likely to occur in the peer to peer context, followed by the job interview, compared to the sociolinguistic interview. Again,

in keeping with previous findings from the possessive pronoun analysis, for the CEHL group, the sociolinguistic interview elicited the highest rates of the most informal variant. The model provides further evidence that for the CEHL group, the sociolinguistic interview is the more informal context, which contrasts with the CWHL group, for whom it was the peer-to-peer context that was most informal.

A Wald χ^2 test revealed a different magnitude of effect for CEHL compared to CWHL, with subject being the strongest predictor of periphrastic use, followed by regularity, followed by context. This is summarised in Table 6.35 below.

Table 6.35: Wald χ^2 test for CEHL periphrastic use

	Wald χ^2	Df	<i>p</i> -value
Subject	12.60	1	<0.001 ***
Regularity	7.57	1	0.005 **
Context	3.20	2	0.20

It appears that the constraints which dictate the use of the periphrastic variant are not the same for CEHL as for CWHL. The parsimonious model (see Table 6.36) for CEHL includes the two significant predictors observed following the Wald χ^2 test; subject and regularity. For CEHL, first person constructions and irregular verbs were significantly more likely to be inflected than periphrastic.

Table 6.36: Parsimonious model for CEHL subset

	Estimate	Std. Error	<i>z</i> -value	<i>p</i> -value	Sig	Observations	
						<i>n</i>	% periph.
Fixed effects:							
(Intercept)	8.87	1.79	4.94	<0.001	***		
SUBJECT (<i>vs other</i>)						185	91
first	-1.41	0.37	-3.82	<0.001	***	417	78
REGULARITY (<i>vs Regular</i>)						389	97
Irregular	-9.08	3.41	-2.66	0.007	**	213	54
Random effects:						<i>sd</i>	<i>n</i>
SPEAKER						0.92	4
MAIN VERB						5.57	119

6.6.8 Hierarchies for subset groups

Figure 6.7 below presents the log-odds (of full subset models) for all four speaker groups, in order to compare the hierarchies of each predictor. As with the previous analysis chapter, whereas the parsimonious models provide us with the first line of evidence, the third line of evidence should be comparable across groups, and thus even factors not considered predictors need to be compared. The line indicating zero represents the estimate for the missing predictor level (namely the sociolinguistic interview, the ‘other’ subject, and regular verbs). The position of the other predictor levels shows whether they are more or less likely to use periphrastic variants than the reference (missing) level. Note that although the x axis is different for each, strong patterns emerge. The most notable is that the hierarchies for subject and regularity are the same across groups. That is to say that for all groups, first person constructions are consistently less likely to be used in periphrastic constructions. Recall that error bars crossing the intercept are not significant, showing that the case of subject is only significant for the CWHL group. As was shown in all the models above, for all speaker groups, regularity was the strongest predictor of periphrastic use, and this was a significant finding for all.

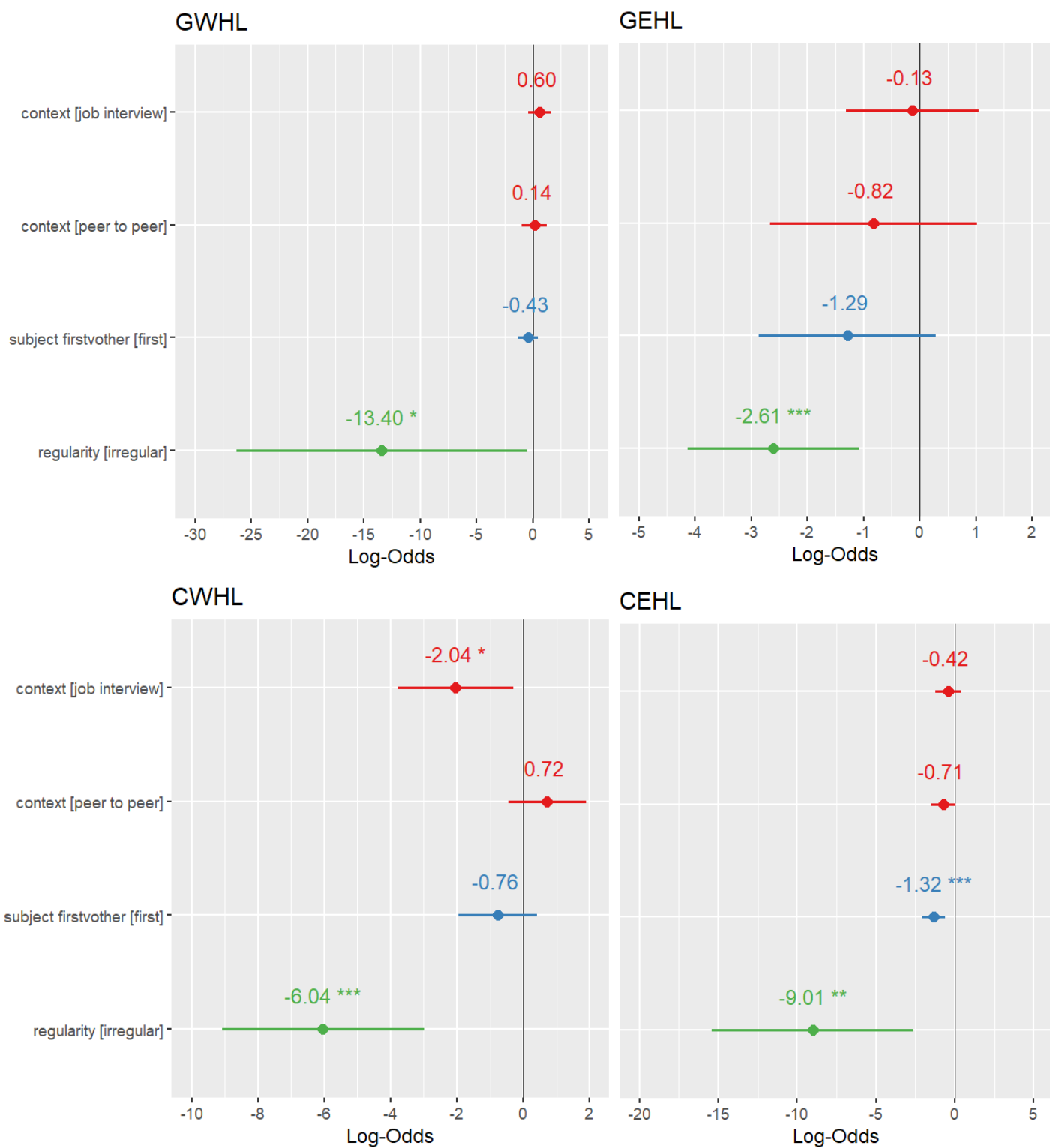


Figure 6.7: Log-odds likelihood of periphrastic use across 4 speaker groups

Context was not found to be a significant predictor of periphrastic use outside of the CWHL group. The log-odds show that the hierarchy of the context levels varied between groups. It was expected that the job interview would elicit the lowest rates of periphrastic use, as it is considered a more colloquial feature. This was the case for all but the GWHL group. The possessive pronoun analysis also showed that Gwynedd participants had higher rates of informal variants across different contexts. It is possible that in Gwynedd, where the Welsh language is more widespread in the community, speakers are more comfortable using more informal language in more formal contexts than in Cardiff. It appears that for WHL speakers, the peer-to-peer context prompts further use of informal variants compared with the sociolinguistic interview. For EHL, on the other hand, the sociolinguistic interview contains more informal variants than the peer-to-peer context. As discussed previously, this is likely to be due to the fact that those who are from EHL backgrounds were more likely to use English with peers, making the Welsh-medium peer-to-peer task for the current study less ‘natural’ and thus slightly more formal than for the WHL groups.

Any error bars that do not overlap between levels within a predictor are likely to be significantly different from one another. There appears not to be any overlap in log-odds error bars between job interview and peer to peer context for CWHL. The mixed effects modelling only allows us to measure the significance of each level against the reference level. To see whether there is a significant difference between the job interview and peer to peer context, a Tukey test was run in R, revealing a significant difference between the two contexts (estimate= 2.76, std. error = 0.94, z -value = 2.95, p -value= 0.009). That is to say that for the CWHL group, there is a significant difference between the use of periphrastic variants in all three contexts, the most variants of this type occurring in the peer-to-peer context, followed by the sociolinguistic interview, followed by the job interview context. Next, I present a discussion of the findings presented thus far.

6.7 Discussion

In terms of the acquisition of sociolinguistic competence, the findings presented in this chapter show evidence of EHL speakers acquiring the norms of WHL speakers. The findings are summarised in the three tables below, one for each line of evidence.

6.7.1 Line of evidence 1: Statistical significance

This first table (Table 6.37) shows the significant results from each subset model. The patterns for the two WHL differs between the two areas. For GWHL, only regularity is significant, and all (non)significant factor constraints match with GEHL, showing how, in Gwynedd, patterns are acquired. For CWHL, regularity and context were significant, but CEHL only match the regularity constraint. Where CEHL had no significant finding for context, they were, however, significantly constrained by subject.

Table 6.37: Line of evidence 1 (verbs)

	Gwynedd		Cardiff	
	WHL	EHL	WHL	EHL
Regularity	Yes	Yes	Yes	Yes
Context	No	No	Yes	No
Subject	No	No	No	Yes

6.7.2 Line of evidence 2: Strength of effect

Table 6.38 presents the findings from the Wald χ^2 tests conducted for each subset model. The strongest effects are at the top of the table and are shown in decreasing order. WHL groups share the pattern in terms of strength of effect; regularity is the strongest predictor of periphrastic use (which is not surprising, considering it was significant in the first line of evidence), followed by context, followed by subject. Compare the size of the effect, however,

between the groups. Although for GWHL regularity has the strongest effect ($\chi^2 = 4.21$), this is far smaller than the effect size for CWHL ($\chi^2 = 15.01$).

For both EHL groups, context is the weakest predictor, however, note the size of the effect for context between the two regions. Though weak, context still has a stronger effect in Cardiff ($\chi^2 = 3.20$) than in Gwynedd ($\chi^2 = 0.77$). For CEHL, subject is the strongest predictor. This line of evidence shows that verb use is constrained by different factors for the two home language groups; EHL is more constrained by internal linguistic factors, whereas the effect of the social factor could be stronger for WHL.

Table 6.38: Line of evidence 2 (verbs)

	Gwynedd				Cardiff			
Strength of effect	WHL		EHL		WHL		EHL	
↓	Regularity	4.12	Regularity	11.22	Regularity	15.01	Subject	12.60
	Context	1.36	Subject	2.55	Context	8.71	Regularity	7.57
	Subject	0.89	Context	0.77	Subject	1.59	Context	3.20

6.7.3 Line of evidence 3: Constraint hierarchy

The third line of evidence is the hierarchy of the levels within the factor. This allows us to assess to what extent patterns of variation have been acquired. The darker shaded cells indicate where the EHL speakers have not acquired the same constraint hierarchy as their WHL counterparts.

Table 6.39: Line of evidence 3 (verbs)

	Gwynedd		Cardiff	
	WHL	EHL	WHL	EHL
Regularity	Irregular verbs least periphrastic	Irregular verbs least periphrastic	Irregular verbs least periphrastic	Irregular verbs least periphrastic
Context	Job interview > peer to peer > sociolinguistic interview	Sociolinguistic interview > job interview > peer to peer	Peer to peer > sociolinguistic interview > job interview	Sociolinguistic interview > job interview > peer to peer
Subject	1 st person least periphrastic	1 st person least periphrastic	1 st person least periphrastic	1 st person least periphrastic

From observing the constraint hierarchy, it is clear that EHL speakers have acquired all the internal patterns of their WHL counterparts, but not the context patterns. WHL speakers in Gwynedd used more periphrastic constructions in the job interview, whereas WHL in Cardiff used the least periphrastic constructions in that context. This points to differences in WHL stylistic variation in the two regions, underlining the importance of examining variation within local community norms. Both EHL groups, on the other hand, have matching patterns. Here, the rates of periphrastic are lowest in the peer-to-peer context, making the effect of the Welsh-medium task more pronounced for these groups who generally use English together. The picture of stylistic variation was clearer for the previous chapter, which is likely to be due to the fact that periphrastic verb constructions are less stigmatised than colloquial pronoun constructions.

Having considered all three lines of evidence, it appears that EHL students have acquired all the internal constraints on Simple Past tense verb use. In Cardiff, however, EHL are further constrained by subject than their WHL counterparts. The local WHL social factor constraints have not been acquired in either region by EHL speakers. Rather, EHL patterns supersede local norms and could be similarly constrained by peer group norms rather than the intended formality of each devised context. Sociolinguistic competence is mostly acquired for this verb feature.

6.7.4 Insights to the feature beyond sociolinguistic competence

Findings from this analysis also show:

- The analysis of the feature showed that the periphrastic was the most frequent variant in the speech of the young people studied, regardless of the area they lived in or their home language background.
- The widespread use of the periphrastic variant among younger speakers supports Borsley *et al.*'s (2007: 12) argument that the periphrastic variant is 'spreading [...] in the modern language'. B. M. Jones claimed that periphrastic constructions are more commonly heard in the Welsh of north Wales than in the traditional southern dialects (1993: 157), and in my analysis, I show that the Cardiff variety, similar to the claim about northern speakers, also used mostly periphrastic overall.
- The periphrastic variant may also be undergoing similar processes of grammaticalization as the future 'mynd i' [going to] variant in Welsh as discussed by Webb-Davies and Shank (2020), and in keeping with previous findings about the historical decline of inflected forms in Welsh over time (Ó Sé 2004; Ronan 2012: 240).
- There is little indication of style-shifting in the current findings, which disputes the impressionistic claims suggesting that the inflected variant belongs more to

written varieties and standard spoken Welsh (King 2016: 226) compared with the more informal periphrastic variant.

- There was a statistically significant difference between the rates of the irregular verbs and the regular ones. It is possible that the frequency of these verbs (as described in 6.5.7) could explain the tendency to use inflected forms. It is possible that these verbs are mostly inflected due to having become conventionalised ‘set phrases’ in the vernacular of speakers, such as Davies (2016: 44) hypothesised with high frequency nouns taking colloquial pronoun variants (see chapter 5).
- The regularity constraint was significant for all four groups, and was also the factor that showed the strongest effect. Although all speakers seemed to avoid regular inflection, this was not the case for irregular (stored) inflection. This is consistent with B. M. Jones (1993). The patterns seen here may not be tied to sociolinguistic competence, but simply with the fact that irregular verbs have to be learnt in a different way. Further work could consider input from teachers or teaching materials to examine the acquisition of this grammatical constraint more closely.
- 1st person constructions were significantly more likely to be inflected. This was also the most frequent subject, showing (as with the case of irregular verbs), that frequency influences the use of inflected verbs.

6.8 Conclusion

Although some impressionistic claims have been made about periphrastic verbs occurring in more informal, spoken Welsh (King 2016), particularly for north Walian speakers (B. M. Jones 1993), this is the first work to quantitatively examine the repertoire of speakers in different areas and from different home languages. This chapter provides evidence that young people use this more informal variant across all settings of formality, and some style-shifting is

observed, but in opposite directions for the two areas. Although EHL students have not acquired the context constraint of their WHL peers, they do replicate the internal constraints observed, and pattern similarly across grammatical constraints. Irregular verbs are more likely to be used in inflected verb constructions in the Simple Past tense for all groups, although no evidence has been found for this in descriptive grammars. Where WHL speakers inflected more first person constructions, this was also replicated by EHL speakers. Overall, it appears that the use of Simple Past verbs is influenced by frequency more than by style.

7 Intensifiers

7.1 Introduction

This chapter presents an analysis of the use of intensifiers in Welsh. Intensifiers are a form of *graddolion ansoddeiriol* [adjectival modifiers] (Thomas 1996). These modifiers are words such as *iawn* [very] and *eithaf* [quite] which modify the degree or strength of another word (Quirk et al. 1985). The effect can be amplifying, such as *cyflym iawn* [very fast] or diminishing, such as *ychydig yn gymhleth* [a little complicated]. Modifiers can be used to intensify a number of different modified heads (which includes verbs, gradable adjectives, nouns and other adverbs) (Fuchs and Gut 2016: 186). The focus of this analysis is the modification of adjectives, as these are frequently studied in the field of variationist sociolinguistics. Following Reichelt and Durham (2017), I use the term ‘intensifier’ as the top-level term for this type of modifier, and I provide an introduction to lower-level terms later in this chapter. I will compare the speaker groups presented in previous chapters to see whether English home language students vary their use of intensifiers in the same way as the Welsh home language students. I will examine the use of intensifiers across three speech contexts in order to discern patterns of stylistic variation.

There is not much literature on intensifier use in Welsh, particularly in terms of variation. The little literature that does exist focuses on the use of English intensifiers in Welsh; Thomas (1996) alludes to the fact that there may be a language contact effect of English on the use of intensifiers in Welsh, namely that the use of English intensifiers can occur in very informal styles in Welsh speech as borrowings from English, such as *blydi twp* [bloody stupid] (Thomas 1996: 220). This chapter will examine intensifier use in young speakers including the use of borrowed intensifiers. Part of this work will determine the stylistic or linguistic constraints on their use. While some previous researchers (see Tagliamonte and Roberts 2005)

have explored variation in the use vs non-use of intensifiers, similarly to Barnfield and Buchstaller (2010) and Reichelt and Durham (2017), the current work instead examines the patterns of use across speaker groups and speech contexts.

The current chapter will be presented as follows; I will introduce intensifiers in Welsh, then I will present some research on intensifier use in other languages, before turning to the method. The analysis will be presented in two parts; a descriptive account of the patterns of intensifier use according to each of the speaker groups, and then an inferential analysis which lays out the factors which can predict the use of more colloquial variants.

7.2 The feature

Contrary to the relatively well-documented intensifier variation in English, to date, there has been no variationist sociolinguistic analysis of the Welsh intensifier system. Indeed, there are very few studies dealing explicitly with the mechanisms of intensification in language contact situations (Fiorentini and Sansò 2017: 176). Analysing this situation will help to determine to what extent intensification strategies in a minority language are subject to influence from a dominant language and what the patterns of use may be in this process. Whereas intensifiers in English generally precede an adjective (such as ‘very slow’, ‘quite fast’) the placement of intensifiers in Welsh can vary, depending on the intensifier (Thomas 1996). Intensifiers can either follow the adjective (as is the case with *iawn*, below), some precede it (as in English), and some precede it with an *o* or *yn* particle. These are demonstrated in the examples below (taken from the dataset).

Table 7.1: Position of intensifiers in Welsh

Intensifier	Mae’r ddau ohono ni yn <i>agos</i>	The two of us are very <i>close</i> (Fiona)
following	iawn (Fiona)	
<i>adjective</i>		

Intensifier preceding <i>adjective</i>	Fel oedd y miwsig yn really <i>uchel</i> (Elliw)	Like the music was really <i>loud</i> (Elliw)
Intensifier preceding <i>adjective + o</i>	Mae chwythu swigen yn andros o amharchus (Catrin)	Blowing a bubble is terribly <i>disrespectful</i> (Catrin)
Intensifier preceding <i>adjective</i>	Mae grandma fi ar Facebook a mae e bach yn <i>od</i> (Deio)	My grandma's on Facebook and it's a bit <i>odd</i> (Deio)

A full description of each intensifier included in this chapter (along with a brief etymology) is provided in 7.4.4. It should be noted that intensification in Welsh can also occur through repetition of the adjective e.g. *bach bach* (small small, or very small), although no examples of this were found in the current dataset.

7.2.1 Function

Thomas' grammar of Welsh mentions that Welsh modifiers have the capacity to *dwysáu* (intensify) or *gwanhau* (weaken) the adjective (Thomas 1996: 217). The functions of intensifiers are shared across languages to a great extent, therefore many of the following references refer to English, as there are none in Welsh. In Quirk *et al.*'s classification of English intensifiers (1985: 589-603), these functions are termed as amplifiers (scaling upwards from a perceived norm), such as 'really good', and downtoners (scaling downwards from a perceived norm), such as 'a little scary'.

The amplifier function can be further subdivided into maximisers, denoting 'the upper extreme of the scale' (e.g. *hollol wych* [totally great]) and boosters, denoting 'a high point on the scale' (e.g. *mor ddiflas* [so boring]) (Quirk et al 1985: 589-590). Downtoners, on the other

hand, can be subdivided into four subcategories, according to the degree of moderation. Approximators serve ‘to express an approximation’ (e.g. *bron yn amhosibl* [almost impossible]), compromisers have a ‘slight lowering effect’ (e.g. *eithaf nerfys* [quite nervous]), diminishers are synonymous with ‘to a small extent’ (e.g. *bach yn betrysgar* [a little worried]) and minimizers serve as ‘negative maximisers’ (e.g. *prin yn llwglyd* [hardly hungry]) (Quirk et al 1985: 589-590). As I mention in section 7.1, the term I use for all these modifiers, regardless of their function is ‘intensifiers’, following work by Reichelt and Durham (2017). The functions of intensifiers (what Quirk et al. 1985 refer to as modifiers) are summarised in Table 7.2.

Table 7.2: Modifiers adapted from Quirk et al. (1985: 589-603)

Type	Subtype	Example	Explanation
amplifiers	maximisers	<i>hollol</i> [totally]	Denote the upper extreme of the scale
	boosters	<i>mor</i> [so]	Denote a high degree, a high point on a scale
downtoners	approximators	<i>bron</i> [almost]	Express an approximation of the force
	compromisers	<i>eithaf</i> [quite]	Have only a slight lowering effect
	diminishers	<i>bach</i> [a little]	Scales downwards to mean to a small extent
	minimizers	<i>prin</i> [hardly]	Negative maximizers to mean not to any extent

Further to this, König (2017: 17) argues that maximizers and diminishers relate to the validity of a description; boosters express emotional reactions to the extent with which the entity compared exceeds a contextually relevant standard; and compromisers are cautious, metalinguistic assessments of degrees expressing quantitative restrictions.

According to previous work on intensifiers in English and Norwegian, amplifiers have been found to occur more frequently than downtoners, and further that boosters are more frequent than maximisers (Tagliamonte 2003 and Stratton and Sundquist 2022) – even in scripted speech (Reichelt and Durham 2016). It is unknown whether the same pattern can be found in Welsh.

7.2.2 Borrowing

In the context of borrowing intensifiers from English into Welsh, Thomas (1996: 220) observes that in very informal styles, some borrowed adjectives which represent swearwords can modify adjectives, e.g. *blydi twp* [bloody stupid]. However, Geiriadur Prifysgol Cymru (2023) states that borrowed intensifiers are represented beyond the use of swearwords. *Cweit* and *reit* (borrowings from the English *quite* and *right*, respectively) both appear represented by Welsh orthography in dictionaries and grammars. Welsh speakers also make use of English intensifiers, such as *really* and *totally*, in their Welsh speech (e.g. *mae o'n really dda* [it's really good], or *dw i'n totally anghytuno* [I totally disagree]), which do not feature in Welsh grammars and dictionaries. It is possible that this is because these are considered code-switches from English, rather than borrowings. Previous work on language contact situations has shown that discourse pragmatic intensifiers can be switched “in a fairly regular manner”, but that this can depend on their function (Fiorentini and Sansò 2017: 175-177).

7.3 Previous work

The current section will present previous work on intensifiers in languages other than Welsh.

7.3.1 Grammaticalization and overuse

It is widely reported that intensifiers begin as content words before gaining a grammatical function, that is, they become grammaticalized (Kanwit *et al.* 2018). For example, *iawn* is documented in the 12th century as an adjective meaning *right*, *correct* or *true* (Geiriadur Prifysgol Cymru, 2023), but since the 13th century has also been used as an intensifier. New intensifiers emerging from lexical items such as *iawn*, are ‘delexicalised’ and are combined with an increasingly larger number of adjectives (Tagliamonte and Roberts 2005: 284-5).

However, once an intensifier has undergone the process of grammaticalization, it can remain in a state of flux. It has been argued that one of the reasons intensifiers fluctuate in terms of their overall ranking throughout time is because the frequency of use, and in some cases, overuse, eventually leads to a loss in their intensifying effect (Tagliamonte, 2012: 334). As a result, in order to maintain intensifying power, these overused intensifiers are replaced by others (Stratton 2021: 208). No historical account of this process exists for Welsh. However, work by Ito and Tagliamonte (2003) on York English found that throughout the 20th century, *very* was being used less and less frequently, having been replaced by *really*, which had previously been deemed “vulgar” (Tagliamonte 2016: 83). Analysing the use of intensifiers in young people reveals points at which language might be changing.

7.3.2 Social factors

7.3.2.1 Style

One of the social factors thought to influence the use of intensifiers is style. Paradis (2000: 147) claimed that intensifiers can display clear “stylistic differences”. In English, intensifiers have often been associated with casual, non-standard and inventive varieties of speech (Tagliamonte 2016), and previous analyses of intensifier use across different registers (written and spoken) has shown that amplifiers such as *really* and *totally* and compromiser *pretty* occur

exclusively in casual conversation, compared with, for example, academic writing (Biber *et al* 1999). Tagliamonte (2016) reported on the intensifiers used across different registers, i.e. spoken conversation and instant messaging. In this comparison, she found that far higher rates of intensification occurred in the spoken context, but that instant messaging contained more rates of *so* (the emergent boosting intensifier).

As stated, the current work will be the first of its kind to record patterns of intensification across different contexts in Welsh, but considering their association in other languages with colloquial and non-standard varieties of speech (Tagliamonte 2016: 223), it is expected that the rates at which intensifiers are used by the participants in this data will vary between the different speech contexts in which they were recorded. In addition, it is possible that English intensifier codeswitches (such as *really* or *totally*) may be observed more frequently in more informal contexts.

7.3.2.2 *Young speakers*

Paradis (2000) claimed that teenagers may be likely to make more use of strongly reinforcing adjectives, because of their tendency to exaggerate rather than modulate. Connected to the discussion of style, it has also been suggested that teenagers would use more informal modifiers (such as *pretty* and *a bit*) to a greater extent than educated adults (Paradis 2000). Intensifiers are therefore expected to occur relatively frequently in the speech of the young people recorded for this research.

7.3.2.3 *Regional variation*

In English, differences have been found between intensifier use in the USA and the UK. A corpus analysis by Sturdy (2019) found that Americans favoured the use of *pretty* instead of *rather*, while British preferred the inverse. Work by Barnfield and Buchstaller (2010) on Tyneside English, and Ito and Tagliamonte's (2003) work on York English found that British

English speakers favoured *very* over *really* and *so*. On the other hand, Tagliamonte and Roberts' (2005) analysis of the scripted English in the American sitcom *Friends* found this pattern was reversed, with *so* being used most frequently followed by *really*, then *very*. Stereotypes about the differences between these varieties of English have been found to be exaggerated in scripted television, with an overuse of the marked intensifier *bloody* for British characters and an overuse of *totally* and *so* by American characters in *Buffy the Vampire Slayer* (Reichert and Durham 2017: 84).

In their exploration of Spanish booster use in Latin America, Kanwit *et al.* (2018) compared *bien* and *muy* and found that *bien* occurs more frequently in Latin America than in Spain, and that it is successfully acquired by study abroad learners. This implies that rates of intensification can be geographically constrained, meaning that speakers in north Wales could use different patterns to those in south Wales. In addition, if, as this Spanish research found, intensifier patterns are acquired by non-native learners, it could be that EHL background speakers might replicate the patterns of WHL background speakers in this dataset.

7.3.3 Linguistic factors

Research from further afield has found that the quality of the adjective can influence the use of intensifiers. For example, the aforementioned research on the Spanish intensifier system in Argentina and Spain found that *bien* was favoured over *muy* with variably positive adjectives (that is, adjectives that have positive meanings in some but not all contexts) (Kanwit et al 2018: 458). Work on English by Frej and Nam (2014) found that the boosting intensifier *very* collocated most frequently with positive adjectives, compared with the maximiser *too*, which, because of its particular semantic function, is predominantly used with negative adjectives instead. Ito and Tagliamonte (2003) argue that wide intensifier collocation with both positive and negative adjectives is evidence of advanced delexicalization. They present the example (cf Partington 1993: 184), of the intensifier *awfully* which is considered to be more advanced in

220

terms of delexicalization than *terrible*, because the former collocates with modifiers having positive connotations (e.g. good, nice, and glad) as well as negative ones (e.g. cruel and bad), whereas the latter is preferred with negative adjectives.

Frequent collocation of intensifiers with predicative adjectives (as in *mae'n hawdd iawn* [it is very easy]) is argued to be indicative of evidence for a later stage in the delexicalization process of the intensifiers (Ito and Tagliamonte 2003: 271), whereas collocation with only attributive adjectives (as in *gwaith hawdd iawn* [very easy work]) is argued to be indicative of either an outgoing receding variant or the arrival of a novel but latent variant (Stratton and Sundquist 2022).

According to previous literature, therefore, intensifiers that are both positively and negatively collocated, and that are present with more predicative than attributive adjectives could be considered to be further delexicalized and thus more grammaticalized as intensifiers.

7.4 Method

I will now present how the feature appears in my data and the method for coding the variants.

7.4.1 Data extraction

As in previous sociolinguistic studies of intensifiers (Fuchs 2017), a list of intensifiers was compiled from previous academic studies of intensifiers in English and Welsh. Additional intensifiers were identified by searching the CorCenCC corpus (Knight *et al.* 2020) for parts of speech corresponding to modifying intensifiers. This resulted in a list of 56 intensifier variants (see Appendix 11.11) which were used to search the dataset. All extracted tokens from the search were extracted using AntConc and coded in Excel. All tokens that were clearly not functioning as intensifiers were removed (e.g. *car bach* [a small car], *really?* etc). In line with previous sociolinguistic work on intensifiers (Ito & Tagliamonte 2003; Stratton and Sundquist 2022), negative tokens (*ddim yn rhy dda* [not too good]) and comparative tokens (*bach gwell*

[a bit better]) were removed as they can block intensification (D’Arcy 2015: 459). Of the list of 56 identified intensifiers in the literature and in initial searches, 28 intensifiers were present in the data. It is acknowledged that there is sometimes difficulty in interpreting the function of some intensifiers; note for example the ambiguity of the English compromiser *quite* which can in certain varieties and in some syntactic contexts be a booster, too. The classification noted below is based on my own interpretation where the function was ambiguous.

7.4.2 *Intensifier classification*

A total of 28 intensifiers ($n = 1465$) were identified in the dataset, after excluding modified adverbs, negations and comparatives as discussed above. These intensifiers are displayed in Table 7.3 below. Intensifiers in bold are English. Following Quirk et al.’s classification (see Table 7.2), I summarise the classification of the intensifiers found. No examples of approximators or diminishers were found in the dataset. Only intensifiers accounting for 1 percent and more were included in the analysis in order to detect patterns relating to style in the participants’ speech (Reichelt and Durham 2017). These accounted for 94.9% of total intensification ($n = 1391$) and are highlighted in the table above. A total of 5.1% ($n = 74$) intensifiers appeared in less than 1% of the corpus and were excluded from the analysis.⁴⁵

Table 7.3: Classification of intensifiers in the dataset

Scale	Subtype	Intensifier	Gloss	n	%
Amplifier	Maximiser	rhy	too	47	3.2
		hollol	totally	29	2.0
		cwbl	completely	13	0.9
		completely	completely	3	0.2
		totally	totally	1	0.1
	Booster	really	really	475	32.4
		iawn	very	197	13.6
		mor	so	113	7.7
		wir	truly	8	0.5
		very	very	6	0.4

⁴⁵ Examples from the dataset along with a brief etymology of the frequent intensifiers (in grey in the table) will be provided in section 7.4.4

		proper	proper	3	0.2
		hynod	remarkably	3	0.2
		mega	mega	2	0.1
		well	well	1	0.1
		so	so	1	0.1
		andros	terribly	1	0.1
Downtoner	Compromisers	eithaf	quite	313	21.4
		cweit	quite/very	33	2.3
		reit	quite/very	22	1.5
		digon	enough	19	1.3
		go	rather	13	0.9
		gweddol	fairly	6	0.4
		pretty	pretty	6	0.4
		braidd	rather	1	0.1
		cymharol	comparatively	1	0.1
	Diminishers	bach	a bit	143	9.8
		ychedig	a little	4	0.3
		cryn	somewhat	1	0.1

7.4.3 Frequent intensifiers

The most frequent intensifier was the borrowed booster *really*, which made up 32.4% of the intensifier system, followed by the compromiser *eithaf* [quite] in second position (21.4%), and the booster *iawn* [very] in third position (13.7%).

Because there are differences with respect to word counts across contexts, and intensifiers are presumed to be used less frequently in more formal settings, the data were normalised and intensifier use per 10,000 words will be reported alongside the proportion of use for each intensifier. Overall, a total of 54.1 intensifiers were found to occur per 10,000 words. These overall distributions are presented in the table below.

Table 7.4: Intensifier rate per 10,000 words

Intensifier	Gloss	Rate per 10,000 words
really	<i>really</i>	18.3
eithaf	<i>quite</i>	12.0
iawn	<i>very</i>	7.7
bach	<i>a bit</i>	5.6
mor	<i>so</i>	4.4

rhy	<i>too</i>	1.8
cweit	<i>quite</i>	1.2
hollol	<i>totally</i>	1.1
reit	<i>quite</i>	0.8
digon	<i>enough</i>	0.7
Total		54.1

7.4.4 The feature in the data

Here (the next page), I present the frequent intensifiers from the dataset, in order of function subheadings below, along with a brief etymology, where available. Examples are drawn from the dataset, to show each intensifier in use in the speech of the participating young Welsh speakers.

Table 7.5: Etymology of frequent intensifiers in the dataset

Maximisers	<i>hollol</i>	[totally]	<p>GPC (2023) reports written examples of <i>hollol</i> being used as an adjective to mean ‘whole’ or ‘complete’ since the 12th century⁴⁶ and although no historical record was found of when it first appeared as serving a maximising function, the following example demonstrates its use in the speech of young people.</p> <p style="text-align: right;"><i>mae 'n hollol normal</i> [it’s totally normal] Nel peer to peer</p>
	<i>rhy</i>	[too]	<p>The maximiser <i>rhy</i> is attested in GPC (2023) as early as the 12th century, but has cognates across Celtic languages, so was inherited and not innovated within Welsh. An example of its modern use can be seen in the example below.</p> <p style="text-align: right;"><i>fel yn rhy addicted</i> [like too addicted] Naomi sociolinguistic interview</p>
Boosters	<i>iawn</i>	[very]	<p><i>Iawn</i> is the only intensifier in Welsh to occur after the adjective. According to Thomas (1996: 210), <i>iawn</i> can precede or follow another adjective, depending on the meaning of the phrase. For example, <i>yn ei iawn bwyll</i> [in his right mind], and <i>araf iawn</i> [very slow]. GPC attests the noun (meaning <i>right, correct</i> or <i>true</i>) from the thirteenth century, the adjective and adverb from the twelfth century (Geiriadur Prifysgol Cymru, 2023), and its position may help distinguish it from its noun and adjective form which is still in use today.</p> <p style="text-align: right;"><i>dw i 'n meddwl bod o 'n bwysig iawn</i> [I think it’s very important] Elinor peer to peer</p>

⁴⁶ Thomas (1996: 218) also references the use of *hollol* as a maximising intensifier, which can also be doubled to create emphasis. No doubled examples were found in the dataset.

	<i>mor</i>	[so]	<p>The use of <i>mor</i> before an adjective in the positive degree is recorded as early as the 9th century in Welsh (GPC 2023), which is far earlier than the use of <i>so</i> as an intensifier in English, reported by Stoffel (cf. Tagliamonte and Roberts 2005: 283) to have emerged in the early 1900s. Records end of <i>mor</i> being used to intensify in GPC after 1688, but it is thought to be prevalent still today.</p> <p style="text-align: right;"><i>bydd hynna mor cŵl</i> [oh that will be so cool] Sally sociolinguistic interview</p>
	<i>really</i>		<p>It is unknown when <i>really</i> started being used in Welsh speech, and there is no record of it in GPC as it is not recognised as a borrowing or a loan word; it is a codeswitch. In the National Corpus of Contemporary Welsh (CorCenCC), <i>really</i> has been transcribed using the English spelling, but also using the Welsh spelling, <i>rili</i>, reflecting the pronunciation and morphology of it being embedded into Welsh discourse (for a similar case see Peterson 2017:123 about <i>pliis</i> in Finnish). <i>Really</i> has been found to be one of the most frequent markers of intensity in American English colloquial conversation (Labov 1984: 44), as well as among younger speakers in the British National Corpus (Lorenz 2002, 153), and in the city of York, (Ito and Tagliamonte 2003) (see Tagliamonte and Roberts 2005: 283). It is thought to have emerged as an intensifier in the 19th century and its function is to boost or strengthen the adjective it modifies.</p> <p style="text-align: right;"><i>dweud pethau really gogleddol</i> [saying really northern things] Deio sociolinguistic interview</p>
Downtoners	<i>cweit</i>	[quite]	<p><i>Cweit</i> is recognised as a borrowing from the English <i>quite</i> and is defined as an adverb (Geiriadur Prifysgol Cymru, 2023). It has been noted that ‘quite’ can mean more than ‘a little’ but less than ‘very’ (Su 2016: 225), and Quirk <i>et al</i> classified ‘quite’ as a compromiser (1985: 590). All cases of <i>cweit</i> in the data served a downtoner function. This intensifier was pronounced as ['kwart] (the expected English pronunciation) and ['kwert] (the expected Welsh pronunciation) by different speakers (regardless of context or language of proceeding adjective). Some Gwynedd speakers (Maddie, Scarlett, Lowri, Awen, Mathew) used the ['kwart] pronunciation, and some used a</p>

			<p>glottalized t ['kwai/?/] for both and ['kwei/?/]. The difference in pronunciation was not always perceptually salient with this intensifier, sometimes caused by poor audio quality or speakers talking at the same time.</p> <p style="text-align: right;"><i>ro'n i'n wneud hynna cweit aml</i> [I used to do that quite often] Mathew sociolinguistic interview</p>
	<i>digon</i>	[enough]	<p>According to GPC (2023), <i>digon</i> appeared first as an adverb in the 14th century, having first appeared in the 13th as a masculine noun. The example below shows how it is used as a downtoner to weaken the adjective by one speaker.</p> <p style="text-align: right;"><i>ro'n i'n ddigon hapus</i> [I was happy enough] Maddie job interview</p>
	<i>eithaf</i>	[quite]	<p><i>Eithaf</i> is inherited within Celtic, but has a cognate formation in the Latin <i>extimus</i> (Geiriadur Prifysgol Cymru, 2023). As a noun meaning <i>extreme</i> or <i>ultimate</i>, it appears in writing in the 9th century, but as a superlative adjective (highest, uttermost), it is reported to date back to the 12th century. Today, <i>eithaf</i> is a downtoning adjective meaning ‘quite’ (see <i>cweit</i> above). The example below presents its use in young speakers of Welsh today.</p> <p style="text-align: right;"><i>fi'n berson eithaf erm cymdeithasol</i> [I’m quite a erm sociable person] Nel job interview</p>
	<i>reit</i>	[quite]	<p><i>Reit</i> [quite] has been identified as a previous borrowing from English right (Geiriadur Prifysgol Cymru). Its translation is quite (a downtoner) or very (an amplifier), even though in English right is predominantly considered a booster in English, rather than a moderator. In the current data, all cases of <i>reit</i> serve to weaken the adjective, and have thus been coded as moderators. One such example can be seen below.</p> <p style="text-align: right;"><i>cadw fo'n reit syml</i> [kept it quite simple]⁴⁷ Nel peer to peer</p>

⁴⁷ This could also be interpreted as a booster ‘keep it really simple’.

	<i>bach</i>	[a little]	<p>Written records of the adjective <i>bach</i> (small or little) being innovated exist from the 14th century (Johnston 2020: 73-75). It is also used in combination with <i>ychydig</i> to form the fuller intensifier <i>ychydig bach</i> (a little bit) which can also occur as an intensifier in Welsh (though not frequent in this dataset). In this analysis, <i>bach</i> is a downtoner, and an example from the dataset can be seen below.</p> <p style="text-align: right;"><i>ie - mae e bach yn eithafol</i> [yeah – it is a little extreme] Naomi peer to peer</p>
--	-------------	------------	--

7.4.5 Social factor coding procedure

The use of intensifiers has been widely reported to be stylistically marked in other languages, however little is known about intensifier use in Welsh. As in the previous chapters, the use of intensifiers will be mapped against three speech contexts of varying formality, for four distinct groups of speakers with either English or Welsh as a home language, and who attend Welsh-medium education in either Gwynedd or Cardiff. The number of intensifiers by speaker group is presented below.

Table 7.6: Intensifiers by speaker group

Speaker group	Number of intensifiers
GWHL	328
GEHL	167
CWHL	412
CEHL	484
Total	1391

7.4.6 Internal factor coding procedure

Internal factors were also coded based on the literature relating to the intensifier feature (presented in 7.3.3), as follows.

7.4.6.1 Intensification function

As mentioned in 7.2.1, intensifiers can serve to amplify or downtone an adjective (Quirk *et al.* 1985: 589-590), and the subtypes of intensifiers have been presented with examples from the dataset in this subsection. It is noted by Quirk *et al.* (1985: 590) that these are only a rough guide to semantic gradients, and these categories often overlap in ways that are not always

readily distinguishable (Reichelt and Durham 2017: 65). each intensifier was coded based on the context of the utterance. Whereas the literature presented in 7.4.4 for *cweit* and *reit* show that its function can be to heighten the adjective to varying degrees, these were combined with the compromisers, taking the view that ‘quite’, although complex, generally means more than *a little* but less than *very* (Su 2016: 225).

7.4.6.2 Language of the adjective

A total of 386 adjectives were intensified in the dataset. Of those, 171 (44.3%) were English adjectives, and 215 (55.7%) were Welsh adjectives, which is evidence of a great deal of codeswitching in the dataset. English adjectives were found to be more likely to be intensified than Welsh adjectives, as can be seen in Table 7.7, however this could be due to the relative frequency of Welsh adjectives overall.

Table 7.7: Rates of intensification with English and Welsh adjectives

Adjective language	Number of intensifiers	%
English adjectives	276	61.9%
Welsh adjectives	1115	19.2%
Total	1391	

By way of demonstrating the adjectives most likely to be intensified in the dataset, Table 7.8 presents some of the adjectives that occur most frequently in the modification data (those accounting for more than 1% of intensification). In the following table, the only two English adjectives were *weird* and *annoying*, marked in bold. The other English adjectives were used infrequently (in some cases only once).

Table 7.8: Most frequently intensified adjectives

Adjective	Gloss	Frequency	<i>N</i>
da	good	13.73%	191
anodd	hard	3.88%	54
neis	nice	3.38%	47

od	odd	3.16%	44
cŵl	cool	2.73%	38
diddorol	interesting	2.66%	37
gwahanol	different	2.01%	28
bach	small	1.87%	26
tebyg	similar	1.80%	25
agos	close	1.65%	23
pwysig	important	1.44%	20
nerfus	nervous	1.29%	18
hawdd	easy	1.22%	17
anffurfiol	informal	1.15%	16
doniol	funny	1.15%	16
hapus	happy	1.08%	15
weird	weird	1.08%	15
annoying	annoying	1.08%	15
gwael	bad	1.08%	15
Cymraeg	Welsh	1.01%	14
cryf	strong	1.01%	14
hir	long	1.01%	14

7.4.6.3 *Quality of the adjective*

The level of delexicalization in intensifiers is directly related to their collocational behaviour. Intensifiers that are more delexicalized tend to have a wider range of collocations. For instance, in English, *awfully* is highly delexicalized because it collocates with both positively and negatively connotated modifiers, while *terribly* exhibits a slight preference for negative modifiers, suggesting it is less delexicalized (Ito and Tagliamonte 2003). Recall that previous research on the Spanish intensifier system in Argentina and Spain found that *bien* was favoured over *muy* with variably positive adjectives (Kanwit et al 2018: 458). In order to ascertain the extent to which adjective quality could impact the intensifier system in Welsh, all adjectives were coded as follows:

- invariably positive (e.g. *pert* [pretty]), adjectives that can only have a positive connotation

- variable positive (e.g. *mawr* [big]), adjectives which can be positive or negative, depending on context
- neutral (e.g. *tebyg* [similar]), adjectives which could be positive or negative, and context was ambiguous.
- variable negative (e.g. *bach* [small]), adjectives which can be positive or negative, depending on context
- invariably negative (e.g. *hyll* [ugly]), adjectives that can only have a negative connotation

7.4.6.4 *Function of the adjective*

Here, I discuss the function of the adjective (predicative or attributive) as an influencing factor on the production of intensifiers. By attributive agreement, I refer to examples such as *ci bach* [a small dog] above, while an example of predicative use is *mae'r ci yn fach* [the dog is small] (Meelen and Nurmio 2020: 2). The use of intensifiers with predicate adjectives (such as *mae'r ci yn fach iawn* [the dog is very small]) has been taken as evidence for the arrival of a novel but latent variant at a later stage in the delexicalization process (Ito and Tagliamonte 2003), or that the intensifier is an outgoing receding variant (Stratton and Sundquist 2022). Higher frequency of use with attributive adjectives, on the other hand (as in *ci bach iawn* [a very small dog]), denotes that the intensifier is less far along the process of delexicalization, and is thus less grammaticalized. We can test for this by examining the distribution of intensifiers according to whether they occurred with attributive adjectives or with predicative adjectives (as in Ito and Tagliamonte 2003: 271).

7.5 Results (distributional analysis)

In this section, I present the descriptive findings of the factors thought to influence the use of intensifiers, starting with function.

7.5.1 Overall distribution of function and context

The first factor I consider here is the function of the intensifier. As mentioned previously, the intensifying modifier indicates a point on a scale of intensity, which can either scale upwards (amplifiers) or downwards (downtoners) from an assumed norm (Quirk *et al.* 1985: 590). As mentioned in 7.4.6, the subtypes of downtoners often overlap in ways that are not always readily distinguishable, and thus, they have been combined into one distinct category of moderator, which, in contrast with maximisers and boosters, will give a general idea of the distribution and patterns of intensifier use. The intensifiers discussed here broadly fit into three categories; boosters, maximisers and moderators. This follows previous work on intensifiers (Reichelt and Durham 2017). The rates (per 10,000 words) of the different functions are presented in the table and figure below. Boosters are most common in the dataset, followed by moderators, followed by maximisers.

Table 7.9: Rate per 10,000 words of intensifier function

Function	Rate per 10,000 words
Booster	30.54
Moderator	20.52
Maximiser	3.07
Total (N)	1391

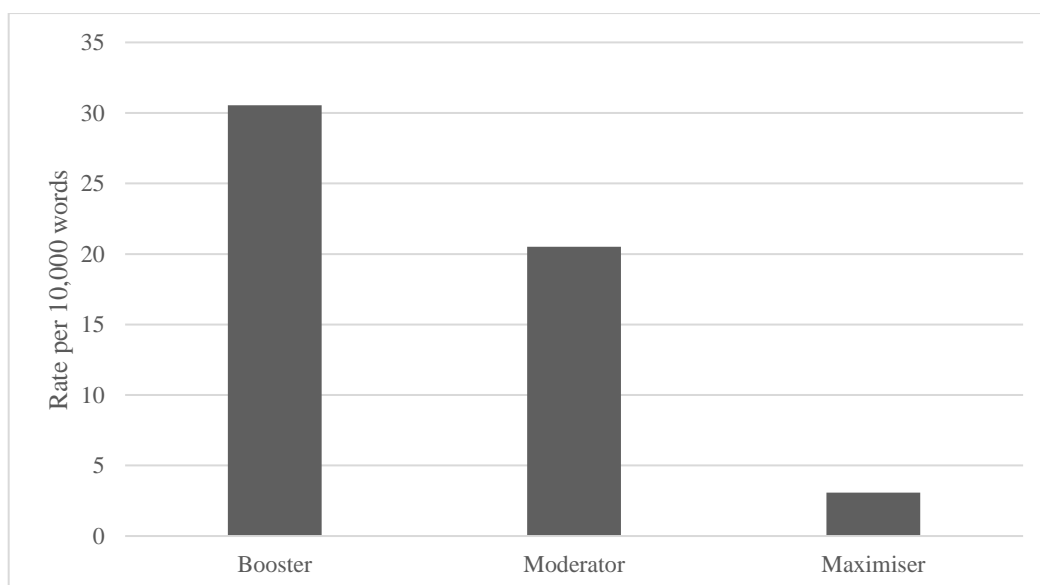


Figure 7.1: Rate per 10,000 words of intensifier by function

In natural speech, boosters have been referred to as ‘the main type of intensifiers ... with moderators and maximisers used less frequently’ (Reichelt and Durham 2017: 83). This was the case in my data too, across all contexts. As expected, the job interview context had the least intensification overall. Boosters were most common in the sociolinguistic interview. The most moderators occurred in the peer to peer context, which occurred twice as often as in the job interview context. Similar rates of maximisers were used across contexts. This is broken down by intensifier in the following section.

Table 7.10: Intensifier function in the three speech contexts per 10,000 words

Context	Booster	Maximiser	Moderator
job interview	5.26	0.80	4.29
peer to peer	10.86	1.20	8.52
sociolinguistic interview	14.43	1.00	7.71

In order to provide a view of overall distributions (before breaking down by speaker group), Table 7.11 shows the intensifiers in each context (per 10,000 words).

Table 7.11: Intensifier rate per 10,000 words by context

Context	Boosters			Maximisers		Moderators					Total
	<i>really</i>	<i>iawn</i>	<i>mor</i>	<i>hollol</i>	<i>rhy</i>	<i>eithaf</i>	<i>bach</i>	<i>digon</i>	<i>quite</i>	<i>reit</i>	
job interview	2.26	2.46	0.54	0.23	0.6	3.22	0.77	0.27	0.04	0	10.36
peer to peer	5.75	3.91	1.19	0.46	0.8	4.29	2.46	0.27	0.65	0.8	20.60
sociolinguistic interview	10.32	1.42	2.69	0.49	0.5	4.49	2.46	0.19	0.58	0	23.17

The highest rates of intensification can be seen in the sociolinguistic interview, where intensifiers occur in 23.17 times in 10,000 words, followed by the peer-to-peer speech context (20.60) and the job interview context (10.36). In the case of most intensifier variants, the rates of use increase from the more formal job interview to the more informal sociolinguistic interview. However, rates of *iawn*, *rhy*, and *digon* decrease in the sociolinguistic interview. In the job interview context, *eithaf* is the most common intensifier per 10,000 words (3.22), followed by *iawn* (2.46), followed by *really* (2.26). In the peer-to-peer context, *really* is most frequent per 10,000 words (5.75), followed by *eithaf* (4.29), followed by *iawn* (3.91). In the sociolinguistic interview, *really* occurs 10.32 times per 10,000 words, followed by *eithaf* (4.49), followed by *mor* (2.69). Following subsections will investigate to what extent different speaker groups use different types of intensifier, as well as exploring linguistic constraints on their patterns of use. As shown previously, maximizers, boosters and moderators have completely different sets of variants and may not be affected by the factors in the same way; I thus present each function in turn separately, rather than together. I begin with an analysis of moderators.

7.5.2 Moderators

Here I present the factors thought to influence the use of 5 frequent moderators in the dataset, *eithaf*, *bach*, *digon*, *cweit* and *reit*. For this set of intensifiers, all the variants are Welsh, but

cweit and *reit* are influenced by English. The first factor under consideration is the social factor, context.

7.5.2.1 *Speech context*

The proportions for each intensifier are calculated within each speaker group in Table 7.12 and presented in Figure 7.2.

Table 7.12: Proportionate use of moderators in different contexts

	eithaf %	bach %	digon %	cweit %	reit %	Total (N)
GWHL	39	26	5	13	16	97
job interview	58	25	13	4	0	24
peer to peer	29	18	4	18	31	51
sociolinguistic interview	41	45	0	14	0	22
GEHL	41	16	0	33	10	58
job interview	92	8	0	0	0	12
peer to peer	24	14	0	33	29	21
sociolinguistic interview	32	20	0	48	0	25
CWHL	62	30	8	-	-	146
job interview	79	9	12			34
peer to peer	48	45	7			60
sociolinguistic interview	67	27	6			52
CEHL	70	28	1	0	-	229
job interview	80	20	0	0		40
peer to peer	72	26	1	1		88
sociolinguistic interview	64	34	2	0		101

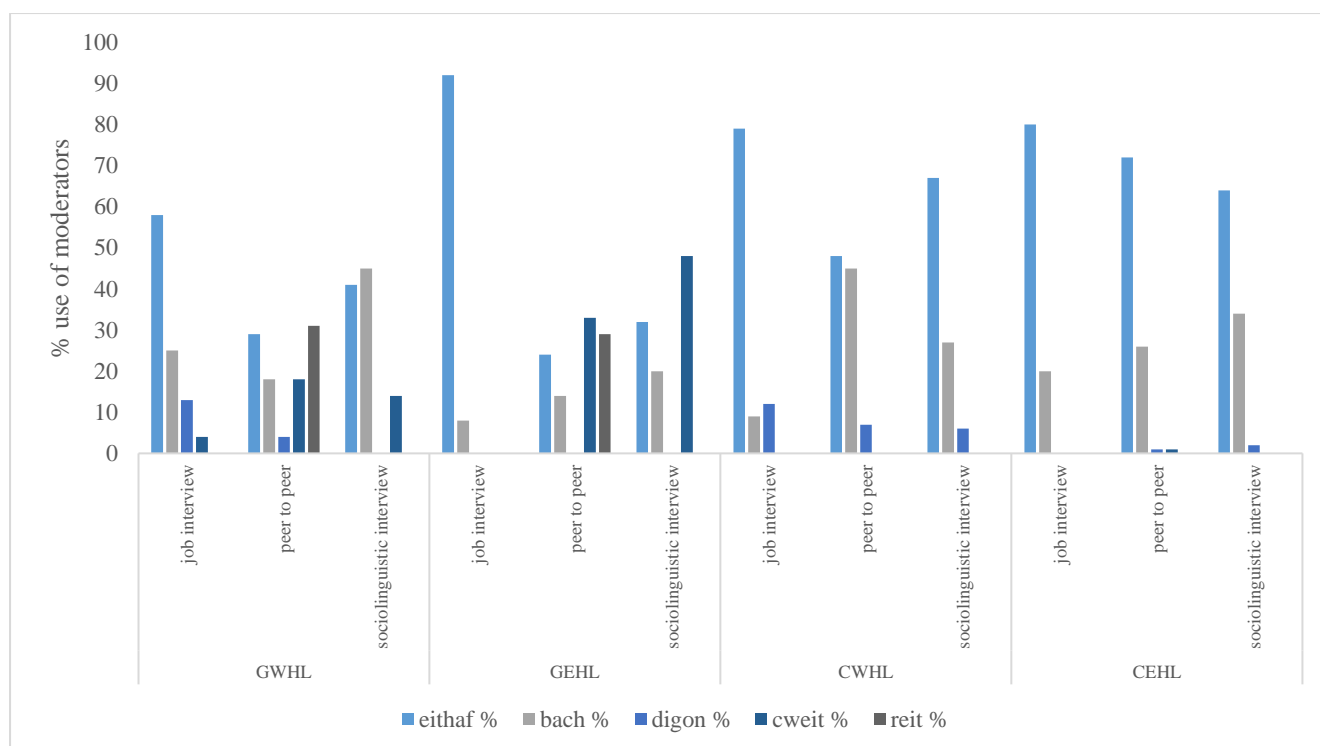


Figure 7.2: Proportionate use of moderators in different contexts

The biggest observable difference between different speaker groups is that only Gwynedd groups use *cweit* and *reit* to any extent. Cardiff groups do not use these two moderators (aside from one token in the peer-to-peer context), and it is clear that EHL participants in both localities are acquiring these local patterns of use for those two moderators.

For all groups, *eithaf* is favoured in all contexts over other moderators, apart from the GWHL sociolinguistic interview context, which contained higher rates of *bach* (45%). For this group, *digon*, pertained mostly to the more formal job interview context (13%). On the other hand, the occurrence of *cweit* and *reit* was minimal in the more formal job interview context.

The GEHL group showed similar patterns; that is, the rates of *bach* increase in more informal contexts. The moderator *eithaf* appears most frequently in the job interview (92%), but *cweit* and *reit* were the most frequent moderators used in the peer-to-peer and sociolinguistic interview registers. A strong Gwynedd pattern emerges here; *eithaf* is clearly

favoured in the formal job interview context, but more informal speech context prompt the use of *bach*, *cweit* or *reit*.

As mentioned above, the Cardiff groups appear to have a narrower repertoire of moderators, opting to use *eithaf* and *bach* in the majority of cases. As a consequence, they show a stronger preference for *eithaf* over other moderators than the Gwynedd groups, but the patterns of use are similar to their Gwynedd counterparts. For CWHL, *eithaf* occurs in the job interview in 79% of cases of moderator use, but in the peer-to-peer context, rates of *eithaf* and *bach* were very similar. The patterns for CEHL are even clearer, the use of *eithaf* decreases as the speech context becomes more informal, and the rates of *bach* increase. In Cardiff, therefore, whereas *eithaf* is favoured in all contexts, *bach* occurs increasingly in more informal contexts.

7.5.2.2 Adjective language

Here, I consider the use of moderators with English and Welsh adjectives (see Table 7.13 and Figure 7.3).

Table 7.13: Proportionate use of moderators with Welsh or English adjectives

	eithaf %	bach %	digon %	cweit %	reit %	N
GWHL	39	26	5	13	16	97
English	12	46	0	38	4	26
Welsh	49	18	7	4	21	71
GEHL	41	16	0	33	10	58
English	11	39	0	39	11	18
Welsh	55	5	0	30	10	40
CWHL	62	30	8	0	0	146
English	32	68	0	0	0	38
Welsh	73	17	10	0	0	108
CEHL	70	28	1	0	0	229
English	47	53	0	0	0	49
Welsh	76	22	2	1	0	180

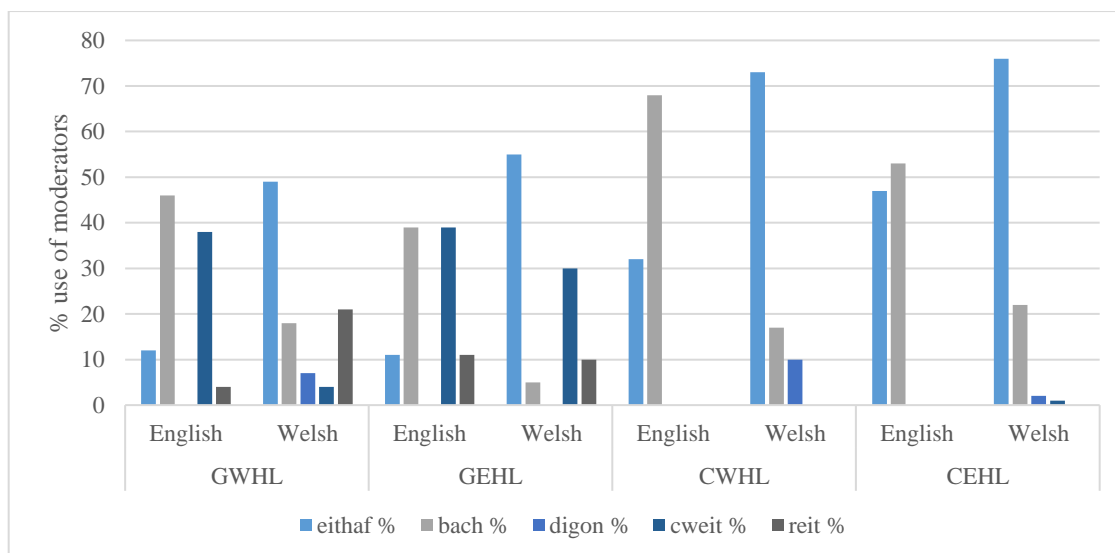


Figure 7.3: Proportionate use of moderators with Welsh or English adjectives

For the GWHL group, a clear pattern can be observed in the proportionate use of moderators with Welsh and English adjectives. When using English adjectives, these participants favoured the use of *bach* or *digon* (e.g. *bach yn weird* [a bit weird]), whereas with Welsh adjectives, *eithaf* was favoured (e.g. *eithaf rhyfedd* [a bit strange]). The GEHL also use this pattern with moderators. A similar pattern is also observed in Cardiff, with CWHL speakers favouring the use of *bach* with English adjectives, and *eithaf* with Welsh adjectives. CEHL follow their CWHL counterparts in their use of *eithaf* and *bach*, however *eithaf* is still prevalent when using English adjectives.

7.5.2.3 Adjective quality

In order to ascertain the extent to which adjective quality could impact the intensifier system in Welsh, all adjectives were coded as invariably positive, variable positive, neutral, variable negative and invariably negative (see section 7.4.6). The *Ns* for this factor were deemed too low to be able to ascertain any discernible patterns for moderators. I therefore turn to the final linguistic factor for moderators; adjective function.

7.5.2.4 *Adjective function*

This next subsection presents the analysis of adjective function (attributive or predicative) with intensifiers. If an intensifier collocates with more predicative than attributive adjectives, it is further delexicalized and thus more grammaticalized as an intensifier. It is unknown whether differences in the grammaticalization of moderators will be observed between speaker groups. Table 7.14 presents the use of moderators with attributive and predicative adjectives. Overall, there were far fewer attributive constructions than predicative.

Table 7.14: Proportionate use of moderators with attributive or predicative adjectives

	eithaf %	bach %	digon %	cweit %	reit %	Total (N)
GWHL	39	26	5	13	16	97
attributive	36	0	0	29	36	14
predicative	40	30	6	11	13	83
GEHL	41	16	0	33	10	58
attributive	43	0	0	57	0	7
predicative	41	18	0	29	12	51
CWHL	62	30	8	0	0	146
attributive	94	0	6	0	0	17
predicative	58	34	8	0	0	129
CEHL	70	28	1	0	0	229
attributive	100	0	0	0	0	28
predicative	66	32	1	0	0	201

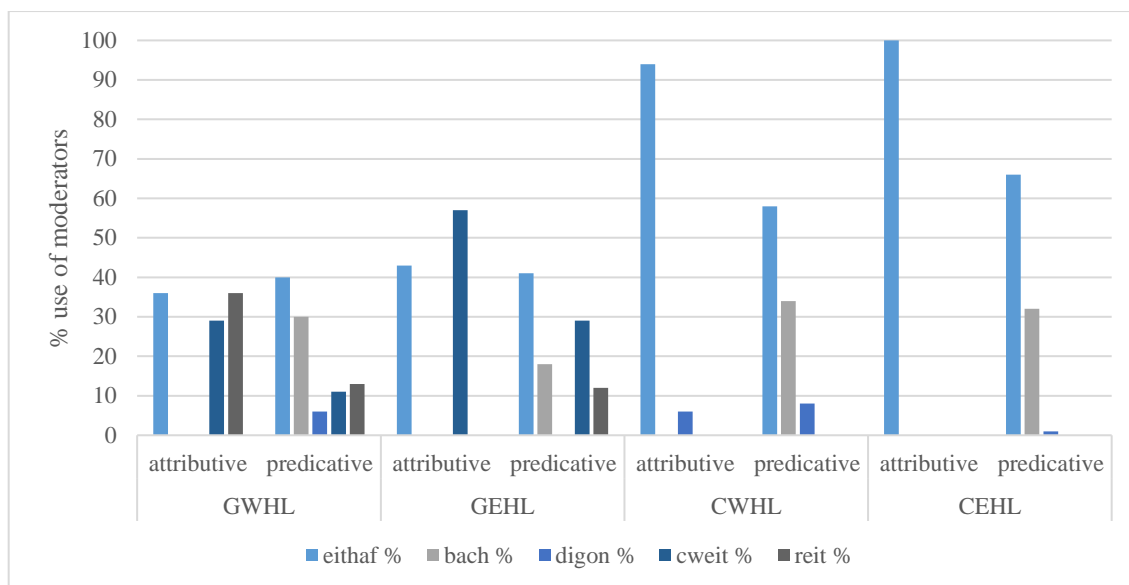


Figure 7.4: Proportionate use of moderators with attributive or predicative adjectives

For the GWHL speaker group, the most frequently used intensifier with predicative adjectives was *eithaf* (40%), followed by *bach* (30%). Although rates were lower, *eithaf* was found to be used with attributive adjectives to a similar extent (36%). *Bach* only appeared with predicative function adjectives, as did *digon*, which suggests that these moderators may be further along in the process of grammaticalization. It is recognised that *bach* is a replica grammaticalization of the English *(a) little*, so the process of grammaticalization is not necessarily happening in Welsh (Heine & Kuteva 2005). On the other hand, both *cweit* and *reit* were used with far more attributive than predicative adjectives.

Similar patterns are observed in GEHL speakers; their rates of use of *eithaf* collocated with attributive and predicative adjectives to a similar extent, although rates of attributive constructions were much lower. The moderators *bach* and *reit* only appeared in predicative constructions. As with their GWHL counterparts, GEHL speakers used *cweit* far more with attributive than predicative adjectives.

For Cardiff speakers from both home language backgrounds, *eithaf* is used more frequently with attributive than predicative adjectives. Rates of use of *bach* are similar for both groups too, and both only use *bach* with predicative adjectives.

In summary, although *eithaf* is frequent in Cardiff, it occurs more often with attributive than predicative adjectives. As discussed in 7.3.3, the use of intensifiers with predicate adjectives could be taken as evidence for a later stage in the delexicalization process. Higher frequency of use with attributive adjectives, on the other hand, could suggest that *eithaf* is a receding variant (Ito and Tagliamonte 2003: 271). The same rationale could suggest that *cweit* could be undergoing a similar process in Gwynedd. The case of *bach* is interesting; *bach* occurs with an *yn* particle and is equivalent to the English ‘a little bit’ or ‘a bit’. In English, this intensifier only occurs in predicative constructions, and it appears as though the pattern is followed in the current Welsh data, although no literature has been found which discusses this as a constraint on its use. Indeed, whereas the English phrase ‘he is a bit strange man’ is ungrammatical, in Welsh, *mae e’n ddyn bach yn rhyfedd* is not, but it is still unused in this dataset.

7.5.2.5 Summary of moderator descriptives

For all groups, *eithaf* appears to be most appropriate in formal settings. Geographical differences were observed, however. Cardiff mainly use *eithaf* and *bach* (*bach* increasingly in more informal settings), whereas Gwynedd have a wider repertoire of moderators and use more *bach*, *cweit* and *reit* (than *eithaf*) in more informal contexts. *Eithaf*, the more formal variant, occurs mostly with Welsh adjectives, whereas both groups use more of their local informal variants with English adjectives, as anticipated. Lower rates of delexicalization occurred with *eithaf* in Cardiff, as was the case in Gwynedd with *cweit*, suggesting that these may be receding variants in the respective areas. In the case of all factors, EHL groups are replicating the WHL trends in their localities.

7.5.3 Boosters

Next, I present the results for the boosting intensifiers. As mentioned in 7.4.3 (and in line with other work on intensifiers in other languages), the boosters were the most common type of intensifier in the data. The three most common boosters identified in the data were *really*, *iawn* [very] and *mor* [so]. *Really* is an English borrowing, whereas the other two are Welsh intensifiers. As with the previous section on moderators, their use will be presented by factor under the following subheadings.

7.5.3.1 Speech context

The use of boosters by context are presented in Table 7.15.

Table 7.15: Proportionate use of boosters in different contexts

	really %	iawn %	mor %	Total (N)
GWHL	74	14	11	214
job interview	70	15	15	46
peer to peer	67	24	9	55
sociolinguistic interview	80	10	11	113
GEHL	66	23	11	105
job interview	44	50	6	18
peer to peer	59	36	5	22
sociolinguistic interview	74	11	15	65
CWHL	61	31	8	233
job interview	30	63	7	46
peer to peer	54	39	7	100
sociolinguistic interview	85	5	10	87
CEHL	45	30	25	233
job interview	21	71	8	24
peer to peer	44	40	17	103
sociolinguistic interview	52	11	37	106

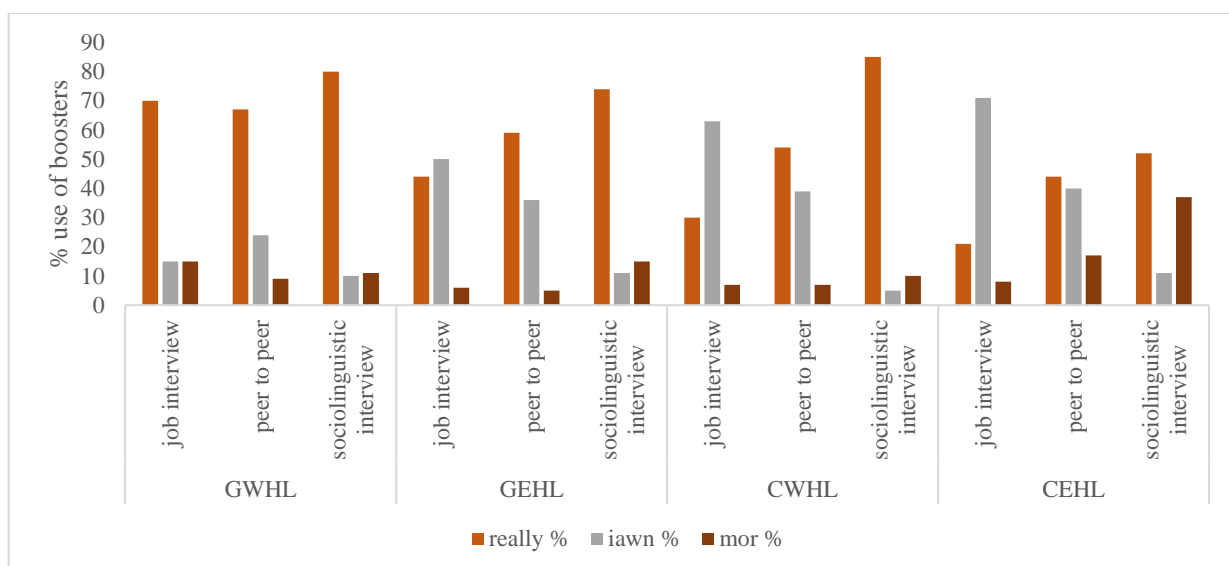


Figure 7.5: Proportionate use of boosters in different contexts

Clear patterns emerge in the data. Overall, *really* and *mor* increase as the speech context becomes more informal, and rates of *iawn* decrease accordingly. This pattern is shared across all groups apart from GWHL. Whereas GEHL, CWHL and CEHL favour the use of *iawn* in the job interview over *really*, the GWHL group use the informal, English-borrowed *really* in this more formal context. The frequency of use of *really* for the GWHL, even in more formal contexts, could indicate that this borrowing is not considered markedly informal by this group. Indeed, it appears as though for GWHL speakers, this booster is firmly embedded in their intensifier repertoire, unaffected by stylistic constraints. The GEHL group do not fully replicate this pattern, and instead rely on the more formal *iawn* variant in formal settings. For the GWHL speaker group, the highest rates of use of *iawn* occur in the informal peer to peer context. CEHL have the highest rates of *mor* in the more informal speech contexts.

7.5.3.2 Adjective language

Although little evidence exists about this phenomenon, it is presumed that intensifiers which are established as Welsh intensifiers (such as the boosters *mor* and *iawn*) may be more likely to be used with Welsh adjectives, and those which are loaned or borrowed from English (such

as *really*) may be more likely to be used with English adjectives. Overall, in the booster data (Table 7.16 and Figure 7.6), *really* is collocated most often with English adjectives rather than Welsh adjectives.

Table 7.16: Proportionate use of boosters with Welsh or English adjectives

	really %	iawn %	mor %	Total (N)
GWHL	74	14	11	214
English	77	6	17	47
Welsh	74	17	10	167
GEHL	66	23	11	105
English	69	12	19	26
Welsh	65	27	9	79
CWHL	61	31	8	233
English	93	0	7	30
Welsh	56	35	8	203
CEHL	45	30	25	233
English	48	26	26	31
Welsh	45	31	25	202

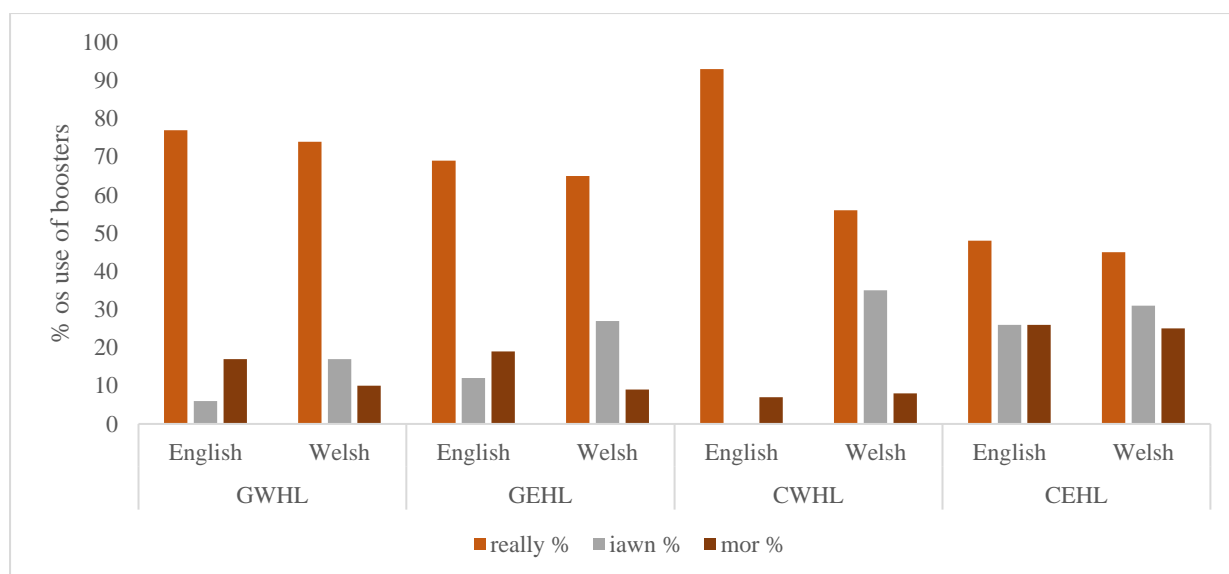


Figure 7.6: Proportionate use of boosters with Welsh or English adjectives

On the other hand, *iawn* is used less with English adjectives than Welsh adjectives across all speaker groups, particularly with CWHL who only use *iawn* with Welsh adjectives. Interesting patterns emerge in the *mor* and *really* data. It appears that Gwynedd speakers (particularly GEHL) seem more likely to use English adjectives with *mor* than their Cardiff counterparts, whose rates of use of English and Welsh adjectives with *mor* are comparable.

A clear trend emerges from the data here, that *iawn* is mostly used in conjunction with Welsh adjectives, and in Gwynedd, *mor* occurs most often with English adjectives. All groups use *really* with adjectives of both languages, but CWHL use a much higher proportion of English adjectives with this intensifier, which is not fully replicated in CEHL speech. Otherwise, EHL speakers seem to follow the same patterns as their WHL counterparts, in both areas.

7.5.3.3 Adjective quality

Next, I turn to present the boosting intensifiers; *iawn* [very], *mor* [so] and *really* [really] and their use with adjectives of varying quality. Although the quality of the adjective was coded as invariably negative, variable negative, neutral, variable positive and invariably positive, these

were grouped into three categories due to low token counts; negative, positive and neutral rates are presented in Table 7.17 (results are also displayed in Figure 7.7).

Table 7.17: Proportionate use of boosters with negative or positive adjectives

	really %	iawn %	mor %	Total (N)
GWHL	75	14	11	213
negative	74	6	20	50
neutral	77	20	3	30
positive	74	16	10	133
GEHL	66	22	12	104
negative	68	13	19	47
neutral	86	0	14	7
positive	62	34	4	50
CWHL	63	29	8	226
negative	76	11	13	79
neutral	65	15	20	20
positive	54	42	4	127
CEHL	47	28	26	225
negative	62	10	28	90
neutral	59	18	24	17
positive	33	42	25	118

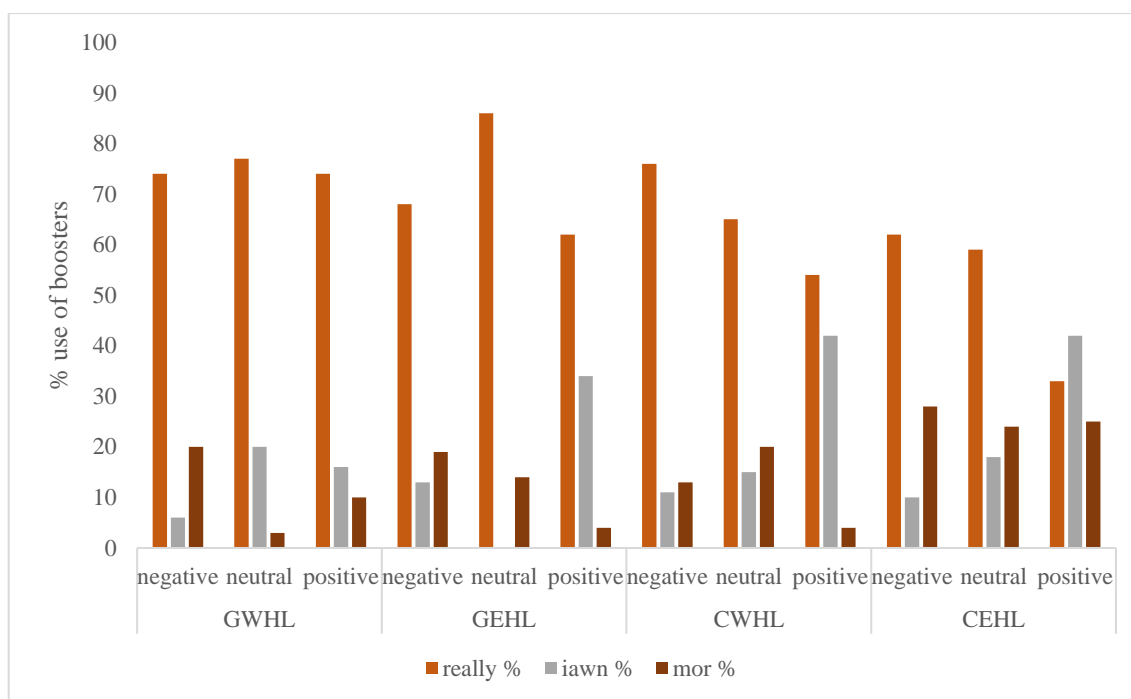


Figure 7.7: Proportionate use of boosters with negative or positive adjectives

As mentioned in 7.3.3, where boosters collocate widely with both positive and negative adjectives, it is assumed that they are further along the process of delexicalization. For the GWHL speaker group, *really* collocates to a similar extent across types of adjective. On the other hand, *iawn* favours collocation with positive and neutral adjectives more than negative. *Mor* is most frequent in negative constructions. For this group, *really* appears to be further along the process of grammaticalization than *iawn* and *mor*.

For GEHL, the pattern with *really* is similar; it occurs evenly with different types of adjective (favouring neutral adjectives somewhat). Similarly to their GWHL counterparts, this group also favoured the use of *iawn* with positive adjectives. *Mor* on the other hand collocated most often with negative and neutral adjectives.

For CWHL, the spread across adjective quality is similar. *Really* occurs most frequently with negative and neutral adjectives. With positive adjectives, however, there is a greater spread of use between *really* and *iawn*. For the CEHL group, the bars cluster more closely together, showing a more even spread between different types of adjectives. *Really* occurred

more in negative adjective cases, favoured over the positive adjective collocation. Again, *iawn* was favoured with negative adjectives, and in this group, to a greater extent than the Gwynedd groups presented above. *Mor* on the other hand, was relatively evenly spread across the types of adjectives. For this group, *really* (which is favoured in more negative collocations than positive) appears less stable than for the Gwynedd speakers. The rates of use of *iawn* also occur most frequently with positive than negative adjectives, showing that this intensifier could also be in a state of flux for this group. *Mor*, on the other hand is widely collocated with both positive and negative adjectives.

The patterns which emerge from this data show that in Gwynedd *really* appears stable and seems to be more grammaticalized than in Cardiff. For all groups (but particularly in Cardiff), *iawn* is favoured in more positive collocations. As its collocation is less varied, it may be an outgoing, or receding variant (Stratton and Sundquist 2022). Two groups use *mor* to a similar extent with different types of adjectives; for CEHL and GWHL speakers *mor* appears to be further grammaticalized. Overall, general patterns are replicated across home language backgrounds in both areas, showing that patterns of WHL use have been acquired by EHL speakers.

7.5.3.4 Adjective function

Next, I turn to review the function of the adjective, in relation to the use of boosters, presented in Table 7.18 (and Figure 7.8) below.

Table 7.18: Proportionate use of boosters with attributive or predicative adjectives

	really %	iawn %	mor %	Total (N)
GWHL	75	14	11	213
attributive	61	39	0	18
predicative	76	12	12	195
GEHL	66	22	12	104

attributive	45	55	0	11
predicative	69	18	13	93
CWHL	63	29	8	226
attributive	55	45	0	38
predicative	64	26	10	188
CEHL	47	28	26	225
attributive	59	38	3	29
predicative	45	26	29	196

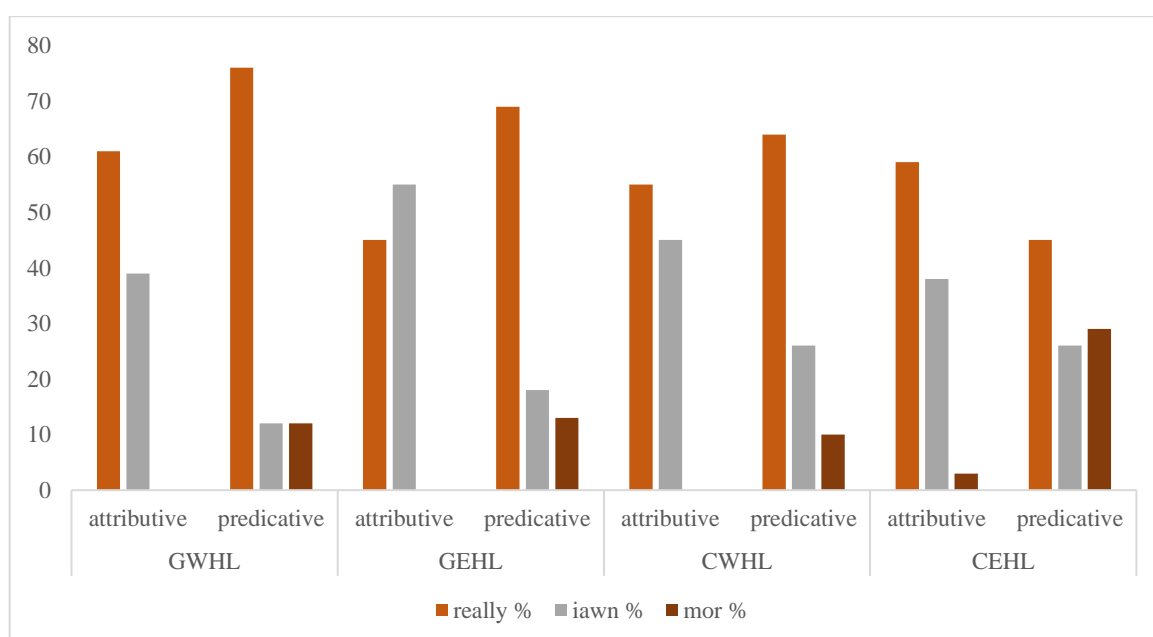


Figure 7.8: Proportionate use of boosters with attributive or predicative adjectives

The most frequent booster, *really*, occurs with mostly predicative adjectives in all speaker groups, apart from the CEHL group. For CEHL, *really* is used in more attributive constructions, implying that for this group, (paired with the fact that it appeared in more negative than positive constructions in the previous section) this variant may be less grammaticalized for this group. *Mor* is most frequently used in predicative constructions for CEHL, showing that for this group, *mor* may be the most grammaticalized booster. All speaker groups use *iawn* more often in attributive constructions than predicative, and this can be seen to a greater extent in Gwynedd than in Cardiff. Whereas acquisition of constraints of booster use have been acquired in

Gwynedd, with GEHL following GWHL patterns, for CEHL, the use of boosters differs from that of CWHL.

7.5.3.5 *Summary of booster descriptives*

Overall, the rates of *really* and *mor* increase as the speech context becomes more informal, and rates of *iawn* decrease accordingly. GWHL over-rely on *really* even in formal speech context, which suggests that for this group, the English borrowing could be less stylistically marked and more embedded than other groups. This pattern is not replicated by the GEHL group who use less of the informal variant in formal contexts. *Really* also occurred more with English than Welsh adjectives, whereas *iawn* was seen to be used more with Welsh than English adjectives, which is consistent with what I expected to find. In Gwynedd *mor* was used with more English than Welsh adjectives, whereas in Cardiff there was a more even distribution between the two. In both cases, the pattern was acquired by EHL speakers in both areas. In Gwynedd, *really* was used frequently with both negative and positive adjectives, whereas in Cardiff, this booster collocated far more with more negative adjectives. In Cardiff, it appears as though *really* is less grammaticalized as an intensifier, further supported by the fact that for the CEHL group, it is used far more in attributive constructions than predicative ones. Some evidence was found that the booster *iawn* could be receding as it was used in far more attributive than predicative constructions across all speaker groups.

7.5.4 *Maximisers*

In this section I explore the use of maximisers by different speaker groups in the data. As mentioned in section 7.5.1, the rates of maximisers in the data were low, and only two frequent variants occurred in the data; *hollol* and *rhy* (both of which are recognised Welsh intensifiers).

7.5.4.1 *Speech context*

The rates use of maximisers in different speech contexts can be seen in Table 7.19 (and Figure 7.9). The rates here are very low, when compared with the boosters presented above.

Table 7.19: Proportionate use of maximisers in different contexts

	rhy %	holloI %	Total (N)
GWHL	82	18	17
job interview	100	0	5
peer to peer	80	20	5
sociolinguistic interview	71	29	7
GEHL	50	50	4
job interview	100	0	1
peer to peer	0	100	1
sociolinguistic interview	50	50	2
CWHL	48	52	33
job interview	83	17	6
peer to peer	50	50	18
sociolinguistic interview	22	78	9
CEHL	68	32	22
job interview	38	63	8
peer to peer	100	0	6
sociolinguistic interview	75	25	8

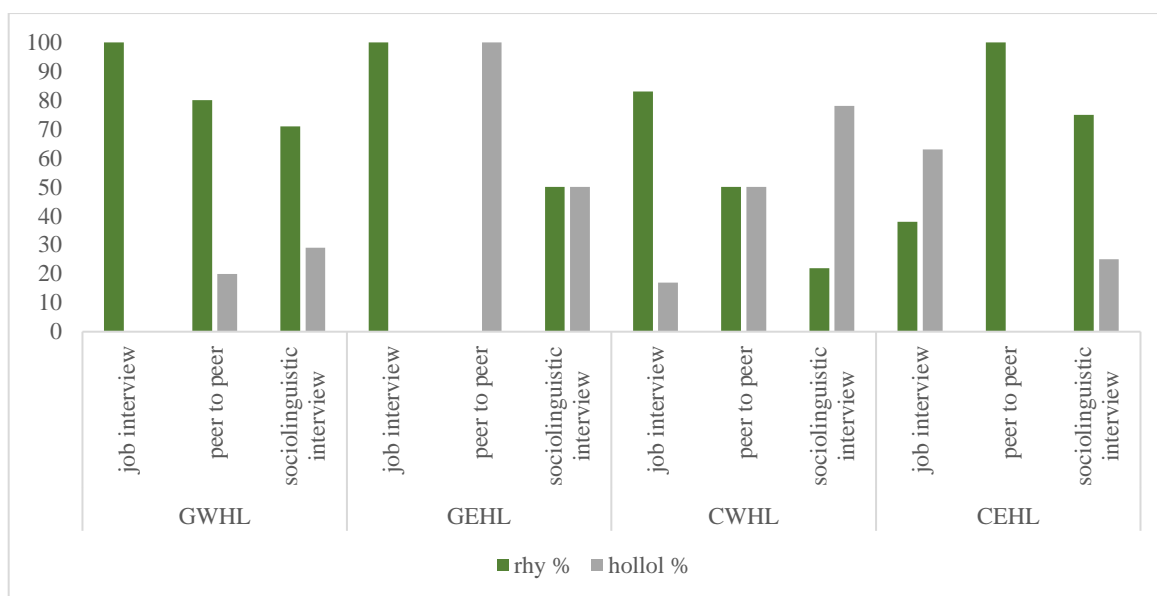


Figure 7.9: Proportionate use of maximisers in different contexts

For GWHL, the rate of use of *rhy* decreases as the speech context becomes more informal, whereas *hollol* grows in frequency. This is replicated to an extent in GEHL speakers, however rates are very low (in some cases only one token), and therefore no clear patterns can be observed.

CWHL show similar trends to GWHL; the job interview contains mostly *rhy* variants, whereas the sociolinguistic interview contains most *hollol* variants. CEHL appear to use a different pattern, with higher rates of *hollol* in the job interview and higher rates of *rhy* in the sociolinguistic interview. Again, however, low rates could account for this difference. Because rates are so low for maximisers, they will not be analysed further against other linguistic factors.

7.5.5 Summary of descriptive findings

This section presented the initial descriptive findings of the analysis of intensifiers by participating speakers. In order to address the research questions, each speaker group will be presented separately, highlighting the stylistic repertoire (with linguistic constraints) for the

WHL groups in both areas, and showing to what extent their EHL counterparts have acquired these patterns:

GWHL

- *eithaf* is more formal and used more with Welsh adjectives.
- Moderator repertoire contains wide range of other informal variants (but *cweit* may be receding as this occurs in less predicative constructions than attributive)
- *really* is used across the board in both formal and informal speech context and occurs mostly with English adjectives.
- *iawn* is more prevalent in more formal settings, and occurs mostly with Welsh adjectives. However, it occurs in more attributive and positive constructions (and is thus less grammaticalized, providing evidence that it could be a receding variant).

GEHL

- Evidence of acquisition of the above, however:
 - More conservative use of *really* in formal contexts

CWHL

- *eithaf* is used in more formal contexts and used more with Welsh adjectives.
- Few alternative moderator variants exist for this group.
- *bach* is also used more widely and in more informal contexts than in Gwynedd. There was insufficient evidence about grammaticalization of this feature, as it may only occur in predicative constructions, as with the English 'a bit')
- *really* is mostly informal and occurs mostly with English adjectives. In Cardiff, this appears to be less grammaticalized than in Gwynedd as it is widely collocated with negative and positive adjectives, and is present in more attributive constructions, too.

- *iawn* is more formal, occurring mostly with Welsh adjectives. However, it is used in attributive and positive constructions, so, as with Gwynedd, this could be a receding boosting variant.
- *mor* is more frequent here than in Gwynedd, and seems grammaticalized and stable, and used with both English and Welsh adjectives.

CEHL

- Evidence of acquisition of the above, however:
 - *really* even less grammaticalized than CWHL (it is more frequently used in more attributive constructions, with negative adjectives)
 - Different patterns of use for maximisers (but low tokens could account for this)

The findings presented here suggest that context, adjective language, quality of the adjective and position of the adjective could be predictive factors influencing the rates of intensification in Welsh for this group of speakers. This summary shows that overall EHL speakers have acquired the patterns of their WHL counterparts. However, the use of *really* by EHL speakers differs to a small extent, with evidence of it being less liberally used in formal contexts in Gwynedd, and being less grammaticalized in Cardiff.

Because the different categories (moderators, boosters, maximisers) of intensifiers in the data serve different functions, the inferential analysis below will not compare rates of intensification across function (for example, comparing the use of *really* vs *eithaf*). It would therefore be necessary to present an inferential analysis of intensifiers within each function. As has been discussed in the distributional analysis so far, the overall tokens of maximisers are very low ($n = 3.07$ per 10,000 words), and thus would not lend itself to inferential analysis. Moderators, on the other hand, occur more frequently. However, some tokens (e.g. *cweit* and *reit*) are only used by some groups. In addition, in keeping with the previous analyses, there would need to be a clear theoretical basis for describing one of the variants as more colloquial

than others, in order to assess whether its use is affected by the formality of the social setting. With the moderator variants, this is not clear. Boosters, on the other hand, contain the highest number of tokens, and the most common variant, *really*, is a codeswitch from English and could arguably be described as the most colloquial variant. *Really* has also been shown to be used differently by WHL and EHL speakers, and an inferential analysis may shed further light on different patterns of its use. The next section presents the results of the mixed effects regression analyses, each of which examined the linguistic and social factors that predicted the use of *really*.

7.6 Results (inferential analysis)

As with the previous analysis chapters looking at possessive pronouns and past tense verbs, mixed effects logistic regression analyses were carried out in order to determine the effect of external and internal factors on the variation observed in the use of the *really* variant. Firstly, I present a maximal model for the entire dataset, then I show the results of a Wald χ^2 test to determine the size of the effect (line of evidence 2), and create a parsimonious model based on only the significant factors (line of evidence 1). Finally, constraint hierarchies (line of evidence 3) are presented in the form of log likelihood figures. This is repeated for each speaker group in order to allow for a comparison between GWHL, GEHL, CWHL and CEHL, to determine whether EHL speakers in two distinct localities have acquired the patterns of their WHL counterparts.

7.6.1 Summary of factors

Table 7.20 presents each of the factors thought to influence the use of intensifiers serving a boosting function in the data. The levels presented here were built into the statistical model. The invariable and variable categories of positive adjectives and negative adjectives were combined for the statistical analysis and adjective quality was categorised as neutral, negative or positive due to low token counts for the factor levels. The baseline of each factor will be discussed in following sections.

Table 7.20: Factors included in the statistical modelling

Factor	Levels
Context	Mock job interview
	Peer to peer
	Sociolinguistic interview

Speaker's home language	English Welsh
Area	Cardiff Gwynedd
Adjective language	Welsh English
Adjective quality	neutral negative positive
Adjective function	predicative attributive

7.6.2 Overall model

In this section I examine the statistical significance and relative weight of the linguistic (adjective quality and language) and social factors (speech context, home language and school), on the likelihood of producing the casual *really* variant. A logistic regression was run, controlling for individual speaker and modified adjective as a random factor. Interactions between the social factors were also included, in order to ascertain whether there is a difference in the effect of the context within the home language and area groups. The maximal model is shown in the below R code:

```
VARIANT ~ CONTEXT * HOME LANGUAGE * REGION + ADJECTIVE LANGUAGE +
ADJECTIVE QUALITY + ADJECTIVE FUNCTION + (1|PARTICIPANT) + (1|MODIFIED
ADJECTIVE)
```

The model with interactions was found to have a significantly lower deviance ($p < 0.001$) than the model without interactions, and therefore the interactions were kept.

The output for the model is reported in Table 7.21. The results tables presented show the fixed factors which were significant predictors of *really* variant use; negative significant coefficients indicate that the named factor level was less likely to result in the production of the *really* variant than the baseline factor level.

Table 7.21: Final generalised mixed effects model for boosting intensifier variants

AIC = 893.6						Observations	
Fixed effects:	Estimate	Std. Error	z-value	p-value	Sig ⁴⁸	n	% really
(Intercept)	2.42	0.78	3.08	0.002	**		
CONTEXT (vs <i>Sociolinguistic interview</i>)						371	72
peer to peer	-1.24	0.44	-2.7	0.005	**	280	53
job interview	-2.54	0.53	-4.74	<0.001	***	134	44
HOME LANGUAGE (vs <i>WHL</i>)						447	67
EHL	-1.69	0.73	-2.29	0.02	*	338	51
REGION (vs <i>Cardiff</i>)						466	53
Gwynedd	-0.29	0.73	-0.41	0.68		319	71
ADJECTIVE LANGUAGE (vs <i>English</i>)						134	72
Welsh	-0.44	0.36	-1.21	0.22		651	58
ADJECTIVE QUALITY (vs <i>Neutral</i>)							
Negative	-0.06	0.46	-0.14	0.88			
Positive	-1.24	0.48	-2.58	0.009	**		
ADJECTIVE POSITION (vs <i>Attributive</i>)						115	47
Predicative	-0.04	0.28	-0.16	0.86		670	63
CONTEXT*HOME LANGUAGE (vs <i>sociolinguistic interview*WHL</i>)							
Peer to peer*EHL	1.05	0.55	1.87	0.06	.		
Jon interview*EHL	1.39	0.81	1.72	0.09	.		
CONTEXT*REGION (vs <i>sociolinguistic interview*Cardiff</i>)							
Peer to peer*Gwynedd	1.08	0.67	1.61	0.10			
Job interview*Gwynedd	1.78	0.74	2.38	0.01	*		
HOME LANGUAGE*REGION (vs <i>WHL*Cardiff</i>)							
EHL*Gwynedd	0.90	1.05	0.86	0.39			
CONTEXT*HOME LANGUAGE*REGION (vs <i>sociolinguistic interview*EHL*Gwynedd</i>)							
Peer to peer*WHL*Cardiff	-1.35	0.99	-1.36	0.17	.		

⁴⁸ No significance where $p > 0.05$; * where $p \leq 0.05$; ** where $p \leq 0.01$; *** where $p \leq 0.001$

Job interview*WHL*Cardiff	-1.27	1.21	-1.04	0.29		
Random effects:					<i>sd</i>	<i>n</i>
SPEAKER					0.28	17
MODIFIED ADJECTIVE					1.15	216

The context factor was found to be a significant predictor of the booster *really*, which was less likely to appear in the peer-to-peer and the job interview contexts, compared with the sociolinguistic interview context. English home language speakers were also significantly less likely to use the *really* variant than their WHL counterparts. The quality of adjectives was a significant predictor of variant use too; speakers were significantly less likely to produce *really* with positive adjectives than with negative or neutral adjectives. The interaction between context and region was also found to be a statistically significant predictor of booster use, which suggests that there is a difference in the effect of the context between the two region groups.

As mentioned in previous analysis chapters, it is not possible to fully assess the hierarchy of constraints until a sparser and more parsimonious model is created, using only factors that contribute significantly to predicting the variation. As with previous analysis chapters, predictors which showed evidence of significant interaction were also included. A Wald χ^2 (chi square) test (which iteratively adds and removes each predictor and compares how well each iteration fits the distribution of the data (Gardner, 2023) is presented in Table 7.22.

Table 7.22: Analysis of deviance, Wald χ^2 test for maximal model

	Wald χ^2	Df	<i>p</i> -value	
Context	20.78	2	<0.001	***
Adjective Quality	14.31	2	<0.001	***
Home language	3.27	1	0.07	
Adjective language	1.49	1	0.22	
Region	1.39	1	0.23	
Adjective position	0.009	1	0.92	

Context has the largest χ^2 value (20.78) indicating it has the greatest effect on the variation. This is followed by adjective quality (14.31), which was also found to be a significant predictor.

Neither home language, adjective language, region nor adjective position were found to be significant predictors of the variation in possessive pronoun constructions. I next present a more parsimonious model, containing only significant predictors of variation in Table 7.23.

Table 7.23: Parsimonious model with only significant predictors of variation

AIC = 886.5						Observations	
	Estimate	Std. Error	z-value	p-value	Sig	<i>n</i>	% really
Fixed effects:							
(Intercept)	1.28	0.49	2.60	0.009	**		
CONTEXT (<i>vs Sociolinguistic interview</i>)						371	72
peer to peer	-0.46	0.23	-1.98	0.04	*	280	53
job interview	-1.40	0.29	-4.70	<0.001	***	134	44
ADJECTIVE QUALITY (<i>vs Neutral</i>)							
Negative	-0.01	0.41	-0.03	0.97			
Positive	-1.26	0.47	-2.68	0.007	**		
Random effects:						<i>sd</i>	<i>n</i>
SPEAKER						1.09	17
MODIFIED ADJECTIVE						1.07	216

The results from the most parsimonious model confirm previous findings that the job interview is significantly less likely to contain the *really* variant compared with the sociolinguistic interview. This was also the case for the peer-to-peer context. A matrix of all the comparisons for the context factor (beyond comparing each level to the reference level) was run using the `glht` function in R (Hothorn, Bretz, and Westfall 2008), and showed that the *really* variant was also significantly less likely to appear in the job interview, compared with the peer-to-peer context ($\beta = -0.94$, $z = -3.20$, $p = 0.004$). There was no significant difference between the use of *really* in the peer-to-peer and sociolinguistic interview contexts.

Positive adjectives were also far less likely to be used with the *really* variant than neutral adjectives. A comparison matrix also revealed that *really* was significantly less likely to be used with positive adjective than negative adjectives ($\beta = -1.25$, $z = -3.73$, $p < 0.001$).

The results from the parsimonious mixed effects model are presented in the log-odds in Figure 7.10. This figure plots the p value results as well as their confidence intervals in order

to display with what certainty any significant factors (in this case only adjective quality) can predict the use of *really*.

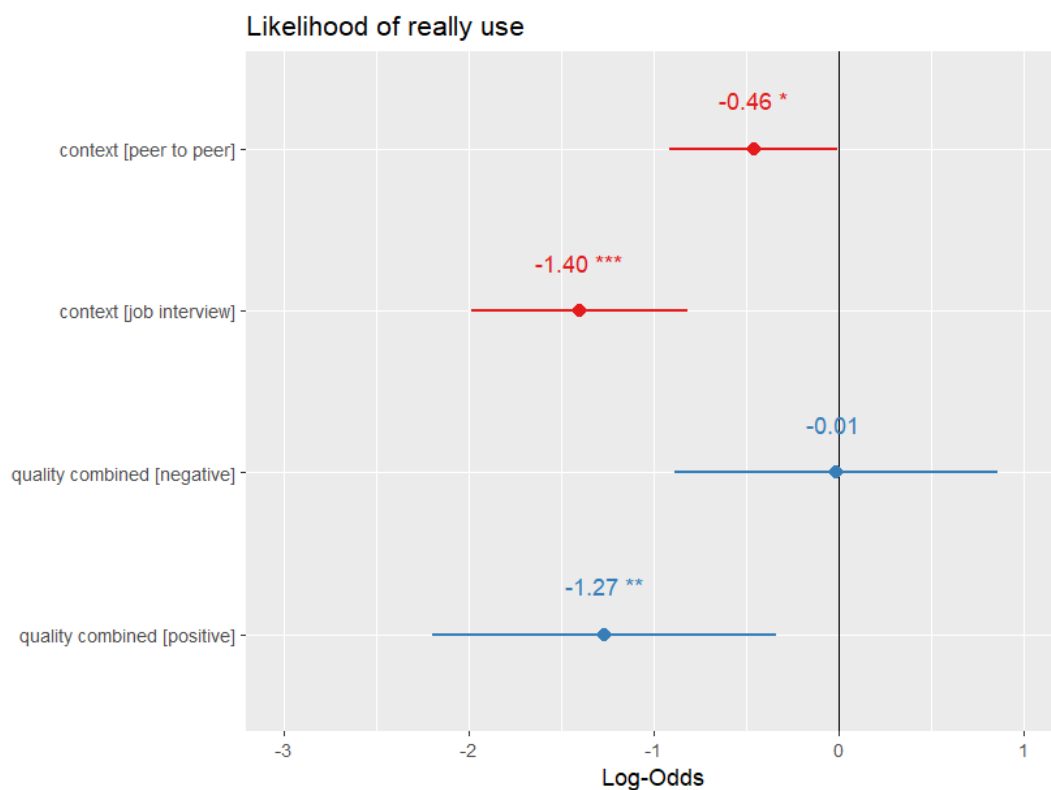


Figure 7.10: Log-odds results for *really* use

7.6.3 Summary

So far, we have seen that context is the strongest predictor of intensifier use; the *really* variant appearing significantly more often in informal speech contexts. Adjective quality was the other significant predictor of *really* use, with negative adjectives being far more likely to be used with the variant (e.g. *really drwg* [really bad]) than positive adjectives (e.g. *really da* [really good]).

Surprisingly, English adjectives are not used with the English calque *really* more often than Welsh adjectives, which is what was hypothesised earlier in this chapter. Though it first appeared that home language groups did differ significantly in their use of *really*, this was not a strong enough predictor to explain the variation. The parsimonious model has shown that

area and home language are not significant predictors of *really* variation. I need to examine individual speaker groups in order to assess the ranking of constraints between the groups.

7.6.4 Gwynedd Welsh home language subset

The results of the regression analysis for this group are presented in Table 7.24.

Table 7.24: Really regression analysis for GWHL

AIC = 196.8						Observations	
	Estimate	Std. Error	z-value	p-value	Sig	n	% really
Fixed effects:							
(Intercept)	4.03	2.96	1.36	0.17			
CONTEXT (<i>vs Sociolinguistic interview</i>)						113	79
peer to peer	0.02	1.09	0.02	0.98		55	67
job interview	-1.74	1.02	-1.70	0.08		46	69
ADJECTIVE LANGUAGE (<i>vs English</i>)						47	76
Welsh	-1.48	1.98	-0.75	0.45		167	74
ADJECTIVE QUALITY (<i>vs Neutral</i>)						23	77
Negative	-2.25	2.08	-1.07	0.28		99	74
Positive	-3.46	2.05	-1.68	0.09		37	74
ADJECTIVE POSITION (<i>vs Attributive</i>)						19	58
Predicative	0.72	1.08	0.67	0.50		195	76
Random effects:						sd	n
SPEAKER						2.57	5
MODIFIED ADJECTIVE						3.94	82

There are no significant results in this model; none of the levels within each factor were significant predictors of the use of possessive pronouns in the current dataset. In order to examine the effect size of the factors above (to enable a comparison with other groups), I present the χ^2 statistics, degrees of freedom and p values from a Wald χ^2 test carried out on this subset model. The Wald χ^2 test for GWHL revealed that context was the strongest predictor (though not to a significant extent) of booster variant use ($\chi^2= 3.01$, $df= 2$, p -value= 0.22), followed by adjective quality and language. Adjective position was the weakest predictor. The results are presented in Table 5.27.

Table 7.25: Wald χ^2 test for GWHL colloquial variant use

	Wald χ^2	Df	<i>p</i> -value
Context	3.01	2	0.22
Adjective Quality	2.90	2	0.23
Adjective language	0.55	1	0.45
Adjective Position	0.45	1	0.50

7.6.5 Gwynedd English home language subset

For this speaker group, no factor was found to be a significant predictor of *really* use (see Table 7.26). In this way, they are similar to their WHL counterparts. Hierarchy patterns within individual factor groups will be discussed alongside all other speaker groups in section (7.6.8).

Table 7.26: *Really* regression analysis for GEHL

AIC = 131.8						Observations	
						<i>n</i>	% really
Fixed effects:	Estimate	Std. Error	z-value	<i>p</i> -value	Sig		
(Intercept)	1.81	2.47	0.74	0.46			
CONTEXT (<i>vs Sociolinguistic interview</i>)						65	74
peer to peer	-0.92	0.86	-1.07	0.28		22	59
job interview	-1.54	1.10	-1.39	0.16		18	44
ADJECTIVE LANGUAGE (<i>vs English</i>)						26	69
Welsh	-0.19	0.90	-0.21	0.83		79	65
ADJECTIVE QUALITY (<i>vs Neutral</i>)						6	68
Negative	-2.32	1.95	-1.19	0.23		32	68
Positive	-2.67	2.12	-1.26	0.20		31	62
ADJECTIVE POSITION (<i>vs Attributive</i>)						12	41
Predicative	1.29	1.00	1.28	0.19		93	69
Random effects:						<i>sd</i>	<i>n</i>
SPEAKER						1.83	4
MODIFIED ADJECTIVE						1.44	57

The Wald χ^2 test for GEHL revealed some similarities between the effect sizes of the factors constraining this group's use of *really* and their WHL counterparts. Firstly, the magnitude of effect for the factors is small; even smaller than was the case for GWHL. Context was also the strongest predictor for this group (though not to a significant extent, $\chi^2 = 2.27$, *df* = 2, *p*-value =

0.32). The position of the adjective had the second strongest effect, however, followed by adjective quality and language. The results are presented in Table 7.27.

Table 7.27: Wald χ^2 test for GEHL colloquial variant use

	Wald χ^2	Df	<i>p</i> -value
Context	2.27	2	0.32
Adjective Position	1.65	1	0.50
Adjective Quality	1.60	2	0.44
Adjective language	0.04	1	0.19

7.6.6 Cardiff Welsh home language subset

Table 7.28 shows that for this group, two factor groups are significant predictors of *really* use. CWHL speakers are highly significantly more likely to use *really* in the sociolinguistic interview context than in the job interview context. The peer-to-peer context is also a significant predictor of *really* use. For this speaker group, Welsh adjectives are also significantly less likely to be used with *really* e.g. *da iawn* [very good] rather than *really da* [really good].

Table 7.28: *Really* regression analysis for CWHL

AIC = 256.6						Observations	
	Estimate	Std. Error	z-value	<i>p</i> -value	Sig	<i>n</i>	% really
Fixed effects:							
(Intercept)	3.30	1.25	2.63	0.008	**		
CONTEXT (<i>vs Sociolinguistic interview</i>)						87	85
peer to peer	-1.52	0.522	-2.92	0.004	**	100	54
job interview	-2.73	0.65	-4.17	<0.001	***	46	30
ADJECTIVE LANGUAGE (<i>vs English</i>)						30	93
Welsh	-2.21	0.95	-2.33	0.01	*	203	56
ADJECTIVE QUALITY (<i>vs Neutral</i>)						13	65
Negative	0.86	0.81	1.06	0.28		60	76
Positive	3.35	0.77	0.46	0.64		69	54
ADJECTIVE POSITION (<i>vs Attributive</i>)						47	45
Predicative	-0.09	0.47	-0.19	0.84		186	65
Random effects:						<i>sd</i>	<i>n</i>
SPEAKER						0.31	4
MODIFIED ADJECTIVE						1.16	97

Similarly to what we have seen in previous chapters, the factor effect size is stronger in the Cardiff data than in Gwynedd. Context in Gwynedd was consistently the strongest predictor on *really* use; the Wald χ^2 test for CWHL reveals the same pattern for this group, too ($\chi^2 = 17.55$, $df = 2$, p -value < 0.001). Where this group differs, however, is that the second strongest (and significant) effect is the language of the adjective. Adjective quality and position, on the other hand are not significant. The results are presented in Table 7.29.

Table 7.29: Wald χ^2 test for CWHL colloquial variant use

	Wald χ^2	Df	p -value
Context	17.55	2	< 0.001 ***
Adjective language	5.45	1	0.01 *
Adjective Quality	1.39	2	0.49
Adjective position	0.03	1	0.84

A parsimonious model, including only significant predictors of variation (outlined in the Wald test above), is presented in Table 7.30. This model confirms that the job interview is the least likely to contain the informal *really* variant, followed by the peer-to-peer context, followed by the sociolinguistic interview. A contrast matrix using the `glht` function in R (Hothorn, Bretz, and Westfall 2008), showed that there were significant differences in the use of the variants across all three contexts. As well as differing significantly from the reference level, the peer-to-peer context contained significantly more instances of *really* than the job interview ($\beta = -1.25$, $z = -2.41$, $p = 0.04$).

Table 7.30: Parsimonious model for CWHL subset

AIC = 252.0						Observations	
						n	% really
Fixed effects:	Estimate	Std. Error	z -value	p -value	Sig		
(Intercept)	3.98	0.98	4.04	< 0.001	***		
CONTEXT (<i>vs Sociolinguistic interview</i>)						87	85
peer to peer	-1.52	0.51	-2.97	0.002	**	100	54
job interview	-2.77	0.65	-4.30	< 0.001	***	46	30
ADJECTIVE LANGUAGE (<i>vs English</i>)						30	93
Welsh	-2.48	0.94	-2.643	0.008	**	203	56
Random effects:						sd	n
SPEAKER						0.28	4

7.6.7 Cardiff English home language subset

Table 7.31 shows that for this group, only context is a significant predictor of *really* use. CEHL speakers were significantly more likely to use *really* in the sociolinguistic interview context than in the job interview context. Differently to their WHL counterparts, CEHL are not significantly constrained by adjective language. Neither adjective quality nor position were found to be significant either.

Table 7.31: Really regression analysis for CEHL

Fixed effects:	Estimate	Std. Error	z-value	p-value	Sig	Observations	
						<i>n</i>	% really
(Intercept)	0.71	0.74	0.91	0.33			
CONTEXT (<i>vs Sociolinguistic interview</i>)						106	52
peer to peer	-0.13	0.30	-0.42	0.67		103	44
job interview	-1.37	0.59	-2.13	0.03	*	24	21
ADJECTIVE LANGUAGE (<i>vs English</i>)						31	48
Welsh	0.39	0.43	0.92	0.35		202	45
ADJECTIVE QUALITY (<i>vs Neutral</i>)						10	59
Negative	0.20	0.56	0.36	0.71		56	62
Positive	-0.91	0.56	-1.63	0.10		39	33
ADJECTIVE POSITION (<i>vs Attributive</i>)						37	46
Predicative	-0.72	0.46	-1.57	0.11		196	45
Random effects:						<i>sd</i>	<i>n</i>
SPEAKER						0.23	4
MODIFIED ADJECTIVE						0.00	92

The Wald χ^2 test for CEHL reveals a different pattern from the three previous subsets. For this group, the quality of the adjective is the factor with the strongest effect ($\chi^2 = 13.21$, $df = 2$, $p = 0.001$). Note that the adjective quality was not significant in the first subset model, and yet appears to have a relatively large effect in the Wald χ^2 test. It is possible for a study to have a large effect size but not achieve statistical significance, which could relate to limited statistical power (due to sample sizes being too small to detect the effect with enough precision) (see

Gardner 2023). Although context is the second strongest effect, it is not significant, and neither is home language, which shows that this group have not acquired all the constraints of CWHL. The results of the test are presented in Table 7.32.

Table 7.32: Wald χ^2 test for CEHL colloquial variant use

	Wald χ^2	Df	<i>p</i> -value
Adjective Quality	13.21	2	0.001 **
Context	4.56	2	0.10
Adjective Position	2.51	1	0.11
Adjective Language	0.85	1	0.35

A parsimonious model for CEHL is presented in Table 7.33, and confirms that positive adjectives are significantly less likely to be used with the *really* variant than neutral adjectives. A comparison matrix showed a significant difference between positive and negative adjectives, too ($\beta = -1.21$, $z = -4.11$, $p = 0.001$), with negative adjectives significantly more likely to be used with the *really* variant.

Table 7.33: Parsimonious model for CEHL subset

AIC = 302.0						Observations	
	Estimate	Std. Error	z-value	<i>p</i> -value	Sig	<i>n</i>	% really
Fixed effects:							
(Intercept)	0.37	0.50	0.75	0.45			
ADJECTIVE QUALITY (<i>vs Neutral</i>)						10	59
Negative	0.12	0.54	0.23	0.81		56	62
Positive	-1.08	0.54	-2.02	0.04	*	39	33
Random effects:						<i>sd</i>	<i>n</i>
SPEAKER						0.12	4
MODIFIED ADJECTIVE						0.00	92

7.6.8 Hierarchies for the subset groups

Here I present the initial patterns of the factor level hierarchy for each of the groups. The data in Figure 7.11 are taken from the full subset models for each speaker group. As with the previous analysis chapters, the full subset models are used rather than the parsimonious ones, to allow a comparison of all factors, to determine patterns across speaker groups.

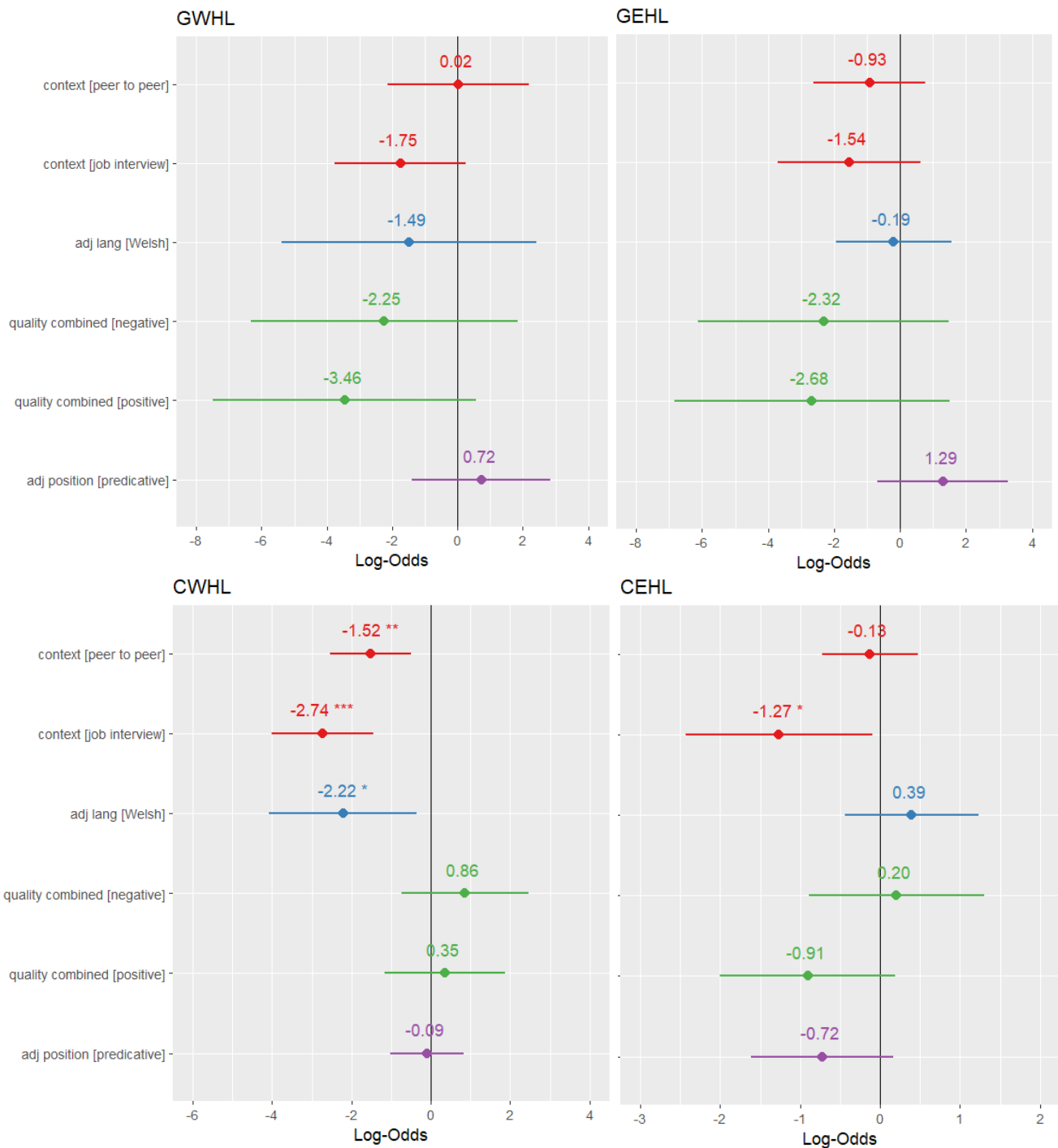


Figure 7.11: Log-odds likelihood of colloquial pronoun use across four speaker groups

The most noteworthy pattern to emerge in this figure is that GEHL patterns are almost identical to GWHL patterns. Recall that the strength of effect for both groups was very weak across all predictors, which is likely to be due to very high rates of the *really* variant across the board in that region. This might contribute to the size of the confidence intervals (error bars) for both groups, which span the intercept. Although neither group were significantly constrained by any social or linguistic factors, the EHL group in Gwynedd have quite clearly acquired the patterns of the WHL group; the job interview context elicited lower rates of *really*, Welsh adjectives, positive and negative adjectives were used with *really* less than neutral adjectives, and predicative constructions elicited more cases of the *really* variant.

The patterns for the Cardiff groups, however, share fewer similarities. Although the hierarchy of context is replicated, Welsh adjectives are significantly less likely to elicit *really* for CWHL, whereas for CEHL the pattern is inverse. CEHL use more of the *really* variant with Welsh adjectives. Cardiff groups are similar in that negative adjectives prompt the use of *really*, however it is a mixed picture for positive adjectives. Indeed, CWHL is the only group where positive adjectives elicit more cases of the *really* variant than neutral adjectives.

7.7 Discussion

Next, I summarise the three lines of evidence for the boosting intensifier variant, to observe where WHL patterns have been acquired (or not) by EHL participants in two regions in Wales.

7.7.1 Line of evidence 1: Statistical significance

As discussed in the analysis of possessive pronouns and Simple Past verbs, the first line of evidence shows us which independent variables are significant predictors of the variation, and which are not. The results from the most parsimonious subset models (where available) are presented in Table 7.34. In Gwynedd, there were no significant predictors of variation (the similarity of patterns discussed above will be covered in line of evidence 3). For CWHL,

context was a significant predictor; the peer to peer and sociolinguistic interview contexts were significantly more likely to elicit the use of the *really* variant. CEHL, on the other hand, were only significantly constrained by the quality of the adjective, i.e. where positive adjectives were significantly less likely to be used with the *really* variant than negative and neutral adjectives. On the whole, significance patterns observed in WHL are not replicated in Cardiff, whereas in Gwynedd, where there are no strong patterns, this is the same for both home language groups. Further to this. Gwynedd and Cardiff are not matched in terms of WHL patterns; north and south Wales have different respective patterns, seemingly, in terms of their use of *really*.

Table 7.34: Line of evidence 1 (intensifiers)

	Gwynedd		Cardiff	
	WHL	EHL	WHL	EHL
Context	No	No	Yes	No
Adjective language	No	No	Yes	No
Adjective quality	No	No	No	Yes
Adjective position	No	No	No	No

7.7.2 Line of evidence 2: Strength of effect

For the second line of evidence, I present the findings from the Wald χ^2 tests conducted for each subset model. The descending WHL pattern for each region has been highlighted in the table to show whether the EHL speakers have acquired the same patterns. The stronger effect (for WHL) is darker, and the weaker effect (for WHL) is a lighter shade.

Table 7.35: Line of evidence 2 (intensifiers)

	Gwynedd				Cardiff			
Strength of effect	WHL		EHL		WHL		EHL	
↓	Context	3.01	Context	2.27	Context	17.55	Adjective Quality	13.21
	Adjective Quality	2.90	Adjective Position	1.65	Adjective language	5.45	Context	4.56
	Adjective language	0.55	Adjective Quality	1.60	Adjective Quality	1.39	Adjective Position	2.51
	Adjective Position	0.45	Adjective language	0.04	Adjective position	0.03	Adjective Language	0.85

For most groups, the context factor has the strongest effect. The strength of effect for both WHL patterns quite similarly, however in Cardiff the effect is far stronger ($\chi^2 = 17.55$) than in Gwynedd ($\chi^2 = 3.01$). This is likely because of high rates of the *really* variant overall in the Gwynedd data. Gwynedd EHL have acquired a similar effect pattern to Gwynedd WHL, where context ($\chi^2 = 2.27$) is the strongest effect. However, adjective position has more explanatory power in GEHL ($\chi^2 = 1.65$) than in GWHL ($\chi^2 = 0.45$).

In Cardiff, the strongest effect for the EHL group is adjective quality ($\chi^2 = 13.21$), which does not match the pattern of CWHL. Indeed, the strength of effect patterns are very different between both Cardiff groups, showing that CEHL have not acquired the same patterns as CWHL. This provides further support to the first line of evidence which shows that EHL speakers not acquiring WHL speaker patterns in Cardiff, and in Gwynedd, where fewer patterns are visible for WHL, this is seemingly replicated.

7.7.3 Line of evidence 3: Constraint hierarchy

For the third line of evidence, I present the hierarchy of the levels within each factor, in order to compare the patterns between the groups. The darker shaded cells indicate where speakers have not acquired the same constraint hierarchy as their WHL counterparts.

Table 7.36: Line of evidence 3 (intensifiers)

	Gwynedd		Cardiff	
	WHL	EHL	WHL	EHL
Context	Peer to peer > sociolinguistic interview > job interview	Sociolinguistic interview > peer to peer > job interview	Sociolinguistic interview > peer to peer > job interview	Sociolinguistic interview > peer to peer > job interview
Adjective Language	Welsh adjectives less <i>really</i> than English	Welsh adjectives less <i>really</i> than English	Welsh adjectives less <i>really</i> than English	Welsh adjectives more <i>really</i> than English
Adjective Quality	Positive and negative adjectives less <i>really</i>	Positive and negative adjectives less <i>really</i>	Positive and negative adjectives more <i>really</i>	Negative adjective more <i>really</i> , positive adjectives <i>less</i> <i>really</i>
Adjective Position	Predicative adjectives more <i>really</i>	Predicative adjectives more <i>really</i>	Predicative adjectives less <i>really</i>	Predicative adjectives less <i>really</i>

Here, we see further suggestions that *really* patterns differently in Gwynedd and in Cardiff, as different patterns can be observed between WHL groups in the two areas. On the whole, local

patterns have been acquired to some extent. In Gwynedd, EHL speakers context constraints are similar, in that the job interview elicits the lowest rates of *really* overall (even though the two more informal settings are inverse). All internal constraints are matched in Gwynedd. In Cardiff, the main difference is that less of the *really* variant was observed for WHL with Welsh adjectives, whereas in Gwynedd, more *really* variants were used with Welsh adjectives. The WHL pattern of adjective quality has also only been partially acquired by CEHL.

Across all three lines of evidence, EHL students pattern the WHL norms more closely in Gwynedd than in Cardiff. In Gwynedd, the social factor (context) has not been fully acquired by EHL students, whereas in Cardiff, EHL are not fully acquiring the linguistic factors of WHL speakers. On the whole, however, with regards to the third line of evidence, we see that sociolinguistic competence is being acquired in relation to the intensifier feature *really*.

7.7.4 Insights to the feature beyond sociolinguistic competence

Considering that this was the first variationist work to look at the use of intensifiers in Welsh, here I summarise other noteworthy findings from the above analysis:

- The findings discussed in this chapter show that boosters are the most frequent form of intensification in the speech of these participants. This finding is consistent with other international sociolinguistic research on naturally occurring English speech (cf. D’Arcy 2015, Reichelt and Durham 2017 for British and American English scripted television, Stratton and Sunquist 2022 for Norwegian, and Stratton 2020 for German). The three common boosters found in the dataset were *really*, *iawn* [very] and *mor* [so].
- *really* is the most frequent intensifier variant in the speech of most of these young Welsh speakers. This is surprising since descriptive grammars (e.g. Thomas 1996) do not make reference to it, and dictionaries (such as GPC) do not recognise *really* as a borrowing into Welsh (even though other intensifiers such as *reit* and *cweit* are). However, recent research has found that the incorporation of English-sourced borrowings are associated

with youth language in Scandinavian languages (see Andersen, 2014 cf. Peterson, 2017). It is possible that *really* could be a recent development into the intensification system of young Welsh speakers. Without comparing these findings to older and younger speakers, it is not known whether this is an age-graded type of variation, or whether it represents a change in progress.

- The frequency of *really* in the dataset could be explained by an ‘off the shelf’ kind of variation (Milroy 2007), where *really* is easy to pick up and use in a different language without having to re-work the syntactic structure (e.g. the majority of intensifiers in Welsh precede the adjective, just as in English).
- The ease with which the *really* borrowing is used also provides evidence for the common phenomenon in bilingual speakers to fail to separate pragmatic markers in their linguistic repertoires; as put by Peterson (2017: 118) ‘when speaking Language A, a bilingual speaker may freely incorporate pragmatic elements from Language B’. Linguistic features such as intensifiers (which are non-obligatory in morphosyntactic structures), can ‘carry signals about speaker attitudes, the speech act performed, discourse structure, information state, politeness, etc.’ (Andersen, 2014:17-18), which does not necessarily add to the propositional content of an utterance.
- Although *really* is an instance of borrowing from English, this analysis has shown that its use is not constrained by the language of the adjective (apart from for the CWHL group, who may be more careful than other groups about code mixing). For three speaker groups it is as likely to be used with an English adjective as a Welsh one.
- Regarding its grammaticalization, in Gwynedd, *really* was found to collocate mostly with predicative adjectives rather than attributive adjectives, suggesting that it has undergone a process of delexicalization in English (from where is arrived, fully formed), paired with the fact that it was as likely to be used with a positive adjective as

a negative one. Only when an intensifier collocates widely with both negative and positive adjectives is it considered to have become fully grammaticalized. The mirrored partial delexicalization for CWHL could point to the fact that this is indeed a change in progress which could stabilise over time. Evidence of this is stronger in Gwynedd than in Cardiff.

- It is worth noting that another booster, *mor* [so], is used to a greater extent with predicative adjectives, and is also widely collocated with negative and positive adjectives, which could point to this intensifier being more strongly grammaticalized (particularly for Cardiff speakers).
- On the other hand, *iawn* [very], the final booster under study, was the intensifier to occur with the most cases of attributive adjectives, which could suggest that this intensifier is also undergoing change, but in this case is receding, with young people favouring *really* and *mor*.

7.8 Conclusion

This work is the first of its kind to map the patterns of young Welsh speakers' intensifier systems. I have examined the use of different intensifiers by the four speaker groups in order to establish a) the repertoire of intensifier use by WHL participants in two areas in Wales, and b) the extent to which the repertoire (and the constraints which govern them) are replicated by their EHL counterparts. This chapter provides evidence of the intensifier *really* occurring frequently overall in young speakers' Welsh. It is stylistically and grammatically constrained by WHL participants, and EHL participants have similar patterns of use, showing that sociolinguistic competence has been acquired to some extent by speakers. In Cardiff, some differences between speaker groups are observed, particularly in that EHL have not acquired some of the linguistic constraints. In Gwynedd, although EHL seem to be following WHL more

closely, there are few discernible patterns here, and only partial acquisition of social constraints was observed. However, more similarities than differences can be observed in the data.

8 Discussion

8.1 Introduction

Broadly, the preceding analysis chapters have shown that sociolinguistic competence has been acquired but to different extents for each feature. The research questions below help us account for this:

1. What are the patterns of stylistic variation in the speech of pupils from Welsh-speaking homes?
2. Do pupils from non-Welsh-speaking homes acquire similar patterns of use across styles, and therefore acquire similar sociolinguistic competence?
3. How does the acquisition of sociolinguistic competence among L2 speakers differ depending on the extent to which Welsh is spoken in the wider community?

The main aim of this research is to examine the extent to which pupils from non-Welsh-speaking homes are acquiring sociolinguistic competence comparable to their Welsh home language peers. In order to do this, it was necessary to establish the extent to which the three features exhibit stylistic variation in the speech of those who come from Welsh-speaking homes in both areas (RQ1). I then examined the extent to which EHL students are matching their counterparts' patterns (RQ2), or whether they show distinct patterns of variation. Finally, I examined the extent to which the acquisition of sociolinguistic competence for EHL speakers is affected by the extent to which Welsh is used in the community (RQ3); i.e. do EHL groups pattern more closely to their WHL in one area over another? The following sections discuss these results more broadly and the possible implications of the findings. Specifically, I present some considerations for Welsh language planning and policy in light of *Cymraeg 2050*'s aim to double the number of speakers (Welsh Government 2017), and the broader context of language acquisition and variation in Welsh education settings and beyond.

8.2 Patterns of stylistic variation in WHL speech (RQ1)

Informal variants were most common for WHL speakers across all three features. Colloquial possessive pronouns e.g. *car fi* [my car], periphrastic verb constructions e.g. *wnes i fynd* [I went], and the borrowed intensifier *really* appear frequently in the dataset across speech contexts, even though two of these (colloquial possessive pronouns and the *really* booster) are stigmatised, and are not extensively mentioned in traditional grammars or dictionaries. On the whole, WHL participants varied their style to match the formality of the speech context, i.e. in more formal contexts there was less frequent use of the more informal/less standard variants, and in more informal contexts, the rates of such variants increased. Table 8.1 presents the results of what was found across the different contexts for each of the three features.

Table 8.1: Summary of lines of evidence for context across features (WHL)

	Gwynedd WHL		Cardiff WHL	
	Pattern (rates)	Sig?	Pattern (rates)	Sig?
Colloquial Possessives	Socio interview 74%	no	Socio interview 56%	yes
	Peer to peer 68%		Peer to peer 53%	
	Job interview 47%		Job interview 6%	
Periphrastic Verbs	Job interview 78%	no	Peer to peer 79%	yes
	Peer to peer 72%		Job interview 68%	
	Socio interview 65%		Socio interview 66%	
<i>Really</i> Intensifier	Socio interview 79%	no	Socio interview 85%	yes
	Job interview 69%		Peer to peer 54%	
	Peer to peer 67%		Job interview 30%	

When looking at the results from all features together, it becomes clear that the extent to which context is an important predictor of variation depends on the feature under study and the geographical location. Here I present an interpretation of why that might be.

Firstly, I will discuss the differences in results according to each feature. WHL speakers in both areas are adhering to overt norms when it comes to possessive pronouns and use of English intensifiers, but, to a lesser extent Simple Past verbs. The extent to which these features are strongly constrained by context varies, to a large extent, according to the perceived stigma associated with the informal variant.

When it came to the use of the colloquial possessive pronoun and the booster *really* variant, WHL students showed strong patterns relating to style; these two more informal variants mostly occurred in informal settings and occurred least of all in the more formal job interview context. These two features have clearly colloquial variants. Recall from chapter 5 that the more informal possessive pronoun variant is traditionally considered ungrammatical (Awbery 1976) non-standard (Borsley *et al.* 2007) or sub-standard (King 2016), and it remains heavily stigmatized and is often absent from grammar books. The colloquial variant is also associated with L2 varieties of Welsh (D. Willis 2016), and its use has also been classed as an oversimplification of the more standard Welsh literary and sandwich variants as a result of language shift (M. C. Jones 1998). Chapter 7 noted that *really* is associated in English with colloquial and non-standard varieties of speech (Tagliamonte 2016: 223), and further, the codeswitched intensifier was expected to be observed more frequently in more informal contexts due to the stigma associated with using English words in standard spoken Welsh (Prys 2016: 341)⁴⁹.

⁴⁹ However, the extent to which the intensifier *really* can be considered a codeswitch is refutable. Indeed, chapter 7 discussed its widespread use (with varying orthography), even though it is not recognised as a Welsh word by Geiriadur Prifysgol Cymru. This is discussed further in section 8.5.4 of this chapter.

On the other hand, there were not many clear findings from the analysis of the role of speech context on the production of formal and informal variants for the Simple Past verbs feature. In Gwynedd, the periphrastic variant is used regularly in even the more formal contexts. Cardiff speakers' usage (more periphrastic in informal contexts) is in keeping with expectations discussed in the literature; the inflected variant is considered more standard than the periphrastic variant (B. M. Jones 1993) and is thus more likely to appear in more formal contexts. When compared to the possessive pronoun and intensifier feature, there is evidence of less systematic stylistic variation with the Simple Past verb variants, particularly for EHL groups. This can be explained by the fact that the periphrastic variant, although thought to occur more in spoken varieties than written, is not stigmatised, likely because it bears no association with language contact with English or Welsh learner speech. Further work could test the theory that stigmatised features are more likely to vary between contexts of spoken Welsh.

There are clear differences between Cardiff and Gwynedd WHL speakers. Gwynedd speakers use more of the more informal variants than Cardiff speakers in all speech contexts. CWHL are also more adherent to style-shifting norms than GWHL, with a more marked distinction in the use of informal variants between different contexts of use. The differences in style-shifting could be explained by the sociolinguistic context of Welsh in both areas, which I present below.

According to these findings, Cardiff speakers are more conservative users of Welsh. Even though they, too, favour the use of non-standard and stigmatised features, they used them less often than speakers in Gwynedd, particularly in more formal contexts. Mougeon *et al.*'s (2010) study, which examined the socio-economic status of French immersion pupils, found that those pupils of a higher social class used more formal and hyperformal variants (pg. 164). The patterns of more careful style-shifting in Cardiff speakers could also be attributed to socio-economic status. In Cardiff, English-speaking parents opt-in for their children to attend a scarce

number of Welsh-medium schools, and their decision-making has been found to intersect with social class and parental cultural capital (Jones 2017: 158). This element of active parental choice in Cardiff could influence the use of Welsh among its pupils. Recall, too, the observed differences between Cardiff and Gwynedd parents in terms of their geographical mobility and jobs, remarked upon in section 4.5. For Welsh-speaking parents in Cardiff, Welsh-medium education may be the default option, but further to this, they are potentially themselves engaged in professional Welsh-speaking jobs where awareness of register is higher, since many of them have moved to the city from elsewhere in Wales. Other than general perceptions based on questionnaire data about parents' choice of Welsh-medium education for their children, I cannot verify this for certain. Future work on stylistic variation in Welsh could examine the rates of use of more informal variants in relation to socio-economic status and social and geographical mobility.

Furthermore, in both contexts, the extent to which the language is undergoing a process of revitalization differs, which contributes to the argument that varieties encountered in different areas of Wales are likely to be connected to their sociolinguistic history with the language. In Gwynedd, where the language is more prominent in the community, there appears to be a greater level of acceptance of informal and stigmatised features, with north-western speakers both using them more frequently and also not discriminating as much between different contexts of use. In Cardiff, however, prescriptivism following the perceived sense of threat to the survival of the language could be contributing to young speakers in the capital endeavouring to appear more 'well-spoken' in Welsh and turning away from features associated with language contact. Extracts from two participant transcripts shows that prescriptive attitudes among family members affect their own use of Welsh in certain settings. In the qualitative extracts below, two of the Cardiff students remarked that their respective Welsh-speaking grandmothers often correct their own use of the language.

jest *posh* Cymraeg 'sa i'n siŵr - does dim *you know* mae hi'n ddynes eithaf *snobby* so ... dw i'n trio bod yn fwy ffurfiol gyda mamgu trio defnyddio mwy o Gymraeg gyda hi lle efo Gerwyn baswn i jest yn - *you know* dweud pethau - fel siarad gyda - fel dw i yn gyda chi nawr erm a gyda mamgu achos mae hi'n gallu bod yn erm *you know* mae hi'n pwyntio mas os ti wedi camdreiglo a pethau fel yna mae hi'n eithaf *annoying* ond dw i yn caru hi

[just posh Welsh I'm not sure – there's no you know she's quite a snobby woman so ... I try to be more formal with granny try to use more Welsh with here whereas with Gerwyn I would just – you know say things – like talking with – like I'm talking with you now erm and with granny because she can be erm you know she points out if you've mutated wrong and stuff she is quite annoying but I love her]

George sociolinguistic interview

mae nain bach fel oh mae *dialect* hi o Cymraeg yw'r un cywir a mae'r ffordd mae pawb arall yn siarad Cymraeg yn anghywir so os 'dych chi ddim yn siarad Cymraeg fel mae nain yn siarad 'dych chi'n wneud o'n anghywir a bydd hi yn cywiro chi erm so fi'n cofio wnaeth mam pan es i lawr pan ro'n i'n bach bydd mam yn dweud os chi am gwylio teledu gallwch chi o leiaf gwylio bach o fel S4C neu Cyw a bydd nain yn rhoi e ar ond wedyn os oedd rhywun bach gogleddol yn siarad byddai hi'n troi e *off* ro'n i fel “Nain!”

[granny is a bit like oh her dialect of Welsh is the correct one and the way everyone else speaks Welsh is wrong so if you don't speak Welsh like granny speaks it you're doing it wrong and she will correct you so I remember mum – when I went down when I was little mum would say if you're going to watch television you can at least watch a bit of S4C or Cyw and granny would put it on but then if there was someone speaking a bit northern she'd turn it off and I was like “Nain!”]

Sally sociolinguistic interview

A detailed qualitative exploration of this data would shed further light on these attitudes, but it is beyond the scope of the current thesis.

However, more use of informal variants and less systematic style-shifting with them between contexts could be attributed to the fact that Gwynedd is leading a process of change. Further work would be needed to compare the use of these features across cohorts of different ages to assess this further. Davies (2016) substantiates the language change claim for the colloquial pronoun variant in north Wales, showing that younger speakers are far more likely than older speakers to use this variant. *Really* could be even further along the process of language change, as its use is more widespread across different speech styles. Further work is needed to ascertain this pattern of generational shift for the *really* variant in both areas. It is possible that generational language change is less far along in its process in Cardiff, where

prescriptivism among the older generation is prominent, as shown in the extracts above. Further work would be required to verify this hypothesis.

To summarise, I have shown that patterns of stylistic variation for WHL speakers are strongest for features which are associated with language contact, and more prominent in Cardiff than in Gwynedd, which could be linked to the socio-economic status of the pupils attending the schools and the revitalisation status of the language in the area. In terms of language change, we can see that Gwynedd are leading the process of change, with colloquial possessive pronouns and the intensifier *really* growing in prominence, and Cardiff are less far along on this process, with stronger systematic variation observed. Next, I turn to answer the second research question and present what we have learnt about patterns of variation in EHL speakers, and whether these match their WHL counterparts' patterns.

8.3 The acquisition of sociolinguistic competence for EHL pupils (RQ2)

Section 4.7.3 noted that this thesis would use a comparative variationist sociolinguistic framework to determine whether or not WHL patterns of variation have been acquired by EHL pupils. This followed previous work by Poplack (e.g. 2000) and Tagliamonte (e.g. 2006) and later Schlee and others (Schlee *et al.* 2011; Schlee 2013). Three lines of evidence have been presented in each analysis chapter in order to show not only whether each factor is a significant constraint on variation for each speaker group, but also the size of the effect and the hierarchies within each factor (Gardner 2023). In order to ascertain whether patterns of variation have been acquired by EHL students, we must first ascertain to what extent there is a pattern in WHL speech to begin with.

The main focus of this thesis was to determine whether stylistic patterns have been replicated by EHL speakers, and overall context is the strongest predictor of variation for these three features (although not always significantly so). The results chapters also presented whether internal, linguistic patterns observed in WHL speech have been acquired by EHL

284

speakers, and in this discussion, I present an argument to explain the strongest patterns observed. Recall, however, that I am discussing factors which have not all been found to be significant predictors of variation. The caveat presented in the Methodology chapter prevails: due to this being introductory and exploratory work, some conjecture about possible patterns in WHL speech still exists following the analysis of this data.

For now, however, I turn to discuss whether stylistic patterns have been acquired. Previous variationist sociolinguistic literature on phonological and phonetic variation in Wales (Mayr *et al.* 2017; Mennen *et al.* 2020) have shown no consistent effect of home language on the production of certain variables, which has been attributed to peer group dynamics overriding home language differences. In some cases, however, phonological variables patterned distinctly in speaker groups with different home languages (Gruffydd 2022; Morris 2013). Based on the literature, therefore, some differences between EHL and WHL speakers were expected. However, what could have been stark contrasts when these pupils were younger (see Hatton's (1988) description of EHL as monostylistic between 7 and 11 years old), may have now been over-ridden by peer group influence, as expected based on the literature. EHL group patterns are shown in Table 8.2 below, and where WHL patterns and significance have been replicated by EHL, these have been shaded (yellow for Gwynedd, green for Cardiff).

Table 8.2: Summary of lines of evidence for context across features (WHL vs EHL)

	Gwynedd WHL		Gwynedd EHL		Cardiff WHL		Cardiff EHL	
	Pattern (rates)	Sig?	Pattern (rates)	Sig?	Pattern (rates)	Sig?	Pattern (rates)	Sig?
Colloquial Possessives	Socio interview 74%	no	Socio interview 98%	no	Socio interview 56%	yes	Socio interview 75%	yes
	Peer to peer 68%		Job interview 84%		peer to peer 53%		Peer to peer 50%	
	Job interview 47%		Peer to peer 80%		job interview 6%		Job interview 17%	
Periphrastic Verbs	Job interview 78%	no	Job interview 89%	no	Peer to peer 79%	yes	Peer to peer 83%	no
	Peer to peer 72%		Peer to peer 87%		Job interview 68%		Job interview 83%	
	Socio interview 65%		Socio interview 87%		Socio interview 66%		Socio interview 81%	
<i>Really</i> Intensifier	Socio interview 79%	no	Socio interview 74%	no	Socio interview 85%	yes	Socio interview 52%	no
	Job interview 69%		Peer to peer 59%		Peer to peer 54%		Peer to peer 44%	
	Peer to peer 67%		Job interview 44%		Job interview 30%		Job interview 21%	

The patterning of rates in Table 8.2 and the extent to which this is significant shows clearly that EHL speakers are style-shifting according to the formality of the context. Having said that, however, in the case of all three features, EHL were less constrained by the speech context (smaller effect sizes, and fewer instances of significance) than WHL speakers (apart from the case of possessive pronouns for Gwynedd EHL). Recall that context consistently emerged as a significant predictor of style-shifting in the CWHL speaker group, and this was not always replicated to a significant extent by CEHL speakers.

As was the case with WHL speakers, the extent to which context is considered an important predictor of variation for EHL depends on the feature under study and stigma could account for this; where more stigma is associated with the feature under question, we see evidence of more careful style-shifting. These replicated patterns of variation are evidence that EHL speakers have acquired key elements of sociolinguistic competence. This clearly disputes findings from studies on French immersion (Mougeon *et al.* 2010) and claims closer to home (Hatton 1988: 251) which reported EHL pupils to be “monostylistic” (in their use of mutations with the literary possessive pronoun) and over-reliant on formal variants. It is clear from this thesis that EHL pupils produce a range of variants in different speech contexts, and in some cases actually use more informal variants overall (in contrast with previous studies). As predicted, therefore, by putting WHL and EHL students in classes together all pupils have the opportunity to learn and replicate patterns of stylistic variation. This underlines the success of the Welsh immersion system; EHL speakers largely match the patterns of WHL speakers.

8.3.1 Use of more informal variants

Table 8.2 shows that for possessive pronouns and periphrastic verbs, EHL pupils use higher rates of informal variants than WHL pupils. This is in contrast with other studies on the acquisition of sociolinguistic competence, which found that Canadian immersion students (Mougeon *et al.* 2010) and French L2 learners in Ireland (Dewaele and Regan 2001) and in the

USA (Kinging 2008) over-rely on more formal variants because using marked informal variants is deemed to be “sociopragmatic risk-taking” (Dewaele and Regan 2001, cf. Mougeon *et al.* 2010: 10). For possessive pronouns and Simple Past periphrastic verbs, EHL pupils seem unaffected by this type of risk. Indeed, in the context of Welsh immersion education, where pupils have shared classes with WHL pupils since the beginning of their school careers, this may not be considered a risk at all.

Recall that for the Simple Past verbs feature, EHL pupils used significantly more informal variants than WHL pupils. Why could this be? It is possible that we are observing a preference among EHL pupils for grammatical simplicity. I have previously presented the argument from grammarians of Welsh who have argued that the morphological complexities of inflecting verbs may prompt the speaker to use the periphrastic construction (B. M. Jones 1993: 156), which King (2016: 30) has alluded to as being beneficial to learners of Welsh. It is possible that EHL pupils favour the periphrastic variants for this reason.

Having said that, the intensifier *really* is being significantly under-used by EHL pupils, by contrast. I cannot argue a lack of grammatical simplicity in the case of *really* because it is a direct loan from the home language of the pupils in question. In both areas, EHL pupils are not using *really* to the same extent as their WHL counterparts, across all contexts of use. EHL pupils may be avoiding this because of the associated stigma with language contact. As (marginally) later bilinguals, the codeswitch may be sociopragmatically ‘risky’ for this group because of its direct link with English. Further work exploring the attitudes of Welsh immersion pupils from the two home language backgrounds towards the use of borrowed items such as these could shed further light on this. I provide a more detailed discussion of the implications of language contact below.

8.3.2 *Language contact and switching*

I have shown that WHL speakers used more instances of the *really* variant, but were highly constrained by the context of use (i.e. it was mainly used in informal situations), and in section 8.2 I argue that this is attached to the stigma associated with language contact. This is supported by further evidence in my analysis chapters of WHL speakers being constrained in their use of all features by English borrowing in the morphosyntactic structure, which prompted more (and in some cases categorical) use of the more colloquial variants for all three features. Recall that WHL were more likely to use the sandwich and literary possessive pronouns when the language of the possessed noun was Welsh, e.g. *fy ngwaith cartref* [my homework], whereas English nouns were more likely to prompt the use of the colloquial variant e.g. *homework fi* [my homework]. Where the main verb in Simple Past tense was English, WHL students always used the periphrastic variant e.g. *wnes i panicio* [I panicked], whereas Welsh verbs prompted the use of inflected verbs e.g. *eisteddais i* [I sat]. Where the adjective being intensified was English, *really* was more frequent than *mor* or *iawn* for WHL speakers. This contributes to the evidence that English borrowings are used in more colloquial styles of Welsh, and WHL speakers' use of informal variants is strongly associated with borrowings. This is the first work of its kind to remark on this pattern.

EHL groups, on the other hand, were less influenced by English borrowings overall. Instead, it appears that EHL speaker groups in Gwynedd and in Cardiff are less concerned about using formal and informal variants with English borrowings. This was particularly the case for the intensifier *really*, where the language of the proceeding adjective was the weakest predictor of variation for both EHL groups. Indeed, CEHL did not even follow the same pattern as their CWHL counterparts with regards to English adjectives; for CEHL, Welsh adjectives were more likely to be used with *really*. This pattern was not as clear with regards to possessive pronouns; GEHL did not vary their use of possessive variant according to the language of the

noun (however, rates of the colloquial variant were high overall), whereas this was more important for CEHL whose use of the variant was significantly predicted by the language of the noun.

Overall, however, we see further evidence here that EHL Welsh speakers are more freely integrating borrowed items into their morphosyntax. Previous work which has considered how children integrate English borrowings into the Welsh mutation system found that EHL speakers were less likely to do this than WHL speakers (Bellin 1988). Bellin claimed that the EHL children had not yet acquired the critical mass of exposure required to generalize mutations across English borrowings. I have found clear evidence that morphosyntactic generalizations are being made by EHL when using English borrowings, yet these generalisations do not reflect the systematic style-shifting observed in their WHL counterparts' speech. It is possible that although a critical mass of exposure has been accessed, they have not yet acquired this as a 'systematically organized system' (Thomas & Mayr 2010: 111).

EHL pupils in Gwynedd seem further along in "systematising the variation" (Schleef 2013: 209) for the *really* variant than Cardiff EHL pupils, because GEHL are more constrained by context, and also follow their GWHL counterparts' pattern in terms of using *really* more with English adjectives. Cardiff EHL's higher use of *really* with Welsh adjectives (compared with WHL speakers) is symptomatic of the type of 'off the shelf' variation (Milroy 2007), where *really* is easy to pick up and use in Welsh without having to re-work the syntactic structure. The ease with which this type of variation can occur, might contribute to the CEHL group's less constrained use of the *really* booster overall, and with adjectives of different languages.

8.3.3 *The role of input*

As presented above, EHL students are acquiring most of the stylistic patterns of their WHL counterparts. This could provide evidence that Welsh immersion education is supporting the

acquisition of sociolinguistic competence. In previous work on Welsh, Hatton (1988) argued that EHL pupils were restricted to a formal educational pattern of Welsh, compared with WHL pupils. Age could explain the difference between Hatton's (1988) work and my own. As has been found to be the case in Nance's (2020) work on pronunciation in Gaelic medium education, differences between home language groups are likely to be levelled out after primary school age, due to continued peer group exposure, particularly in adolescence. The continued longitudinal input (as well as peer group effect) is likely to have influenced the increased use of informal variants by the end of their school careers, for the pupils under study in this thesis.

Beyond continued exposure to Welsh through peers, the acquisition of sociolinguistic competence observed here could be related to the expectations and behaviour of teachers in the classroom. As discussed in 2.2.2.3, teachers report that they correct casual speech features in more formal contexts, indicating an expectation for students' sociolinguistic competence to develop (Young 2019). It is possible that in Welsh classrooms, the amount of variation in teacher input is sufficient to allow students to acquire variation. It was not the aim of this thesis to examine the role of teacher input, but further work could address this.

However, even stigmatised features which have been shown to be disfavoured by teachers (Young 2019) are acquired by students. Work by Young found that colloquial possessive pronouns were significantly likely to be corrected by teachers (based on self-report data) if used in a context deemed to be inappropriate. This suggests that peer group could influence the production of this feature more than education influence. Indeed, it seems unlikely that teacher expectations and corrective feedback are the driving force behind the style-shifting observed in students.

8.3.4 The registers

As previously discussed in the feature analysis chapters, for EHL speakers in Gwynedd, it is possible that the peer-to-peer context was not deemed a particularly informal context, due to

the forced nature of the Welsh language task in what would otherwise be a naturally English-speaking peer group. Indeed, during their sociolinguistic interview with the researcher, students in two participating peer groups in Gwynedd alluded to never speaking Welsh with their friends (of whom some were in the task peer group and others in their wider friendship groups) outside of the school setting. It is therefore possible that the imposed peer group task felt artificial for these speakers, rather than this speaker group not having acquired the sociolinguistic variation observed in their WHL counterparts.

On another note, relating to register, a key finding of this work is that not only are EHL speakers acquiring variation, but they even demonstrate variation with a single interlocutor. In this thesis, greater differences were observed between the sociolinguistic interview and mock job interview, which involved the same interlocutors, compared with the difference between the peer-to-peer context and sociolinguistic interview, which involved different interlocutors. It is possible that audience design (Bell 1984) is a weaker model of variation to explain the findings in this work. Indeed, as in Rickford and McNair-Knox's (1994) "Foxy Boston" study, variants varied more between topics than between the interviews overall, showing that more than one element of register can affect stylistic variation in any given situation.

Although this research design was successful in eliciting stylistic patterns across contexts with a single interlocutor, these speech contexts were merely an online proxy for register, because of limitations faced by COVID-19. Further work could confirm whether face-to-face speech contexts with different interlocutors also predict the use of different speech styles for young speakers of Welsh. Other speech contexts would need to be comparable to the 'life-worlds' (Sharma 2011: 474) of participants, meaning that we can make further generalisations about the stylistic practice of participants.

8.3.5 Awareness

As discussed, overall, most constraints have been acquired and replicated across groups, providing evidence that sociolinguistic competence has been matched by EHL students. This can be somewhat explained by the level of awareness speakers have towards the constraint. The acquisition of high-awareness variables (associated with social meaning and identity) differs in character from the acquisition of low-awareness variables (which are acquired more automatically through contact) (Tarone 2007: 844; Howley 2015: 119; Ryan 2018: 265). Context, which has been replicated for possessive pronouns and intensifiers, could be argued to be a high-awareness variable. These constraints are likely to have been acquired because of the speakers' understanding that there is a socially-constructed requirement for them to adjust their language to suit the speech context. The awareness in Welsh-medium education could be linked with higher levels of stigma associated with the more informal variants.

8.3.6 Frequency

In order to ascertain whether the frequency of each feature in the data was likely to affect the extent to which its use is replicated across groups, the reader is reminded of the total number of tokens per feature in Table 8.3.

Table 8.3: Tokens per feature

Feature	Tokens
Pronouns	1968
Verbs	1492
Boosters	785

As has been remarked, the patterns of style-shifting are stronger in Cardiff overall than in Gwynedd, but the close replication of the possessive pronouns feature stylistic pattern by CEHL pupils could be connected to the frequency with which the feature appears in spoken

Welsh. Boosters, on the other hand, appear less frequently in the dataset. However, this does not lead to weaker patterns of replication by EHL groups in either area and I therefore posit that frequency of the feature does not play an important explanatory role in the acquisition of sociolinguistic competence in Welsh for the features under study.

8.3.7 Transformation under transfer

The regression models in the three analysis chapters showed that EHL speakers in Cardiff replicated constraints from WHL speakers, but also showed an additional constraint for each feature. For the possessive pronouns feature, noun language was a significant predictor of variation, for the verbs feature, grammatical person was a significant predictor of variation, and for the intensifiers feature, adjective quality predicted patterns of variation for the EHL group. These findings are consistent with previous literature on the acquisition of sociolinguistic variation, which show that constraints not found in ‘native speech’ do appear among ‘new speakers’ (Mougeon *et al.*, 2004; Rodríguez-Ordóñez & Sainmaza-Lecanda, 2018; Rodríguez-Ordóñez 2022). These reallocations of constraints are explained by Meyerhoff and Schlee (2012: 405) as undergoing transformation under transfer; this describes the “reallocation of the relative importance of variable input constraints in the output variation”. This has in part been explained by the interference of governing constraints of the home language on the acquisition of English variability, and also by limited exposure to the language. However, on analysing evidence of the emergence of grammatical category constraints in migrant learners’ use of English not found in native speakers’ English, Meyerhoff and Schlee (2012: 406), also attribute this to exposure to other varieties of English beyond the speech of their locally-born peers. Further work is required in order to ascertain whether CEHL are replicating patterns beyond their peer groups’, in Welsh encountered in the media, etc.

8.3.8 Summary of acquisition of sociolinguistic competence

In summary, EHL students are replicating many of the constraints observed in WHL speech, particularly in relation to style-shifting, showing that they have acquired a level of sociolinguistic competence. This is likely to be related to continued input from peers with varying degrees of exposure to Welsh outside of the classroom, rather than the educational input itself. However, EHL students on the whole are using higher rates of informal variants than their WHL counterparts, which could be tied to the grammatical simplicity in forming the morphosyntactic structures (for possessive pronouns and periphrastic verb variants). EHL students are also less constrained in their use of codeswitches than they are with other established Welsh features, showing that they have yet to systematise the variation of features strongly associated with language contact. Their acquisition of varying constraints can be explained in part by the level of speaker awareness towards each predicting factor, and the high awareness of the context variable could, in part, be related to the level of stigma associated with the style-shift. Where constraints have not been acquired, this could relate to the perception of the formality of the task at hand, and could be addressed by tweaking the research design in future work.

The next section answers the final research question by addressing the influence of community exposure to the language on the acquisition of sociolinguistic competence in EHL pupils.

8.4 Differences according to Welsh in the community (RQ3)

Recall that in Gwynedd, 68% of the population reported using the language on a daily basis, compared with Cardiff which stood at 6.6% (Annual Population Survey 2023). This data shows us that in Cardiff there are far fewer opportunities to use the language in the community than in Gwynedd. The discussion so far has shown that, on the whole, sociolinguistic competence is demonstrated by EHL groups in both areas. Here I ask whether community exposure to the

language affects the acquisition of sociolinguistic competence among pupils in Welsh immersion education. Table 8.2 highlighted in yellow and green where style-shifting patterns had been replicated by EHL speakers in the two areas. It is clear to see that more cases of replication were found in Cardiff than in Gwynedd. This is a somewhat surprising finding, given that community exposure is so much higher in Gwynedd.

This could be due to the fact that CWHL demonstrate stronger (and thus more easily replicable) patterns than GWHL. I argued in section 8.2 that Gwynedd are leading in the process of language change for the two more stigmatised features, and thus patterns of stylistic and linguistic variation are less pronounced. In Cardiff, on the other hand, WHL speakers are more conservative, and the closer patterning of EHL speakers could, in part, be more strongly tied to socio-economic status than to community exposure. The fact that GWHL style-shift less could be the main explanation for the comparable patterns seen in the GEHL pupils. What is vital from these findings is that GWHL and CWHL pattern in different ways, and that this is likely to explain the EHL group differences.

It is recognised that ‘community’ for young speakers can be defined beyond the reported use associated with the area in population surveys. For example, it is difficult to argue the significance of meaningful community exposure in a context where many EHL pupils in Gwynedd claimed not to speak Welsh with their peers. In contrast, even though Cardiff community use of Welsh is reportedly low, all participating Cardiff speakers reported using Welsh ‘some of the time’ with peers. It can be argued therefore that the micro-level school community, may in actual fact, contain more Welsh than is reported at local authority level. Whereas communities of practice were not found to be significant predictors of phonological variation in recent work on Welsh immersion pupils (Gruffydd 2022), future research could examine the extent to which peer group use of Welsh predicts levels of stylistic variation.

Previous work has systematically analysed the levels of extra-curricular exposure to the target language; for example, Rehner *et al.* (2003) and Mougeon *et al.* (2010) found trips to Francophone areas and stays with Francophone families to be important predictors of immersion students' acquisition of sociolinguistic competence. Future work on the acquisition of sociolinguistic competence in Welsh could further examine the role of extra-curricular exposure to the target language by measuring engagement with Welsh-language media (as in Morris 2013), the *Eisteddfod*, and *Urdd* activities, which promote the use of Welsh among younger speakers.

I have answered the three research questions set out in section 1.4, and conclude that overall, sociolinguistic competence is acquired in Welsh-immersion education. Clear distinctions have been found in use of Welsh across contexts, with rates of more informal, non-standard, or stigmatised features increasing in contexts which are more casual. This has implications for the way Welsh is used and taught. I discuss the implications of this finding below.

8.5 Implications of the research findings

8.5.1 Implications for social use

Work by Mougeon *et al.* (2010) posited that students might be less likely to use French socially because they were not acquiring the informal varieties required for colloquial conversations with traditional speakers. In contrast with that work, we have seen evidence that students can grasp different styles in Welsh after having attended Welsh-medium education, however, we know that there are still some barriers to young people's social use of Welsh. Indeed, most of the participating students stated in their sociolinguistic interviews that they either spoke a combination of Welsh and English or mostly English with their peers. Evidence from Welsh

Government⁵⁰ (see Figure 8.1) shows that only a small minority (10%) of students use Welsh ‘always/almost always’, even though many of them report that there are enough opportunities to do so in their communities (Welsh Government 2022).

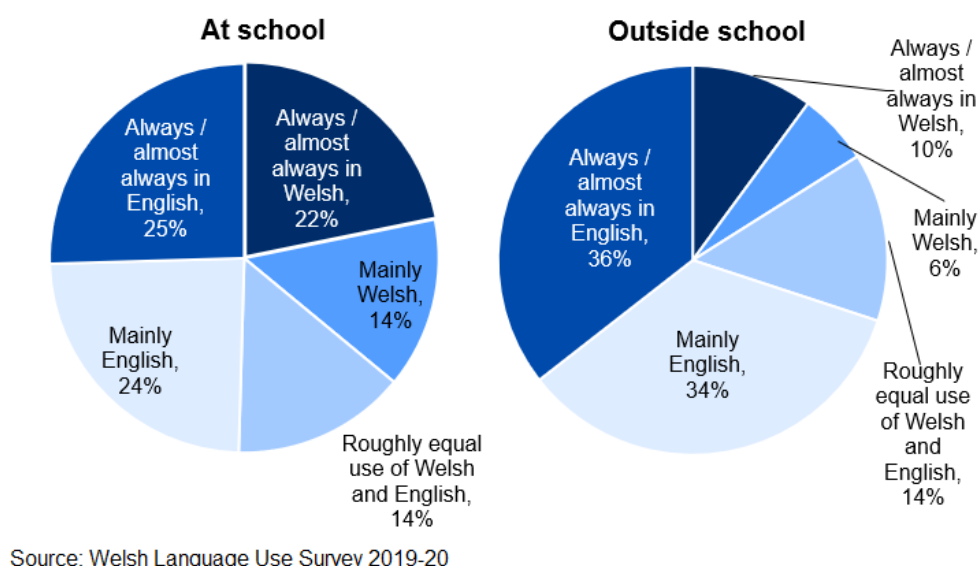


Figure 8.1: Reported use of Welsh at school and outside school (Welsh Language Use Survey 2019-2020)

The question therefore remains as to why students are opting out of using their Welsh outside of school, if they have acquired the ability to vary their language appropriately. It has been suggested by some research that the use of Welsh between peers outside of formal school contexts appears to be highly stigmatised (Morris 2014: 85). Further work is needed to ascertain what barriers remain which limit EHL pupil’s social use of Welsh. Although my work has found that sociolinguistic competence has been acquired, it remains to be seen whether these speakers identify themselves as sociolinguistically competent users of the language. It might be that the perception of competence differs from observed competence, and that this contributes to reduced social use. Future work could compare actual competence with perceived competence in EHL and WHL students.

⁵⁰ <https://www.gov.wales/social-use-welsh-language-july-2019-march-2020-html#109595>

8.5.2 *Parent concerns*

Non-Welsh-speaking parents who choose to send their children to Welsh-medium school can be reassured, based on these findings, that in terms of the acquisition of the variable features mentioned in this thesis, their children will share patterns of variation with WHL pupils. This will provide positive encouragement to some, as the Welsh-medium education sector grows with increasing numbers of pupils from non-Welsh-speaking backgrounds (Jones 2012). In this thesis, I argue that Welsh-medium schools do succeed in creating an environment where home language background does not have an effect on the development of a stylistic repertoire.

*Recommendation 1: During school admission activities
local authority officials could communicate to parents the
potential for EHL pupils to reach high levels of
sociolinguistic competence in Welsh-medium schools*

8.5.3 *Moving away from ‘the standard’*

It has been argued that Welsh-medium education is having a standardizing effect on Welsh. As Sayers (2009: 293) notes, ‘If education is the main reason for increasing Welsh use, then the kind of Welsh being used is more likely to be influenced by that education’. M. C. Jones’ (1998) study shows some evidence for this; whilst comparing the Welsh local dialects of two areas, she found that those who attended Welsh-medium schools were using more standard features than children of a similar age in English-medium education, pointing to the standardizing effect of Welsh immersion.

However, the analysis in this thesis found evidence of a large proportion of non-standard variants in possessive pronoun use by young people in Welsh-medium education. These were overall more evident in the speech of EHL students, who supposedly have limited access to

Welsh outside of the education setting. The fact that all groups used non-standard forms overall, and that EHL are using stigmatised features in some cases more than their WHL counterparts, shows that Welsh immersion education is not having the clear standardizing effect proposed in the previous literature (Sayers 2009; M. C. Jones 1998). We see in this Welsh data that speakers do not adhere consistently to either "standard" or "non-standard" forms. Instead, their speech tends to exhibit a mixture of both variants. This inherent inconsistency in language use reflects the dynamic nature of linguistic variation and the diverse ways speakers employ linguistic features.

In June 2023, Welsh Government published a policy⁵¹ priority to coordinate and develop Welsh linguistic infrastructure. Part of this policy plan involves establishing a Welsh Language Standardization Panel to “facilitate the development of easy-to-use language guidance. This could include guidance, advice and answers concerning, for instance, conjugating verbs, mutations, appropriate use of register in certain contexts” (Welsh Government 2023). It is important at all levels of language planning that due consideration is given to the harmful nature of standard language ideologies, particularly in minoritized settings. As previously mentioned, standard varieties of language are associated with historically entrenched elites and often associated with prestigious written registers (M. C. Jones 1998) and are legitimized by standard language ideologies (SLIs) which promote negative attitudes towards deviations from the idealised standard norm, drawn primarily from the spoken language of the upper middle class (Lippi-Green 1997: 64). Evidence of these SLIs can be seen in Welsh teachers’ negative attitudes to informal or colloquial variants (Young 2019).

⁵¹ <https://www.gov.wales/welsh-linguistic-infrastructure-policy-html>

I would argue that there is plenty of evidence in this thesis that young Welsh speakers are acquiring language appropriate to different settings. I recommend that SLIs are challenged at all levels of Welsh education provision.

Recommendation 2: Academic work on Welsh should continue to challenge SLIs surrounding the use of Welsh in education

8.5.4 Teaching resources

Mougeon *et al.* (2010) found that French Language teaching materials did not include marked informal variants, which the authors associated with the L2 speakers' limited use of those marked informal variants. I have shown that EHL students in Welsh-medium education are acquiring marked informal variants. Following reports by teachers about how they tend to disfavour non-standard varieties and promote more formal language in more formal settings (Young 2019), it is unclear to what extent teacher input and teacher resources contribute to the development of sociolinguistic competence in Welsh immersion contexts. Further work could examine teacher talk and resources in these contexts to assess the extent to which students are acquiring or learning these norms from their teachers.

In previous chapters, I explained that L2 teaching materials for adult learners (e.g. Learn Welsh 2023a; 2023b) do not reference such informal variants as were found in the speech of young people. The possessive pronoun variant is clearly undergoing a process of language

Recommendation 3: Learn Welsh resources should consider including non-standard variants which are frequent in L1 speech.

change and becoming the most frequent pronoun variant in the speech of young people (Davies 2016). Mougeon *et al.* (2010) make a case for materials reflecting L1 speakers' variation so as to promote similar acquisition in L2 language immersion speakers. It is vital that L2 learners who are acquiring the language later in life are exposed to the varieties present in the speech of young Welsh speakers – if not, we are deepening the divide between 'learner' speech and 'native' speech. Newcombe (2007: 111) makes the case for more naturalistic teaching materials which can help bridge the divide between being an L2 learner and an L2 speaker. I thus recommend that Learn Welsh resources include the colloquial pronoun variant and other non-standard variants which are known to be in common use.

In a similar vein, reference books and grammars should reflect the actual use of Welsh by looking at the process of language change and should make more mention of features which are subject to variation. Young speakers are using very high rates of the intensifier *really* and further work could compare this to older speakers (by using the Siarad corpus, for example) to rule out age grading. The method I employed to determine whether a historical borrowing from English has been incorporated into Welsh by using Geiriadur Prifysgol Cymru appears arbitrary, when comparing the frequency of the intensifier *really* (which is not considered a Welsh item) and infrequency of *cweit* and *reit* (which are both considered to be Welsh items). It has been argued that the use of English calques and borrowings is a feature which distinguishes colloquial from standard usage (A. R. Thomas 1987). According to A. R. Thomas, 'doublets' (1987: 107) exist in Welsh, where a loan and a source language word can occur in either casual or formal contexts respectively, for example *miwsig* [music] which is the more colloquial form of the formal variant *cerddoriaeth* [music].

As I noted in chapter 7, *really* has been found to be one of the most frequent markers of intensity in American and British English colloquial conversation (Labov 1984; Lorenz 2002; Ito and Tagliamonte 2003; Tagliamonte and Roberts 2005). Considering the rise in its frequency in the speech of young people (as demonstrated in this thesis) and older people (Welsh-speaking readers of this thesis will attest to its widespread use across age-groups), I recommend that *really* (or *rili*) is recognised as a borrowing in Welsh and incorporated into reference materials such as Geiriadur Prifysgol Cymru.

Recommendation 4: Reference materials should include frequent and stylistically constrained loanwords.

8.5.5 Welsh ‘immersion’

In Wales, because home language backgrounds are mixed in classes, the term ‘immersion’ in fact applies to the requirements of the individual student, rather than the education context itself. If students come from a EHL background, they are being immersed, but if they are from a WHL background it is just Welsh-medium education.

I have shown that EHL are acquiring sociolinguistic competence in terms of the main stylistic and internal patterns, which has also been found in other recent work, looking at pupils of a similar age to the ones in this thesis (Gruffydd 2022). Although older research found stark differences between pupils of home language backgrounds at a younger age (Hatton 1988)⁵² my work suggests that by the end of compulsory school age, these differences are levelled out, drawing EHL pupils closer to the norms of WHL pupils.

⁵² It is recognised, however, that since this research was undertaken, Welsh education and wider attitudes towards Welsh have changed.

In relation to other contexts where sociolinguistic competence can be acquired, I suggest that Welsh-medium education sits further along the continuum than previously alluded to (see literature review, Table 2.1). Table 8.4 presents an adaptation of the continuum.

Table 8.4: An (adapted) continuum for acquiring sociolinguistic competence in various settings

Context of acquiring sociolinguistic competence	Home language	Welsh “immersion”	Study abroad / migrant language	French immersion education	SLA
	<div> <div>more sociolinguistic competence acquired</div> <div> <div></div> <div>less sociolinguistic competence acquired</div> </div> </div>				

It is clear that the term immersion does not fully apply to the type of education attended by the participants in this thesis, and it is certainly not comparable to French immersion contexts whereby students fail to acquire sociolinguistic competence.

Recommendation 5: Due consideration should be given to the continued use of the term ‘immersion’ in Wales, which is not consistent with international contexts of immersion.

Having presented a discussion of the main findings and their implications, the final chapter presents the conclusion to this work.

9 Conclusion

This doctoral work set out to examine the patterns of sociolinguistic variation in pupils from Gwynedd and Cardiff, from EHL and WHL backgrounds. The purpose of this was to determine to what extent EHL pupils are acquiring sociolinguistic competence which is comparable to their WHL counterparts. The current chapter summarises the main conclusions of this work.

For the first time, I have been able to show that EHL pupils are acquiring the more salient WHL patterns of variation, even when those variants are stigmatised, and they are likely to be discouraged from using them in the classroom. The fact that there is stylistic variation in all groups suggests that the overt ideology surrounding what constitutes standard and non-standard and casual and non-casual variants of speech, (thus, their sociolinguistic competence) is being acquired even if there are minor differences between areas and home languages.

Having undertaken this work, we now know that Welsh immersion students who do not speak Welsh at home are acquiring the stylistic patterns of their WHL peers. Although these two speaker groups have been found to diverge in a number of phonological studies (Morris 2013, Gruffydd 2022), in terms of morphosyntax, they share far more patterns. Those variable patterns of use have been investigated in detail throughout the results and discussion chapters of this thesis. From the features examined in this work, we can rule out there being morphosyntactic patterns of variant use which are “typical” of EHL or WHL use – overwhelmingly, their patterns are shared. This finding is different to other studies which have looked at the acquisition of sociolinguistic competence in classroom settings (e.g. Mougeon *et al.* 2010).

Because of the reportedly stigmatised nature of two of the features which are used overwhelmingly by all groups, we can surmise, for the first time, that EHL students must be modelling their speech on their peers’ speech, rather than on the standard target variety they are exposed to at school. This supports previous arguments about differences between home

305

language backgrounds levelling out over time, where pupils from different language backgrounds are taught in the same classrooms, resulting from long-term language contact (Mayr *et al.* 2017; Nance 2020).

Differences between the areas were observed for all three features analysed, and this thesis found strong evidence for Gwynedd leading the process of language change. Speakers in Gwynedd were less constrained in their use of informal, non-standard and stigmatised variants. Cardiff on the other hand, showed more significant variation in their use, which is tied to more contextual and linguistic factors. Further to this, the patterns analysed in this work show that CEHL are patterning more closely with CWHL than GEHL are with GWHL, despite dominance of Welsh in the community in Gwynedd. I have suggested that future work might consider community at a more micro level in order to determine whether peer group language use might affect this more than reported use at local authority level.

Preconceptions and generalisations about monostylistic EHL speakers (see Hatton 1988; Hughes 2013) could be rooted in variables other than morphosyntactic features (which have been shown to vary across contexts). This raises a question about the variables which give this impression. Both Robert (2009) and Gruffydd *et al.* (2024) found that the Welsh accents of WHL pupils are highly likely to be identified correctly by other speakers. On the other hand, those who acquired Welsh at school (EHL) scored less well in social measures such as openness, friendliness and politeness. I suggest that these phonological variables can explain the generalisations made about EHL speakers, because as I have shown, in terms of morphosyntax, they are mostly matched with their WHL peers.

The current thesis makes a significant contribution to the study of stylistic variation in Welsh, and further variationist work of this kind (instead of over-relying on impressionistic accounts of Welsh language use) is required to build a broader picture of patterns relating to style, particularly in relation to whether WHL features are replicated by EHL speakers. Indeed,

the widespread use of the colloquial possessive pronoun and the *really* booster variant suggest that reference materials need updating; according to the use by young people, these are not stigmatised, these are just considered less formal. Further work could examine other features described as ‘stigmatised’ which are thought to vary according to style, as reported in the literature. Furthermore, the distinct areas of Wales examined in this thesis were found to show some evidence of community-specific norms, which should be a consideration for all future variationist work on Welsh.

More research is needed to understand how teachers use these variants in their classroom talk. However, the findings of this thesis have implications beyond the acquisition of Welsh in immersion classes, too. With the rise of online learning in recent years leading to an increase in the number of people learning Welsh as adults, particularly outside of Wales (Morris *et al.* 2024), we need to understand whether the patterns being taught match the patterns being used. Learn Welsh resources do not always mirror wider patterns of use, which bring into question whether the ‘learner’ varieties of Welsh are fit for purpose. I also make a number of other recommendations in the discussion chapter, which highlight the potential impact of this work.

Within immersion school contexts, however, the current thesis builds a picture of an immersion context which succeeds (at least in terms of morphosyntactic features), to create sociolinguistically competent speakers. To conclude, therefore, I highlight that learning with L1 peers is an excellent way to help L2 speakers reach sociolinguistic competence, which could be implemented in other L2 learning contexts, too. Although there is still a great deal to be done in order to investigate the acquisition of sociolinguistic competence of Welsh speakers, this research has systematically mapped patterns of variation in four speaker groups and has shown that EHL speakers are closely matching their WHL peers’ speech. Understanding this has important implications for Welsh language and education policy, and for the revitalisation of the language itself.

10 References

- Aitchison, J. and Carter, H. 1987. The Welsh language in Cardiff: A quiet revolution. *Transactions of the Institute of British Geographers*, 12, pp. 482-92.
- Aitchison, J. and Carter, H. 1994. *A Geography of the Welsh language 1961-1991*. Oxford: Alden Press.
- Aitchison, J. and Carter, H., 2004. *Spreading the Word: The Welsh Language 2001*. Talybont, Ceredigion: Y Lolfa Cyf.
- Armstrong, N. 2001. *Social and Stylistic Variation in Spoken French: A Comparative Approach*. Amsterdam: Benjamins.
- Arnau, J. 2000. Catalan Immersion Teachers: Principles of Language Teaching, *International Journal of Bilingual Education and Bilingualism*, 3(2), pp. 79-100, doi: 10.1080/13670050008667701.
- Awbery, G. 1976. *The Syntax of Welsh: a transformational study of the passive*. Cambridge Studies in Linguistics 18, Cambridge, Cambridge University Press.
- Awbery, G. 1994. Echo pronouns in a Welsh dialect: A system in crisis? *Bangor Research Papers in Linguistics*, 5, pp. 1-29.
- Ball, M. J. 1984. *Sociolinguistic Aspects of the Welsh Mutation System*. Unpublished doctoral thesis, University of Wales.
- Ball, M. J., Griffiths, T. and Jones, G. E. 1988. 'Broadcast Welsh'. In: M. Ball (ed.) *The use of Welsh: A contribution to sociolinguistics*, pp. 182-199. Clevedon: Multilingual Matters.
- Barakos, E. 2012. Language policy and planning in urban professional settings: bilingualism in Cardiff businesses, *Current Issues in Language Planning*, 13(3), pp. 167-186, doi: 10.1080/14664208.2012.722374
- Barakos, E. 2012. Language policy and planning in urban professional settings: bilingualism in Cardiff businesses, *Current Issues in Language Planning*, 13(3), pp. 167-186, doi: 10.1080/14664208.2012.722374
- Barnfield, K. and Buchstaller, I. 2010. Intensifiers in Tyneside: Longitudinal developments and new trends. *English World-Wide* 31(3), pp. 252-287.
- Bayley, R. and Regan, V. 2004. Introduction: The acquisition of sociolinguistic competence. *Journal of Sociolinguistics*, 8(3), pp. 323-338. doi: 10.1111/j.1467-9841.2004.00263.x.
- Bayley, R., Holland, C., and Ware, K. 2013. Lexical frequency and syntactic variation: A test of a linguistic hypothesis. *University of Pennsylvania Working Papers in Linguistics*, 19(2), pp. 21-30.

- Ball, M. J., Griffiths, T. and Jones, G. E. 1988. Broadcast Welsh. In Martin Ball (ed.), *The use of Welsh*, 182–99.
- Bell, A. 1977. *The language of radio news in Auckland: A sociolinguistic study of style, audience, and subediting variation*. Unpublished doctoral thesis, University of Auckland.
- Bell, A. 1984. Language Style as Audience Design. *Language in Society*, 13(2), pp. 145-204.
- Bell, A. 2001. ‘Back in style: Reworking audience design’. In: P. Eckert and J. R. Rickford (eds.) *Style and sociolinguistic variation*. Cambridge, UK: Cambridge University Press, pp. 139-169.
- Bellin, W. 1988. ‘The development of pronunciation’. In: M. Ball (ed.) *The use of Welsh: A contribution to sociolinguistics*, pp. 212-228. Clevedon: Multilingual Matters
- Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. 1999. *Longman grammar of spoken and written English*. Harlow, UK: Longman.
- Binks, H.L. and Thomas, E.M. 2019. Long-term outcomes for bilinguals in minority language contexts: Welsh-English teenagers’ performance on measures of grammatical gender and plural morphology in Welsh. *Applied Psycholinguistics*, 40(4), pp. 1019–1049. doi: 10.1017/S0142716419000110.
- Birnie, I. 2022. Blended learning to support minority language acquisition in primary school pupils’ lessons from the “taking Gaelic Home Study”, *Australian and International Journal of Rural Education*, 32(2), pp. 126–141. doi:10.47381/aijre.v32i2.329.
- Borsley, R. D., Tallerman, M, Willis, D. 2007. *The syntax of Welsh*. Cambridge: Cambridge University Press.
- British Educational Research Association. 2018 *Ethical Guidelines for Educational Research, fourth edition*. Available at: <https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018> [Accessed 20/07/2024].
- Bybee, J. and Hopper, P. 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins
- Bybee, J., Perkins, R. D., and Pagliuca, W. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.
- Canale, M., and Swain, M. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), pp. 1-47.
- Cardiff Council. 2024. *Ein dinas: ein hiaith* <https://www.ourcityyourlanguage.wales/3-to-11-years-old/?cn-reloaded=1> [Accessed 29/07/2024]

- Cardiff University. 2020. *Cardiff University weekly update*. Available at: <https://undergraduate-enews.cardiff.ac.uk/campaigns/reports/viewCampaign.aspx?d=t&c=F99FB036943F7985&ID=1220E33DA75191CF2540EF23F30FEDED&temp=False&tx=0&source=Report> [Accessed 20/07/2024].
- Chambers, J. K. 2003. *Sociolinguistic Theory: Linguistic Variation and Its Social Significance*, 2nd edition. Maiden, MA. Blackwell Publishers.
- Chan, J. Y. H. 2017. Stakeholders' perceptions of language variation, English language teaching and language use: The case of Hong Kong. *Journal of Multilingual and Multicultural Development*, 38(1), pp. 2–18.
- Clarke, E. (2019 March 14) Fourth Welsh secondary school hope for Cardiff as demand grows. *BBC News*. Available at: <https://www.bbc.co.uk/news/uk-wales-47528827>
- Clayton, P. 1978. 'Domain and register in the use of Welsh'. In: G. Williams, (ed.), *Social and cultural change in contemporary Wales*. London: Routledge & Kegan Paul, pp. 206–18.
- Cloud, N., Genesee, F. and Hamayan, E. 2000. *Dual Language Instruction: A Hand-book for Enriched Education*. Boston, MA: Heinle & Heinle.
- Coupland, N. 1980. Style-shifting in a Cardiff work setting. *Language in Society* 9, pp. 1–12.
- Coupland, N. 1985. 'Hark, hark, the lark': social motivations for phonological style- shifting. *Language & Communication*, 5(3), pp. 153-171.
- Coupland, N. 1988. *Dialect in Use: Sociolinguistic Variation in Cardiff English*. Cardiff: University of Wales Press.
- Coupland, N. 2001. 'Language, situation, and the relational self: Theorizing dialect-style in sociolinguistics'. In P. Eckert and J. Rickford (eds.), *Style and sociolinguistic variation*. Cambridge: Cambridge University Press, pp. 185- 210.
- 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, pp. 1–27.
- Cummins, J. 1998. Immersion education for the millennium: What we have learned from 30 years of research on second language immersion? Learning through two languages: Research and practice. *Second Katoh Gakuen International Symposium on Immersion and Bilingual Education* , pp. 34–37.
- Cymraeg Byw, rhifyn 1. 1964. Cyhoeddwyd ar ran y Gyfadran Addysg, Coleg Prifysgol Abertawe, gan Llyfrau'r Dryw, Llandybïe.

- Cymraeg Byw*, rhifyn 2. 1967. Caerdydd: Cyd-bwyllgor Addysg Cymru.
- Cymraeg Byw*, rhifyn 3. 1970. Caerdydd: Cyd-bwyllgor Addysg Cymru.
- Davies, J., 2014. *The Welsh Language: A History*. Cardiff: University of Wales Press.
- Davies, P. 2016. Age variation and language change in Welsh: Auxiliary deletion and possessive constructions. In: Durham, M. and Morris, J. (eds.) *Sociolinguistics in Wales*. London: Palgrave Macmillan, pp. 31-60.
- Denis, D. and Tagliamonte, S.A., 2018. The changing future: competition, specialization and reorganization in the contemporary English future temporal reference system. *English Language and Linguistics*, 22(3), pp. 403-430.
- Denis, D., Gardner, M. H., Brook, M., and Tagliamonte, S.A. 2019. Peaks and arrowheads of vernacular reorganization. *Language Variation and Change*, 31(1), pp. 43–67. doi:10.1017/S095439451900005X
- Dewaele, J-M. 2004. The acquisition of sociolinguistic competence in French as a foreign language: An overview. *French Language Studies* 14, 301–319. doi: 10.1017/S0959269504001814
- Dewaele, J-M. and Regan, V. 2001. The use of colloquial words in advanced French interlanguage. In: S. Foster-Cohen and A. Nizegorodcew (eds.) *EUROSLA Yearbook 1*, pp. 51-68. Amsterdam: Benjamins.
- Díaz-Campos, M. 2005. The Emergence of Adult-like Command of Sociolinguistic Variables: A Study of Consonant Weakening in Spanish-Speaking Children. *Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, pp. 56–65.
- Drager, K. 2009. *A Sociophonetic Ethnography of Selwyn Girls' High*. Unpublished doctoral thesis, University of Canterbury, Canterbury, NZ.
- Drummond, R. 2010. *Sociolinguistic variation in a second language: The influence of local accent on the pronunciation of non-native English speakers living in Manchester*. Unpublished doctoral thesis, University of Manchester, Manchester, UK.
- Drummond, R. 2012. Aspects of identity in a second language: ING variation in the speech of Polish migrants living in Manchester, UK. *Language Variation and Change*. 24, pp. 107-133
- Drummond, R. 2013. The Manchester Polish STRUT: Dialect Acquisition in a Second Language. *Journal of English Linguistics*, 41(1), pp. 65–93. doi: 10.1177/0075424212449172.

- Duggan, B., Thomas, H. and Jones, E., 2014. *Evaluation of the Welsh-medium Education Strategy - Final Report*. Cardiff: Welsh Government Social Research.
- Durham, M. 2014. *The acquisition of sociolinguistic competence in a lingua franca context*. Multilingual Matters; Bristol, UK.
- Durham, M. and Morris, J. 2016. 'An overview of sociolinguistics in Wales'. In: M, Durham and J, Morris (eds.), *Sociolinguistics in Wales*, pp. 3-28, London: Palgrave Macmillan.
- Durham, M. and Morris, J. 2016. *Sociolinguistics in Wales*. London: Palgrave Macmillan UK.
- Eckert, P. 1988. Adolescent social structure and the spread of linguistic change. *Language in society*, 17(2), pp. 183-207.
- Eckert, P. 2000. *Linguistic variation as a social practice*. Malden, MA: Blackwell Publishers.
- Eckert, P. 2000. *Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Oxford: Blackwell.
- Eckert, P. 2012. Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, 41, pp. 87-100.
- Eckert, P. and Rickford, J. R. 2001. *Style and sociolinguistic variation*. Cambridge: Cambridge University Press.
- Egan, D. 1987. *Coal Society: a history of the South Wales mining valleys 1840-1980*. Llandysul: Gomer
- Erker, D. and Guy, G. 2012. The Role of Lexical Frequency in Syntactic Variability: Variable Subject Personal Pronoun Expression in Spanish. *Language*, 88(3), pp. 526-557.
- Estyn. 2017a. *Inspection report School X*. Available at:
[https://www.estyn.gov.wales/provider/\[REDACTED\]](https://www.estyn.gov.wales/provider/[REDACTED])
- Estyn. 2017b. *Inspection report School Y*. Available at:
[https://www.estyn.gov.wales/provider/\[REDACTED\]](https://www.estyn.gov.wales/provider/[REDACTED])
- Estyn. 2017c. *Careers: The implementation of the careers and world of work framework in secondary schools*. Cardiff: Estyn. [Online]. Available at:
https://www.estyn.gov.wales/sites/default/files/documents/Careers%20-%20Thematic%20survey%20report%20%28002%29_0.pdf [Accessed April 2022]
- Estyn. 2022. Welsh Immersion Education - Strategies and approaches to support 3 to 11- year-old learners. Available at: <https://estyn.gov.wales/improvement-resources/welsh-immersion-education-strategies-and-approaches-to-support-3-to-11-year-old-learners/>

- Evas, J., Mac Giolla Chríost, D., and Williams, C. H. 2014. *A Review of the Work of Mentrau Iaith, Language Action Plans and the Aman Tawe Language Promotion Scheme*. Cardiff: Welsh Government.
- Fehringer, C. and Corrigan, K. 2015. The rise of the going to future in Tyneside English: Evidence for further grammaticalization. *English World-Wide* 36(2), pp. 198–227. doi:10.1075/eww.36.2.03feh
- Fife, J. 1986. Literary vs. colloquial Welsh: problems of definition. *Word*, 37(3), pp. 141-151.
- 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.
- 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*. Cambridge: Cambridge University Press, pp. 235- 267.
- Fortune, T. W., and Tedick, D. J. 2008. *Pathways to Multilingualism : Evolving Perspectives on Immersion Education*. Clevedon: Multilingual Matters.
- Foulkes, P., Gerard D. and Watt, D. 2005. Phonological variation in child-directed speech. *Language*, 81(1), pp. 177–206.
- Frej, M. Y. and Nam, J. 2014. Study on the Sentiment Polarity Types of Collocations for too and very. *Athens Journal of Philology*. 1(1), pp. 23-32
- French, L.M. and Beaulieu, S. 2016. Effects of sociolinguistic awareness on French L2 learners' planned and unplanned oral production of stylistic variation. *Language Awareness*, 25(1-2), pp. 55-71, doi: 10.1080/09658416.2015.1122024.
- Fuchs, R. 2017. Do women (still) use More intensifiers than men? Recent change in the sociolinguistics of intensifiers in British English. *International Journal of Corpus Linguistics*. 22, pp. 345-374. doi: 10.1075/ijcl.22.3.03fuc.
- Fuchs, R., and Gut, U. 2016. 'Register variation in intensifier usage across Asian Englishes' In: H. Pichler, (ed.), *Discourse-Pragmatic Variation and Change: Insights from English: New Methods and Insights*. Cambridge: Cambridge University Press.
- Gardner, M. H. 2023. Mixed-Effects Logistic Regression Analysis. Linguistics Methods Hub: Doing LVC with R. (https://lingmethodshub.github.io/content/R/lvc_r/112_lvc_r.html). doi: 10.5281/zenodo.7160718 (Accessed May 2024)

- Gardner, M., and Tagliamonte, S. A. 2020. The bike, the back, and the boyfriend: Confronting the “definite article conspiracy” in Canadian and British English. *English World-Wide* 41(2), pp. 225–254.
- Gathercole, V.C.M. and Thomas, E.M. 2009. Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism*, 12(2), pp. 213–237. doi: 10.1017/S1366728909004015.
- Geeslin, K. L., Gudmestad, A., Kanwit, M., Linford, B., Yim Long, A., Schmidt, L. and Solon, M. 2018. ‘Sociolinguistic competence and the acquisition of speaking.’ In: *Speaking in a Second Language*. R. Alonso Alonson, (ed). Amsterdam: John Benjamins.
- Geeslin, K.L., and Long, A.Y. 2014. *Sociolinguistics and Second Language Acquisition: Learning to Use Language in Context*. Oxon: Routledge.
- Goalabré, F. 2015. Immersion education and the revitalisation of Breton and Gaelic as community languages. In M. C. Jones (Ed.), *Policy and Planning for Endangered Languages* (pp. 48–66). chapter, Cambridge: Cambridge University Press.
- Gondra, A. 2024. Ideological effects of language revitalisation and standardisation on traditional home language speakers: the case of Basque, *Journal of Multilingual and Multicultural Development*, doi: 10.1080/01434632.2024.2326506
- Gruffydd, I. 2022. *Astudiaeth o amrywio ieithyddol yng nghyd-destun adfywio ieithyddol yng Nghymraeg Caerdydd* [A study of language variation in the context of language revitalisation in Cardiff Welsh]. Unpublished doctoral thesis, Cardiff University.
- Gruffydd, I., Morris, J. and Mayr, R. 2024. The Social Perceptions of Traditional and New Speaker Accents in Welsh. *BAAP Colloquium 2024*, Cardiff University, 25 - 27 March 2024.
- Hansen-Edwards, J. 2011. Deletion of /t, d/ and the Acquisition of Linguistic Variation by Second Language Learners of English. *Language Learning*, 61(4), pp. 1256-1301. doi: 10.1111/j.1467-9922.2011.00672.x
- Hatton, L. 1988. ‘The Development of the Nasal Mutation in the Speech of School-Children’. In: M. J. Ball (ed.) *The use of Welsh: A contribution to sociolinguistics*, pp. 239-257, Clevedon: Multilingual Matters
- Heidemann, K. A. 2014. In the Name of Language: School-Based Language Revitalization, Strategic Solidarities, and State Power in the French Basque Country, *Journal of Language, Identity & Education*, 13(1), pp. 53-69, doi: 10.1080/15348458.2014.864213

- Heine, B. and Kuteva, T. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Hickey, T. M. Lewis, G. and Baker, C. 2014. How deep is your immersion? Policy and practice in Welsh-medium preschools with children from different language backgrounds. *International Journal of Bilingual Education and Bilingualism*, 17(2), pp. 215-234, doi: 10.1080/13670050.2013.866629
- Hickey, T. M., Lewis, G. and Baker, C. 2014. How deep is your immersion? Policy and practice in Welsh-medium preschools with children from different language backgrounds, *International Journal of Bilingual Education and Bilingualism*, 17(2), pp. 215-234, doi: 10.1080/13670050.2013.866629
- Hodges, R., 2009. Welsh Language Use Among Young People in the Rhymney Valley. *Contemporary Wales* 22, pp. 16-35.
- Hothorn, T., Bretz, F. and Westfall, P. 2008. Simultaneous Inference in General Parametric Models. *Biometrical Journal* 50: 346–363. <https://doi.org/10.1002/bimj.200810425>.
- Howley, G. 2015. *The acquisition of Manchester dialect variants by adolescent Roma migrants*. Unpublished doctoral thesis, University of Salford, Salford, UK.
- Hughes, B. 2013. *Patrymau defnydd iaith disgyblion cyfnod allweddol 2 a chyfnod allweddol 3* [Language use patterns of key stage 2 and key stage 3 pupils]. Unpublished doctoral thesis, Bangor University.
- Hymes, D. 1972. ‘On communicative competence’. In: J. B. Pride, and J. Holmes, (eds.) *Sociolinguistics*, pp. 269-293, Harmondsworth: Penguin.
- Ito, R. & Tagliamonte, S. A. 2003. Well weird, right dodgy, very strange, really cool: Layering and recycling in English intensifiers. *Language in Society*, 32(2), pp. 257–79.
- Jenkins, G.H. Parry, G. and Williams, M. A. 1999. ‘More people speak it than ever before’. In: G. Parry and M. A. Williams (eds) *The Welsh Language and the 1891 Census*. University of Wales.
- Johnston, D. 2020. *“Iaith oleulawn”: Geirfa Dafydd ap Gwilym*. Caerdydd: Gwasg Prifysgol Cymru.
- Johnson, D. E. 2009. Getting off the GoldVarb standard: introducing Rbrul for mixed– effects variable rule analysis. *Language and Linguistics Compass* 3 pp. 359-383
- Johnson, R. K. and Swain, M. 1997. *Immersion education: International perspectives*. Cambridge: Cambridge University Press.

- Joint Planning Policy Unit (Gwynedd and Anglesey). 2015. *Topic Paper 10: Welsh Language and Culture*. Bangor. Available at:
<https://www.gwynedd.llyw.cymru/en/Council/Documents---Council/Strategies-and-policies/Environment-and-planning/Planning-policy/Supporting-documents/Supporting-documents-2015/Welsh-Language-and-Culture.pdf> [Accessed 29/07/2024]
- Jones, B. M. 1993. *Ar Lafar ac ar Bapur [In speech and on paper]*. Aberystwyth: Argraffwyr Cambrian.
- Jones, D. G. 1988. 'Literary Welsh'. In: M. J. Ball (ed.) *The Use of Welsh: A Contribution to Sociolinguistics*, pp. 125–172. Clevedon: Multilingual Matters.
- Jones, G. E. 2000. 'The Welsh Language in the Blue Books of 1847'. In: G. H. Jenkins (ed.) *Welsh language and its social domains, 1801-1911*. pp. 431-457, Cardiff : University of Wales Press.
- Jones, H. 2012. *A statistical overview of the Welsh language*. Cardiff: Welsh Language Board.
- Jones, M. C. 1998. *Language obsolescence and revitalization: Linguistic change in two sociolinguistically contrasting Welsh communities*. Oxford: Oxford University Press.
- Jones, S. L. 2019: Perspectives and attitudes towards Welsh-medium study at post-compulsory level among 15–16-year-old students in the South Wales Valleys, *International Journal of Bilingual Education and Bilingualism*, 25(1), pp. 261-271. doi: 10.1080/13670050.2019.1668349
- Jones, S. L., 2017. What Do We Know and Not Know about Choice of Medium of Education in South-East Wales?, *Wales Journal of Education* 19(2), pp. 143-162. doi: <https://doi.org/10.16922/wje.19.2.8>
- Kanwit, M., & Geeslin, K. L. (2020). Sociolinguistic competence and interpreting variable structures in a second language: a study of the copula contrast in native and second-language Spanish. *Studies in Second Language Acquisition*, 42(4), pp. 775–799. doi:10.1017/S0272263119000718
- Kanwit, M., Elias, V., and Clay, R. 2018. Acquiring intensifier variation abroad: Exploring muy and bien in Spain and Mexico. *Foreign Language Annals*, 51, pp. 455–471. doi: <https://doi.org/10.1111/flan.12335>
- Kasstan, J. R. 2020. 'Modelling stylistic variation in threatened and under-documented languages', *Language Ecology*, 4(1), pp. 73-94.
- Kerswill, P. 1996. Children, adolescents, and language change. *Language Variation and Change*, 8(2), pp. 177–202.

- Kerswill, P. and Williams, A. 2000. Creating a new town koine: children and language change in Milton Keynes. *Language in Society* 29, pp. 65-115.
- King, G. 2014. *Basic Welsh: a grammar and workbook*. Oxon: Routledge.
- King, G. 2016. *Modern Welsh: A Comprehensive Grammar*. London: Routledge.
- Kinging, C. 2008. Language learning in study abroad: Case studies of Americans in France. *Modern Language Journal* 92(1), pp. 1-24.
- Klinger, D. and Véronique, D. 2006. La grammaire du français fondamental: interrogations historiques et didactiques. *Documents pour l'histoire du français langue étrangère ou seconde*, 36, pp. 61-75.
- Knight, D., Morris, S., Fitzpatrick, T., Rayson, P., Spasić, I., Thomas, E-M., Lovell, A., Morris, J., Evas, J., Stonelake, M., Arman, L., Davies, J., Ezeani, I., Neale, S., Needs, J., Piao, S., Rees, M., Watkins, G., Williams, L., Muralidaran, V., Tovey-Walsh, B., Anthony, L., Cobb, T., Deuchar, M., Donnelly, K., McCarthy, M. and Scannell, K. (2020). *CorCenCC: Corpws Cenedlaethol Cymraeg Cyfoes – the National Corpus of Contemporary Welsh*. Cardiff University. doi: <http://doi.org/10.17035/d.2020.0119878>
- Konig, E. 2017. 'The comparative basis of intensification'. In: M. Napoli, and M. Ravetto (eds.) *Exploring Intensification. Synchronic, diachronic and crosslinguistic perspectives*. pp. 15-32, Amsterdam: John Benjamins.
- Kovac, C. and Adamson, H. D. 1981. 'Variation theory and first language acquisition'. In: D. Sankoff and H. Cedergren (eds.), *Variation Omnibus*, pp. 403-410, Edmonton, Alberta: Linguistic Research Inc.
- Labov, W. 1964. 'Stages in the acquisition of Standard English'. In: R. Shuy, A. Davis & R. Hogan (eds.), *Social Dialects and Language Learning*, pp. 77-104. Champaign: National Council of Teachers of English.
- Labov, W. 1966. *The Social Stratification of English in New York City*. Washington: Centre for Applied Linguistics.
- Labov, W. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. 1984, 'Intensity'. In: D. Schrifin (ed.), *Meaning, Form and Use in Context: Linguistic Applications*, pp. 43-70, Washington DC: Georgetown University Press.
- Labov, W. 1994. *Principles of Linguistic Change, Volume 1: Internal Factors*. Oxford: Blackwell.
- Labov, W. 2001. 'A dissection of style-shifting'. In: J. R. Rickford and P. Eckert (eds.), *Style and sociolinguistic variation*, pp. 85-105 Cambridge: Cambridge University Press.

- Labov, W. 2007. Transmission and diffusion. *Language* 83(2), pp. 344–387. doi: 10.1353/lan.2007.0082.
- Labov, W. 2010. *Principles of Linguistic Change, Volume 3: Cognitive and Cultural Factors*. Oxford: Blackwell.
- Labov, W. 2013. Preface: The acquisition of sociolinguistic variation. *Linguistics*, 51(2), pp. 247–250. doi: 10.1515/ling-2013-0009.
- Ladegaard, H.J. and Bleses, D. 2003. Gender differences in young children’s speech: The acquisition of sociolinguistic competence. *International Journal of Applied Linguistics* 13(2), pp. 222–233. doi: 10.1111/1473-4192.00045.
- Learn Welsh. 2023a. *Mynediad 1: Units 1-15*. Available at: <https://learnwelsh.cymru/learning/digital-resources/entry/>
- Learn Welsh. 2023b. *Mynediad 2: Units 16-28*. Available at: <https://learnwelsh.cymru/learning/digital-resources/entry/>
- Lewis, D. G. 1995. *Y Llyfr Berfau*. [The Verb Book]. Talybont, Ceredigion: Y Lolfa.
- Lichtenberk, F., Vaid, J. and Chen, H. 2011. On the interpretation of alienable vs. inalienable possession: A psycholinguistic investigation. *Cognitive Linguistics*, 22(4), pp. 659-689.
- Linford, B, Long, A.Y., Solon, M., Whatley, M, Geeslin, K. 2016. Lexical frequency and subject expression in native and non-native Spanish. In: S. Sessarego, F. Tejedo-Herrero (eds.) *Spanish Language and Sociolinguistic Analysis*, pp. 197-215, Amsterdam: John Benjamins.
- Linford, B., and Shin, N. L. 2013. ‘Lexical frequency effects on L2 Spanish subject pronoun expression’. In : J. Cabreilli Amaro, G. Lord, A. de Prada Perez and J.F. Aaron (eds.) *Selected proceedings of the 16th Hispanic Linguistics Symposium*, pp. 175-189, Somerville, MA: Cascadilla Proceedings Project.
- Lippi-Green, R. 1997. *English with an Accent: Language, Ideology, and Discrimination in the United States*. London and New York: Routledge.
- Loeffler, M. 2008. ‘English in Wales’. In: H, Momma, and M, Matto, (eds.) *A Companion to the History of the English Language*, pp. 350-357, Oxford: Blackwell.
- Lorenz, G. 2002. ‘Really Worthwhile or Not Really Significant? A Corpus-based Approach to the Delexicalization and Grammaticalization of Intensifiers in Modern English’. In: I. Wischer and G. Diewald (eds.) *New Reflections on Grammaticalization*, pp. 143-161, Amsterdam: Benjamins.
- Lyster, R. 1993. *The effect of functional-analytic teaching on aspects of sociolinguistic*

- competence: A study in French immersion classrooms at the Grade 8 level*. Unpublished doctoral dissertation, University of Toronto, Toronto, Ontario, Canada.
- Lyster, R. 1994. The effect of functional-analytic teaching on aspects of sociolinguistic competence. *Applied Linguistics*, 15, pp. 263-287.
- May, S. 2000. Accommodating and resisting minority language policy: The case of Wales. *International Journal of Bilingual Education and Bilingualism*, 3(2), 101–128.
- Mayr, R., Morris, J., Mennen, I. and Williams, D. 2017. Disentangling the effects of long-term language contact and individual bilingualism: The case of monophthongs in Welsh and English. *International Journal of Bilingualism* 21(3), pp. 245–267. doi: 10.1177/1367006915614921.
- Meelen, M. and Nurmio, S. 2020. Adjectival Agreement in Middle and Early Modern Welsh Native and Translated Prose. *Journal of Celtic Linguistics*, 21(1), pp. 1-28
- Mennen, I., Kelly, N., Mayr, R. and Morris, J. 2020. The Effects of Home Language and Bilingualism on the Realization of Lexical Stress in Welsh and Welsh English. *Frontiers in Psychology*, 10, pp. 1–17. doi: 10.3389/fpsyg.2019.03038.
- Meyerhoff, M. 2011. *Introducing Sociolinguistics*. Oxon: Routledge.
- Meyerhoff, M. and Schlee, E. 2012. Variation, contact and social indexicality in the acquisition of (ing) by teenage migrants. *Journal of Sociolinguistics*, 16(3), pp. 398–416. doi: 10.1111/j.1467-9841.2012.00535.x.
- Milroy, J. 1995. William Labov, Principles of linguistic change. Volume I: Internal factors (Language in Society 20). Oxford: Blackwell, 1994, pp. xix + 641. *Journal of Linguistics*, 31(2), pp. 435–439. doi:10.1017/S0022226700015693
- Milroy, L. 1987. *Language and Social Networks*. Oxford: Blackwell.
- Milroy, L. 2007. ‘Off the shelf or under the counter? On the social dynamics of sound changes’. In: C. M. Cain and G. Russom (eds.) *Managing Chaos: Strategies for Identifying Change in English*, pp. 149–172, New York: Mouton de Gruyter.
- Milroy, L. and Gordon, M. 2003. *Sociolinguistics: Method and Interpretation*. Malden, MA: Blackwell.
- Morris, J. 2013. *Sociolinguistic variation and regional minority language bilingualism : an investigation of Welsh-English bilinguals in North Wales*. Unpublished doctoral thesis, The University of Manchester, Manchester, UK.
- Morris, J. 2014. The influence of social factors on minority language engagement amongst young people: An investigation of Welsh-English bilinguals in North Wales. *International*

- Journal of the Sociology of Language*, 230, pp. 65–89. doi: 10.1515/ijsl-2014-0027.
- Morris, J. 2014. *Trawsgrifio data ieithyddol Cymraeg* [Transcribing Welsh linguistic data]. Unpublished Manuscript.
- Morris, J. 2017. Sociophonetic variation in a long-term language contact situation: /l/-darkening in Welsh-English bilingual speech, *Journal of Sociolinguistics*, 21(2), pp. 183–207. doi: 10.1111/josl.12231.
- Morris, J. 2021. Social Influences on Phonological Transfer: /r/ Variation in the Repertoire of Welsh-English Bilinguals. *Languages*, 6(2), 97. <https://doi.org/10.3390/languages6020097>
- Morris, J. 2022. Fundamental frequency range in the bilingual repertoire of traditional and new Welsh speakers. *International Journal of Bilingualism*, 26(5), pp. 564–583. doi: 10.1177/136700692211110.
- Morris, J. and Hejrná, M. 2019. Pre-aspiration in Bethesda Welsh: A sociophonetic analysis. *Journal of the International Phonetic Association*, pp. 1–25. doi: 10.1017/S0025100318000221.
- Morris, J. and Young, K. 2022. Iaith – beth mae cyfnewid cod yn ei gyfleu? [Language – what does codeswitching convey?] [Online]. Golwg 360: Available at: <https://golwg.360.cymru/gwerddon/2108980-iaith-beth-cyfnewid-gyfleu> [Access 29/07/2024]
- Mougeon, R. and Rehner, K. 2019. Patterns of sociolinguistic variation in teacher classroom speech, *Journal of Sociolinguistics*, 23(2), pp. 163–185. doi: 10.1111/josl.12336.
- Mougeon, R., and Rehner, K. (2001). Acquisition of Sociolinguistic Variants by French Immersion Students: The Case of Restrictive Expressions, and More. *The Modern Language Journal*, 85(3), pp. 398–415. <http://www.jstor.org/stable/1193108>.
- Mougeon, R., Nadasdi, T. and Rehner, K. 2010. *The Sociolinguistic Competence of Immersion Students*. Bristol: Multilingual Matters.
- Murray, H. 2003. Swiss English teachers and Euro-English: Attitudes to a non-native variety. *Bulletin Suisse de Linguistique Appliquée*. 77, pp. 147–165.
- Murtagh, L. 2007. Out-of-school Use of Irish, Motivation and Proficiency in Immersion and Subject-only Post-primary Programmes, *International Journal of Bilingual Education and Bilingualism*, 10(4), pp. 428–453, DOI: 10.2167/beb453.0
- Musk, N. 2010. Code-switching and code-mixing in Welsh bilinguals' talk: confirming or refuting the maintenance of language boundaries? *Language, Culture and Curriculum*, 23(3), pp. 179–197, doi: 10.1080/07908318.2010.515993

- Nadasdi, T., Mougeon, R. and Rehner, K., 2003. Emploi du 'futur' dans le français parlé des élèves d'immersion française. *Journal of French Language Studies*, 13(2), pp. 195-219.
- Nance, C. 2013. *Phonetic variation, sound change, and identity in Scottish Gaelic*. Unpublished doctoral thesis, University of Glasgow, Glasgow, UK.
- Nance, C. 2020. Bilingual language exposure and the peer group: Acquiring phonetics and phonology in Gaelic Medium Education. *International Journal of Bilingualism*, 24(2) pp. 360–375. doi: 10.1177/1367006919826872.
- National Careers Service. 2020. The STAR method. Available at: <https://nationalcareers.service.gov.uk/careers-advice/interview-advice/the-star-method> [Accessed 20/07/2024].
- Nevalainen, T., and Palander-Collin, M. 2011. 'Grammaticalization and sociolinguistics'. In: H. Narrog, & B. Heine (Eds.), *The Oxford Handbook of Grammaticalization*, pp. 118-129, Oxford: Oxford University Press.
- Newcombe, L.P. 2007. *Social Context and Fluency in L2 Learners: The Case of Wales*. Multilingual Matters.
- Ní Dhiorbháin, A., Nic Aindriú, S., Connaughton-Crean, L., & Ó Duibhir, P. 2023. It's more the invisible benefits – multilingual parents' experiences of immersion education and their reasons for choosing immersion. *Language and Education*, 38(5), pp. 746–765. <https://doi.org/10.1080/09500782.2023.2238680>
- Nichols, J. 1988. On alienable and inalienable possession. *Proceedings From the Haas Festival Conference On Native American Linguistics*, Berlin: De Gruyter Mouton.
- Ó Sé, D. 2004. The “After” Perfect and Related Constructions in Gaelic Dialects. *Ériu*, 54, pp. 179–248.
- O'Shannessy, C. 2013. 'Researching Children's Acquisition of Sociolinguistic Competence'. In: J. Holmes, K. Hazen (eds.) *Research methods in sociolinguistics : A practical guide*. Oxford: Wiley.
- O'Hanlon, F. 2015. Choice of Scottish Gaelic-medium and Welsh-medium education at the primary and secondary school stages: parent and pupil perspectives, *International Journal of Bilingual Education and Bilingualism*, 18(2), pp. 242-259
- Owen, J. A. 2018. *Factors influencing Welsh medium school pupils' social use of Welsh*. Unpublished doctoral thesis, Bangor University.

- Paradis, C. 2000. 'It's well weird. Degree modifiers of adjectives revisited: the nineties'. In: J. M. Kirk (ed.) *Corpora galore: Analyses and techniques in describing English*, pp. 147-160 (Language and computers; Vol. 30). Rodopi.
- Parry, G, and Williams, M. A. 1999. *The Welsh Language and the 1891 Census*. Cardiff: University of Wales Press.
- Peterson, E. 2017. The nativization of pragmatic borrowings in remote language contact situations. *Journal of pragmatics*, 113, pp.116-126
- Podesva, R..J. 2007. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*, 11(4), pp. 478–504. doi: 10.1111/j.1467-9841.2007.00334.x.
- Poplack, S. 2000. *The English History of African American English*. Oxford: Blackwell.
- Poplack, S. and Sankoff, D. 1984. Borrowing: The synchrony of integration. *Linguistics*, 22, pp. 99-135.
- Poplack, S. and St-Amand, A. 2007. A real-time window on 19th-century vernacular French: The Récits du français Québécois d'autrefois. *Language in Society* 36 (5), pp. 707-734.
- Poplack, S. and Tagliamonte, S. A. 2001. *African American English in the Diaspora: Tense and Aspect*. Malden, MA: Blackwell.
- Poplack, S. and Turpin, D. 1999. Does the futur have a future in (Canadian) French? *Probus: International Journal of Latin and Romance Linguistics*, 11(1) pp.133–164
- Prys, M. 2016. *Style in the vernacular and on the radio : code-switching and mutation as stylistic and social markers in Welsh*. Unpublished doctoral thesis, Bangor University, Bangor, UK.
- Purvis, K. (2020 July 28). Recruiters inundated as coronavirus takes toll on UK labour market. *The Guardian*. Available at: <https://www.theguardian.com/society/2020/jul/28/recruiters-inundated-coronavirus-toll-job-market-employment>. [Accessed 20/07/2024].
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Redknap, Catrin. 2006. 'Welsh-medium and bilingual education and training: steps towards a holistic strategy', In: C. Redknap,, W. Gwyn S. R. Lewis and J. L. Williams, *Welsh-medium and Bilingual Education*, pp. 1–20, Bangor: School of Education, University of Wales.
- Regan, V., Howard, M. and Lemée, I. 2009. *The Acquisition of Sociolinguistic Competence in a Study Abroad Context*. Bristol: Multilingual Matters.

- Rehner, K. 2011. The sociolinguistic competence of former immersion students at the post-secondary level: The case of lexical variation. *International Journal of Bilingual Education and Bilingualism*, 14(3), pp. 243–259. doi: 10.1080/13670051003787566.
- Rehner, K., & Mougeon, R. 1999. Variation in the spoken French of immersion students: To ne or not to ne, that is the sociolinguistic question. *The Canadian Modern Language Review*, 56, pp. 124–154.
- Rehner, K., Mougeon, R. and Nadasdi, T. 2003. The learning of sociostylistic variation by advanced French as a second language learners: The case of nous versus on in immersion French. *Studies in Second Language Acquisition* 25(1), pp. 127-156.
- Reichelt, S and Durham, M 2017. Adjective Intensification as a Means of Characterization: Portraying In-Group Membership and Britishness in Buffy the Vampire Slayer. *Journal of English Linguistics*, 45(1), pp. 60-87. doi: 10.1177/0075424216669747
- 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), pp. 633–656.
- Rickford, J. and Price, M. 2013. Girlz II women: Age-grading, language change and stylistic variation, *Journal of Sociolinguistics*, 17(2), pp. 143–179. doi:10.1111/josl.12017.
- 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, pp. 93–116. doi: 10.1515/IJSL.2009.007.
- 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. 135–141, Oslo: Novus Press.
- Roberts, A. E. 1988. 'Age-related variation in the Welsh dialect of Pwllheli'. In: M. J. Ball (ed.) *The use of Welsh: A contribution to sociolinguistics*, pp. 104-124. Clevedon: Multilingual Matters.
- Roberts, J. 2005. 'Acquisition of Sociolinguistic Variation'. In: M. J. Ball (ed.) *Clinical Sociolinguistics*, pp. 151-164, Oxford: Blackwell.

- Rodríguez-Ordóñez I., Sainmaza-Lecanda L. 2018. Bilingualism effects in Basque subject pronoun expression: Evidence from L2 Basque. *Linguistic Approaches to Bilingualism*, 8(5), pp. 523–560.
- Rodríguez-Ordóñez, I. 2022. The role of frequency in the acquisition of structured variation: The case of Basque ergativity. *International Journal of Bilingualism*, 26(5), 656-672. <https://doi.org/10.1177/13670069221110401>
- Romaine, S. 1984. *The Language of children and adolescents: The Acquisition of Communicative Competence*. Oxford: Blackwell.
- Ronan, P. 2012. ‘Development of periphrastic tense and aspect constructions in Irish and Welsh’. In: K. Davidse, T. Breban, L. Brems, and T. Mortelmans, (eds.) *Grammaticalization and Language Change: New Reflections*, pp. 227-248. Amsterdam: John Benjamins.
- Rottet, K. J. and Morris, S. 2018. *Comparative Stylistics of Welsh and English : Arddulleg Y Gymraeg*. Cardiff: University of Wales Press.
- Ryan, S.D. 2018. *Language, migration and identity at school : A sociolinguistic study with Polish adolescents in Glasgow*. Unpublished doctoral thesis, University of Glasgow, Glasgow, UK.
- Sankoff, G. and Vincent, D. 1977. L’emploi productif du ne dans le français parlé à Montreal. *Le Francais Moderne* 45, pp. 243-256.
- Sankoff, G., and Blondeau, H. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83, pp. 560–88. doi: 10.1353/lan.2007.0106.
- Sansò, A and Fiorentini, I (2017) Intensifiers between grammar and pragmatics. A lesson from a language contact situation. In: M. Napoli and M. Ravetto (eds.), *Exploring Intensification Synchronic, diachronic and cross-linguistic perspectives*, pp. 173-192. Amsterdam: John Benjamins.
- Savignon, S. J. 1997. *Communicative competence: Theory and classroom practice* (2nd ed.). New York: McGraw-Hill.
- Saville-Troike, M. 2006. *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.
- Sayers, D. 2009. Reversing Babel: Declining linguistic diversity and the flawed attempts to protect it. Unpublished doctoral thesis, University of Essex.
- Schilling-Estes, N. 2002. On the nature of isolated and post-isolated dialects: Innovation, variation and differentiation. *Journal of Sociolinguistics*, 6(1), pp. 64–85. doi: 10.1111/1467-9481.00177.

- 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.
- Schleef, E. 2013. Migrant teenagers' acquisition of sociolinguistic variation: The variables (ing) and (t). In: P. Auer, J. C. Reina and G. Kaufman (eds.) *Language Variation - European Perspectives IV: Selected papers from the Sixth International Conference on Language Variation in Europe (ICLaVE 6), Freiburg, June 2011*.
- Schleef, E., Meyerhoff, M. and Clark, L. 2011. Teenagers' acquisition of variation: A comparison of locally-born and migrant teens' realisation of English (ing) in Edinburgh and London. *English World-Wide: A Journal of Varieties of English* 32(2), pp. 206–236. doi: 10.1075/eww.32.2.04sch.
- Senedd Research (2022) *Coronavirus timeline: Welsh and UK governments' response*. Available at: <https://research.senedd.wales/research-articles/coronavirus-timeline-welsh-and-uk-governments-response/> [Accessed 20/07/2024].
- Sharma, D. 2011. Style repertoire and social change in British Asian English. *Journal of Sociolinguistics*, 15(4), pp. 464–492. doi: 10.1111/j.1467-9841.2011.00503.x.
- Smith, J. and Durham, M. 2019. *Sociolinguistic Variation in Children's Language: acquiring Community Norms*. Cambridge: Cambridge University Press.
- Smith, J., Durham, M. and Fortune, L. 2007. 'Mam, ma troosers is fa'in doon!': community, caregiver and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change*, 19(1), pp. 63-99.
- Smith, J., Durham, M. and Richards, H. 2013. The social and linguistic in the acquisition of sociolinguistic norms: Caregivers, children, and variation. *Linguistics* 51(2), pp. 285–324. doi: 10.1515/ling-2013-0012.
- Smith, R. 2000. 'Broadcasting and the Welsh Language'. In: G. Jenkins (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.
- Stardy, R 2019. A comparative analysis of the intensifiers quite, rather, and pretty used by Americans and British people: a corpus-based study. *Journal of English Language and Culture*, 9(2), pp. 160-171
- Starr, R. L. 2017. *Sociolinguistic variation and acquisition in two-way language immersion: Negotiating the standard*. Bristol: Multilingual Matters.

- StatsWales 2021. *Pupil Level Annual School Census 'Ability to Speak Welsh' & 'Speaking Welsh at Home', as assessed by parents, of pupils aged 11-15 in secondary schools by local authority* (<https://statswales.gov.wales/Catalogue/Education-and-Skills/Schools-and-Teachers/Schools-Census/Pupil-Level-Annual-School-Census/Welsh-Language/pupils11to15secondary-speakwelsh-homewelsh-by-localauthority>) [Accessed: 15/06/2021].
- StatsWales, 2023. *Annual Population Survey - Ability to speak Welsh by local authority and year*. Available at: <https://statswales.gov.wales/Catalogue/Welsh-Language/Annual-Population-Survey-Welsh-Language> [Accessed: 14/03/2024].
- StatsWales. 2021. *Number of pupils in primary, middle and secondary school classes by local authority and Welsh category* (<https://statswales.gov.wales/Catalogue/Education-and-Skills/Schools-and-Teachers/Schools-Census/Pupil-Level-Annual-School-Census/Welsh-Language/pupilswelshclasses-by-localauthority-welshcategory>) [Accessed: 16/05/2021].
- Stratton, J. 2021. 'That's proper cool': The emerging intensifier proper in British English. *English Today*, 37(4), pp. 206-213. doi:10.1017/S0266078420000036
- Stratton, J. and Sundquist, J. 2022. A Variationist Sociolinguistic Analysis of Intensifiers in Oslo Norwegian. *Journal of Germanic Linguistics*. 34. Pp. 385-419. doi: 10.1017/S1470542722000022.
- Su, Y. 2016. Corpus-based comparative study of intensifiers: quite, pretty, rather and fairly, *Journal of World Languages*, 3(3), pp. 224-236, doi: 10.1080/21698252.2017.1308306
- Sundell, L.-G. 1991. *Le temps futur en français moderne. Uppsala: Textgruppen i Uppsala AB.*
- Sundqvist, P. 2009. *Extramural English matters: Out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary*. Unpublished doctoral thesis, Karlstad University.
- Swain, M., and Lapkin, S. 1990. 'Aspects of the sociolinguistic performance of early and later French immersion students'. In: R. C. Scarcella, E. S. Anderson & S. D. Krashen (eds.), *Developing communicative competence in a second language*, pp. 41-54. New York: Newbury House.
- Tagliamonte, S. A and Roberts, C. 2005. So weird; so cool; so innovative: The use of intensifiers in the television series Friends. *American Speech*, 80(3), pp. 280-300. <https://doi.org/10.1215/00031283-80-3-280>.
- Tagliamonte, S. A. 2008. So different and pretty cool! Recycling intensifiers in Toronto, Canada. *English Language and Linguistics*, 12(2), pp. 361–394. doi: 10.1017/S1360674308002669

- Tagliamonte, S. A. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Oxford: Wiley–Blackwell.
- Tagliamonte, S. A. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Oxford: Wiley-Blackwell.
- Tagliamonte, S. A. 2013. ‘Comparative Sociolinguistics’ In: J. K. Chambers and N. Schilling-Estes (eds.), *The handbook of language variation and change*, pp. 376-401. Malden: Blackwell.
- Tagliamonte, S. A. 2013. ‘Comparative Sociolinguistics’. In: J. K. Chambers, and N. Schilling (eds.), *The Handbook of Language Variation and Change*, pp. 128-156, Oxford: Wiley.
- Tagliamonte, S. A. 2013. *Roots of English. Exploring the History of Dialects*. Cambridge: Cambridge University Press.
- Tagliamonte, S. A. 2016. *Teen Talk: The Language of Adolescents*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139583800
- Tagliamonte, S. A. and Baayen, R. H. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2), pp. 135–178.
- Tagliamonte, S. A. and D’Arcy, A. 2007. Frequency and Variation in the Community Grammar: Tracking a New Change Through the Generations. *Language Variation and Change*, 19, pp. 199–217.
- Tagliamonte, S.A., Durham, M. and Smith, J., 2014. Grammaticalization at an early stage: future be going to in conservative British dialects. *English Language and Linguistics*, 18(1), pp. 75-108.
- Tagliamonte, S. & Roberts, C., 2005. *So weird; so cool; so innovative: The use of intensifiers in the television series Friends*. *American Speech*, 80(3), pp.280–300.
<https://doi.org/10.1215/00031283-80-3-280>.
- Tarone, E. 2007. Sociolinguistic approaches to second language acquisition research - 1997-2007. *The Modern Language Journal* 91, pp. 837-848.
- Terry, K., M., K. 2017. Contact, Context, and Collocation. *Studies in Second Language Acquisition* 39(3), pp. 553–578. doi: 10.1017/S0272263116000061.
- Thomas, A. R, 1973. *The Linguistic Geography of Wales: A Contribution to Welsh Dialectology*. Cardiff: University of Wales Press.
- Thomas, A. R. 1987. A Spoken Standard for Welsh: Description and Pedagogy. *International Journal of the Sociology of Language* 66, pp. 99–113.

- Thomas, A. R. 1988b. A spoken standard for Welsh: Description and pedagogy. Cardiff: Welsh Joint Education Committee.
- Thomas, B. 1980. Cymrêg, Cymraeg: Cyweiriau iaith siaradwraig o Ddyffryn Afan, BBCS XX, pp. 579-92.
- Thomas, B. 1984. Linguistic and non-linguistic boundaries in north-east Wales. Welsh Phonology: Selected Readings, 1984, pp. 189-207.
- Thomas, B. & Thomas, P. W., 1989. *Cymraeg, Cymrâg, Cymrêg - Cyflwyno'r Tafodieithoedd*. Paperback ed. Cardiff: University of Wales Press.
- Thomas, E. and Gathercole, V. 2007. Children's productive command of grammatical gender and mutation in Welsh: An alternative to rule-based learning. *First Language*, 27(3), pp. 251-278
- Thomas, E. M. and Mayr, R. 2010. 'Children's Acquisition of Welsh in a Bilingual Setting: A Psycholinguistic Perspective' In: D. Morris (ed.) *Welsh in the Twenty-First Century*. Cardiff: University of Wales Press
- Thomas, H. S. 2013. 'A Dynamic Profile: Perceptions and Facts'. In: H. S. Thomas and C. H Williams (eds.) *Parents, Personalities and Power*, pp. 25-63, Cardiff University of Wales Press.
- Thomas, P. W. 1996. Gramadeg y Gymraeg. Cardiff: Gwasg Prifysgol Cymru.
- Torres-Cacoulllos, R. and Walker, J. A. 2009. The Present of the English Future: Grammatical Variation and Collocations in Discourse. *Language* 85, pp. 321–354. doi: 10.1353/lan.0.0110
- Tristram, A. 2020. Variation and change in future temporal reference with avoir and être, *Journal of French Language Studies*, 31, pp. 25–49.
- Trudgill, P. 1974. *The social differentiation of English in Norwich*. London: Cambridge University Press.
- Trudgill, P. 2002. *Sociolinguistic Variation and Change*. Edinburgh: Edinburgh University Press.
- Ure, J. N., and Ellis, J. O. 1977. Register in descriptive linguistics and linguistic sociology. *Issues in Sociolinguistics*, 197–244. <https://doi.org/10.1515/9783110806687.197>.
- Von Compennolle, R.A., Gomez-Laich, M.P. and Weber, A. 2016. Teaching L2 Spanish Sociopragmatics Through Concepts: A Classroom-Based Study. *Modern Language Journal* 100(1), pp. 341–361. doi: 10.1111/modl.12318.
- Wagner S.E., Sankoff, G. 2011. Age grading in the Montréal French inflected future. *Language Variation and Change*. 23(3) pp. 275-313. doi:10.1017/S0954394511000111

- Watkins, T. A. 1977. The Welsh personal pronoun. *Word* 28(1-2), pp. 146-165.
- Webb-Davies, P. a Shank, C. 2020, Defnydd hanesyddol a chyfoes o ‘mynd i’ yn y Gymraeg: Astudiaeth o ramadegoli fel newid ieithyddol [Historical and contemporary use of ‘mynd i’ in Welsh: A study of grammaticalization as language change], *Gwerddon*, 30, pp. 23-39.
- Welsh Audit Office. 2019. The well-being of young people: skills and employability. Available at: https://www.wao.gov.uk/sites/default/files-old/download_documents/skills-and-employability-english.pdf [Accessed 20/07/2024].
- Welsh Government. 2007. Defining schools according to Welsh medium provision https://dera.ioe.ac.uk/7302/2/defining-schools-welsh-medium2_Redacted.pdf [accessed 25/08/21]
- Welsh Government. 2017. Welsh Language Transmission and Use in Families: Research into conditions influencing Welsh language transmission and use in families. Cardiff: Welsh Government.
- Welsh Government. 2019. *Schools’ Census Results: as at January 2019*. Available at: <https://www.gov.wales/sites/default/files/statistics-and-research/2019-07/school-census-results-2019-764.pdf> [Accessed: 29/07/2024]
- Welsh Government. 2020. The Independent Review Report to re-define schools according to their Welsh-medium education provision <https://gov.wales/sites/default/files/publications/2020-12/the-independent-review-report-to-re-define-schools-according-to-their-welsh-medium-education-provision.pdf>
- Welsh Government. 2022a. *Welsh Language in Wales (Census 2021)*. Available at: <https://www.gov.wales/welsh-language-wales-census-2021> [Accessed: 27/06/2023].
- Welsh Government. 2022b. Curriculum for Wales. <https://hwb.gov.wales/curriculum-for-wales>.
- Welsh Government. 2023a. Data am y Gymraeg yn ol awdurdod lleol [Data about Welsh according to local authority]. Available at: <https://www.llyw.cymru/data-am-y-gymraeg-yn-ol-awdurdod-lleol> [Accessed 20/07/2024].
- Welsh Government. 2023b. *Welsh language composition of households in Wales (Census 2021)*. Available at [https://www.gov.wales/welsh-language-composition-households-wales-census-2021.html#:~:text=For%20one%20parent%20households%20where,in%20Newport%20\(at%2022.2%25\).](https://www.gov.wales/welsh-language-composition-households-wales-census-2021.html#:~:text=For%20one%20parent%20households%20where,in%20Newport%20(at%2022.2%25).) [Accessed 20/07/2024].
- Williams, E. 2009. Language attitudes and identity in a North Wales town: “something different about Caernarfon”? *International Journal of the Sociology of Language* (195), pp. 63-91.

- Willis, P. 1988. Is the Welsh verbal noun a verb or a noun?, *Word*, 39(3), pp. 201-224, doi: 10.1080/00437956.1988.11435790
- Willis, D. 2007. Specifier-to-head reanalyses in the complementizer domain: Evidence from Welsh. *Transactions of the Philological Society* 105. 432–80.
- Willis, D. (2016). *Cyfieithu iaith y caethweision yn Uncle Tom's Cabin a darluniadau o siaradwyr ail iaith mewn llenyddiaeth Gymraeg* ['Translating the language of the slaves in Uncle Tom's cabin and representations of second-language speakers in Welsh literature']. <https://doi.org/10.17863/CAM.5157>.
- Willis, D. 2019. Dialect syntax as a testbed for models of innovation and change: Modals and negative concord in the *Syntactic Atlas of Welsh Dialects*. *Glossa* 4(1). 31. 1–30.
- Willis, D. 2014. '[Newid ac amrywiaeth mewn Cymraeg cyfoes](#) ['Variation and change in contemporary Welsh']'. In D. Prys (ed.), [Ysgrifau ar ieithyddiaeth a geiriaduraeth Gymraeg](#), pp.84-102, Carmarthen: Coleg Cymraeg Cenedlaethol.
- Wolfram, W, and Schilling, N. 2015. *American English: Dialects and Variation*, (3rd ed.) Oxford: Wiley.
- Young, K. 2019. *Perceptions of socio-stylistic variation among teachers of Welsh*. Unpublished MA dissertation, Cardiff University, Cardiff, UK.
- Young, K., Durham, M., & Morris, J. 2024. Possessive pronouns in Welsh: Stylistic variation and the acquisition of sociolinguistic competence. *Language Variation and Change*, 36(1), pp. 25–48. doi:10.1017/S0954394523000273
- Yu, M.-C. 2008. Teaching and Learning Sociolinguistic Skills in University EFL Classes in Taiwan. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect*, 42(1), pp. 31-53.

11 Appendices

11.1 Ethics approval



School of English,
Communication and Philosophy
Ysgol Saesneg, Cyfathrebu
ac Athroniaeth

Cardiff University

John Percival Building, Colum Drive
Cardiff, CF10 3EU, Wales, UK
Tel +44(0)29 2087 6049
Email encap@cardiff.ac.uk
www.cardiff.ac.uk/english-communication-philosophy

Prifysgol Caerdydd

Adeilad John Percival, Rhodfa Colum
Caerdydd, CF10 3EU, Cymru, DU
Ffôn +44(0)29 2087 6049
E-bost encap@caerdydd.ac.uk
www.caerdydd.ac.uk/saesneg-cyfathrebu-athroniaeth

8th January 2021

Dear Ms Katharine Young,

Research project title: The Acquisition of Sociolinguistic Competence in a Welsh Immersion Context

SREC reference: ENCAP/Young/23-07-2020

The ENCAP Research Ethics and Unfair Practice Committee reviewed electronically your application to introduce a number of changes to the data-gathering procedures for the aforementioned project (originally approved on the 23rd of July 2020).

Ethical Opinion

The Committee gave a favourable ethical opinion of the above application on the basis described in the application form, protocol and supporting documentation.

Additional approvals

This letter provides an ethical opinion only. You must not start your research project until all appropriate approvals are in place.

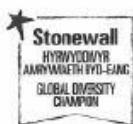
Amendments

Any substantial amendments to documents previously reviewed by the Committee must be submitted to the Committee via Dr Sara M. Pons-Sanz for consideration and cannot be implemented until the Committee has confirmed it is satisfied with the proposed amendments.

You are permitted to implement non-substantial amendments to the documents previously reviewed by the Committee but you must provide a copy of any updated documents to the Committee via Dr Sara M. Pons-Sanz for its records.

Monitoring requirements

The Committee must be informed of any unexpected ethical issues or unexpected adverse events that arise during the research project. Researchers are responsible for ensuring that they adhere to the procedures set out in approved applications. The Committee must be informed when your research project has ended. This notification should be made to ENCAP's Research Office (encap-res@cardiff.ac.uk) within three months of research project completion. You should confirm that data collection has ended and submit a final report of any amendments to the research procedure.



Registered Charity, no. 1136855
Eusen Gofrestredig, rhif 1136855

Documents reviewed by Committee

The documents reviewed by the Committee were:

Document	Date
Research Ethics Approval Application Form	08/01/2021
Participant, Gatekeepers and Parental Information Sheets	08/01/2021
Participant and Gatekeepers Consent Forms	08/01/2021
Participant recruitment message	08/01/2021
Data collection tools	08/01/2021

Complaints/Appeals

If you are dissatisfied with the decision made by the Committee, please contact Dr Sara M. Pons-Sanz in the first instance to discuss your complaint. If this discussion does not resolve the issue, you are entitled to refer the matter to the Head of School for further consideration. The Head of School may refer the matter to the University Research Integrity and Ethics Committee (URIEC), where this is appropriate. Please be advised that URIEC will not normally interfere with a decision of the Committee and is concerned only with the general principles of natural justice, reasonableness and fairness of the decision.

Please use the Committee reference number on all future correspondence.

The Committee reminds you that it is your responsibility to conduct your research project to the highest ethical standards and to keep all ethical issues arising from your research project under regular review.

You are expected to comply with Cardiff University's policies, procedures and guidance at all times, including, but not limited to, its Policy on the Ethical Conduct of Research involving Human Participants, Human Material or Human Data and our Research Integrity and Governance Code of Practice.

Yours sincerely,

11.2 Email to approach schools

FAO Head of School and Head of 6th form:

Dear ***,

I am writing to invite you to engage your students in some careers preparation workshops, and at the same time contribute to some important new research into the use of the Welsh language.

I am a PhD student at Cardiff University, working on a project funded by Welsh Government and the ESRC on the acquisition of sociolinguistic competence in Welsh immersion schools; that is, how pupils vary the way they talk in accordance with a particular register (formal/informal) or situation. The work is important in understanding the type of Welsh used from day-to-day by young people, following Welsh Government's targets of creating a million Welsh speakers by 2050.

The aim of the work is to look at 6th form students' language variation in two Welsh-medium schools in different areas. I believe *** (name of school) would be an appropriate and interesting site for this exciting new research considering the prevalence of the Welsh language in the local authority, and also the mix of different home language backgrounds of your pupils, and would therefore like to invite your school to participate.

Full information about the project is in the 'Gatekeeper information sheet' document attached to this email. I would be grateful if you could read this before you decide whether or not you would like to put forward your school to participate. Due to the time-constrained nature of this research, it would be ideal if you could confirm whether you wish to participate by *** (date 2 weeks from sending). Currently, the aim (which is subject to change depending on your availability) is to collect data between February and May 2021, but the exact period can be discussed with you if your school chooses to participate.

If you have any questions, you are welcome to contact me through the details given on the attached document. If you decide to take part in the project, and after you have confirmed your interest, I will contact you to discuss further details about the project and your participation in it.

Kind regards,

Katharine Young



11.3 Information letter for gatekeepers

THE ACQUISITION OF SOCIOLINGUISTIC COMPETENCE IN A WELSH IMMERSION CONTEXT

Have you noticed how you change the way you speak depending on who you're talking to, or where you are, or what you're talking about?

In my research field, we call this sociolinguistic competence; that is, the ability of a speaker to vary their language (not always consciously) between more formal contexts and more informal contexts.

The current study examines the types of variation which occur in different speech contexts, some more formal and some more casual. It will look at the extent to which social factors, such as home language and use of Welsh in the community affect sociolinguistic competence in Welsh in children attending Welsh-medium schools. Audio recordings will be made of naturally-occurring speech in different registers to elicit varying styles from 20 participants in a sixth form where Welsh is a dominant language in the community, and 20 participants in a sixth form where Welsh is not a dominant language in the community. Questionnaires will also be used to determine the attitudes towards the Welsh language whilst also shedding light on the background of the speaker, the speakers' engagement with Welsh in the school and the wider community.

Your and your students' time is precious and so I want to ensure you have a clear idea about what taking part entails. I'd particularly like to draw your attention to sections 3 and 4, which focus on the practicalities. However, all the details covered below will contribute to your being able to make a fully informed decision. Please take time to read the following information carefully and discuss it with any other relevant gatekeepers (teachers, headteachers and other school staff), as necessary. If you and the students under your care agree to participate, they will be asked to sign a consent form.

Thank you for reading this.

1. What is the purpose of this research project?

The purpose of this project is to understand to what extent home language and community language affect the way young Welsh speakers vary their language according to the situation they're talking in (i.e. more formal language for more formal situations).

2. Why has my school been invited to take part?

Your school has been invited because enrolled students are known to use Welsh in a variety of different situations at school, some more formal and some more casual.

3. Does my school have to take part?

No, your school's participation in this research project is entirely voluntary and it is up to you to decide whether or not to take part. If you decide to take part, you will be asked to select a cohort of 25-30 students (demographic to be discussed with the researcher). We will then discuss the research project with the students and ask them to sign a consent form. If you decide not to take part, you do not have to explain your reasons and it will not affect your students' legal rights. Participation in this project will not be linked to anything that the students are studying and it is not anticipated that there will be any detrimental effects on their sixth form studies. It is hoped that the activities they do as part of the project may have positive benefits for their personal development, particularly in relation to gaining experience in interview techniques for a job/university and knowledge about the preparation of CVs etc.

The students will be free to withdraw their consent to participate in the research project at any time, without giving a reason, even after signing the consent form. If they wish to withdraw their data, they should contact Katharine Young through youngks2@cardiff.ac.uk and it will take up to a month to delete the data. The pre-recorded Careers module will be offered to your school, so you may disseminate the resources at your will to students who have chosen not to participate.

4. What will taking part involve?

The students will complete a prerecorded 'Careers module' using Cardiff University outreach resources, with the aim of providing a space for pupils to learn about the process of applying to university or for jobs, and then having an opportunity to practice the interview process with me. With cuts to Careers Wales funding and the responsibility of exploring career opportunities falling on already-busy teachers, I hope this series of workshops will be useful to your students, to staff and the school.

Students will participate in pairs or groups of three (of their choice, and in their own time), audio-recording their discussions as they complete tasks and activities from the module. Students will send their audio-recorded discussions to the researcher using a dedicated upload link provided to them for the purpose. On completion of the module, participants will undertake a mock job/University interview with the researcher. A semi-informal interview with the researcher will follow, as an opportunity for students to reflect on their performance in the mock interview and to discuss how their hobbies and interests might be drawn upon to improve their interview performance. The audio-recordings in different settings will amount to approx. 1.5 hours in total. Interviews will take place using a virtual learning environment (such as Google Meet), and I would be grateful if you could support the participating students by arranging a space such as a classroom or quiet area to complete this. I am hoping that you will be able to help me minimize disruption to lessons by arranging recording times around free periods, breaks and lunch, where possible. It is anticipated that the peer to peer recordings will take place for 15 minutes for each workshop over the course of 4 weeks, the job interview will take approx. 40 minutes following the completion of the module and the semi-informal reflective interview is expected to take 30 minutes. In total, it is anticipated that I will be working with 20 of your students over the course of 6-8 weeks, once a week or so. They will also complete a questionnaire on the topic of the Welsh language in their lives.

5. Will my students be paid for taking part?

No. You should understand that any data the students give will be as a gift and they will not benefit financially in the future.

6. What are the possible benefits of taking part?

As well as following a module of careers preparation, students will benefit from one-to-one practice in an interview setting, and tailored feedback on their performance with me (an ex-secondary school teacher and post-graduate researcher). On top of this, their contribution will help us understand how the Welsh language is used by young speakers in different registers of formality, providing essential insight into the vitality of the Welsh language.

7. What are the possible risks of taking part?

There are no foreseeable risks to students taking part in this study.

8. Will my school's taking part in this research project be kept confidential?

All information collected from (or about) students and schools during the research project will be kept confidential and any personal information provided by the students will be managed in accordance with data protection legislation. Please see 'What will happen to participants' Personal Data?' (below) for further information.

9. What will happen to participants' Personal Data?

Cardiff University is the Data Controller and is committed to respecting and protecting participants' personal data (name, date of birth, gender, signed consent form, audio recordings and any other information from which an individual could be identified) in accordance with your expectations and Data Protection legislation.

Personal data

The following table shows the types of data that will be collected from students, and shows how the data can be pseudonymised to protect the anonymity of participating students:

<u>The data collected</u>	<u>Identifiable personal data</u>	<u>Pseudonymized personal data</u>
Audio recorded data of interviews and casual conversation with peers	Audio recordings	Pseudonymized transcripts (participant pseudonym given and other identifiable information – such as family member names etc. – removed)
Questionnaire data (name, gender, place of birth and places lived, primary school name, first language, national identity, parents' first language and occupation, Welsh use)	Original questionnaires continent participant names	Pseudonymized questionnaire data (participant pseudonym given)
Consent forms	Names, email addresses and signatures of participants	To be kept securely for as long as anonymized data to show consent was given. Following pseudonymization of above data, consent forms will not be able to be matched to individual participants.

All collected data will be pseudonymized as above (as well as the names of schools which they attend to further protect the identity of all participants in all publications).

Following efforts to pseudonymize all personal data as well as possible, the risk of identifying participants from pseudonymized data is remote, according to the Information Commissioner's Office's test of whether it is 'reasonably likely' that a participant can be recognized. There may, however, still be a risk of identifying participants (particularly in smaller, more tight-knit communities), therefore participants will be made aware of this small risk when signing the consent form. The signed consent forms will be retained indefinitely and may be accessed by members of the research project team and, where necessary, by members of the University's

governance and audit teams or by regulatory authorities, who will maintain the confidentiality of the data provided. This is to ensure that we can always prove that we did gain consent from our participants.

The researcher recognizes that according to GDPR, even pseudonymized information is still classed as personal data. The storage of personal data will be discussed below.

Data storage

The two categories below have been created in order to show how I will store the students' data safely and how I will keep their data safe and protect their anonymity:

Pseudonymized data

Pseudonymized transcripts and questionnaire data will be held indefinitely by the research team, ensuring the material is preserved and made available for further sociolinguistic research. Pseudonymized data will be stored on password-protected computer network and external hard drives.

Identifiable personal data

Identifiable personal data will be held indefinitely by the research team, stored on password-protected computers and external hard drives, ensuring the material is preserved and made available for further sociolinguistic research.

Students will also be asked whether they are happy to consent to their data being shared more widely under certain conditions (a) with other researchers in the field and (b) placed in an online repository with the ESRC (the funding body), as long as anyone accessing the data signs a release form stating that they will use the data in accordance with the consent form signed by the participants. The students' decisions about the use of their data will be respected.

Data sharing

All personal data will be kept solely by the researcher and project team unless the participant agrees to have their data shared with other researchers on an online repository by choosing the relevant options on the consent form.

Publications and conferences

Participants will be asked to provide consent for their transcript and audio recorded data (only isolated whole words and phrases, and no personal details which could identify them) to be used in future publications and conferences.

Further information about Data Protection, including:

- your rights
- the legal basis under which Cardiff University processes your personal data for research
- Cardiff University's Data Protection Policy
- how to contact the Cardiff University Data Protection Officer
- how to contact the Information Commissioner's Office

may be found at <https://www.cardiff.ac.uk/public-information/policies-and-procedures/data-protection>

10. What happens to the data at the end of the research project?

The anonymized and pseudonymized data may be shared within the University and/or shared outside of the University (e.g. in publications, presentations). If participant consent is granted, recordings and pseudonymized metadata will be placed in the ESRC data repository ReShare to enable its re-use in other research projects.

11. What will happen to the results of the research project?

The results will be used for a University assessment and are likely to be published in academic journals and presented at conferences. Participants will not be identifiable in any report, publication or presentation. Where verbatim quotes from participants are used, pseudonyms will be employed.

12. What if there is a problem?

If you wish to complain, or have grounds for concerns about any aspect of the manner in which you have been approached or treated during the course of this research, please contact the Chief Investigator/Supervisor, Dr Mercedes Durham (DurhamM@cardiff.ac.uk). If your complaint is not managed to your satisfaction, please contact the School of English Communication and Philosophy (ENCAP) Research Ethics Officer, Dr Sara Pons-Sanz (pons-sanzs@cardiff.ac.uk). If you or your students are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, you may have grounds for legal action, but you may have to pay for it.

13. Who is organising and funding this research project?

The research is organised by Katharine Young at the School of English Communication and Philosophy (ENCAP) and the School of Welsh in Cardiff University. The research is currently jointly funded by the ESRC and Welsh Government. This PhD is part of a wider project run by supervisors Dr Mercedes Durham and Dr Jonathan Morris (who along with Katharine Young form the 'project team'), funded by the Economic and Social Research Council (ESRC).

14. Who has reviewed this research project?

This research project has been reviewed and given a favourable opinion by ENCAP's Research Ethics Committee.

15. Further information and contact details

Should you have any questions relating to this research project, you may contact us during normal working hours:

Katharine Young

youngks2@cardiff.ac.uk

Room 1.57

John Percival Building

Colum Drive

CF10 3EG



11.4 Gatekeeper consent form

Title of research project: *The acquisition of sociolinguistic competence in a Welsh immersion context*

SREC reference: [Insert reference number from approval letter, once approved]

Name of Chief/Principal Investigator: *Katharine Young*

**Please
initial box**

I confirm that I have read the information sheet dated 22/07/20 version 1 for the above research project.	
I confirm that I have understood the information sheet dated 22/07/20 version 3 for the above research project and that I have had the opportunity to ask questions and that these have been answered satisfactorily.	
I understand that my school's participation is voluntary, non-participation will not have any adverse consequences to students or staff and all participants are free to withdraw any of their data at any time without giving a reason and without any adverse consequences.	
I understand that data collected during the research project may be looked at by individuals from Cardiff University or from regulatory authorities, where it is relevant to my taking part in the research project.	
I understand that participants' personal information (name, gender, ethnicity, age, place of birth, home language, parent first language, information on Welsh use outside of school, recordings of speech and consent form) will be used for the purposes explained to me. I understand that such information will be held in accordance with all applicable data protection legislation and in strict confidence, unless disclosure is required by law or professional obligation.	
I understand that pseudonymized data will be held indefinitely by the research team and that it will be securely stored on password protected computers and hard drives.	
I agree for participants take part in this research project if they are willing to do so after being asked for their free and informed consent.	

Name of participant (print)

Date

Signature

THANK YOU FOR PARTICIPATING IN OUR RESEARCH
YOU WILL BE GIVEN A COPY OF THIS CONSENT FORM TO KEEP

11.5 Information letter for parents

THE ACQUISITION OF SOCIOLINGUISTIC COMPETENCE IN A WELSH IMMERSION CONTEXT

Your child is being invited to take part in a research project. You are encouraged to discuss the project with your child before they decide to take part.

Thank you for reading this.

1. What is the purpose of this research project?

The purpose of this project is to understand to what extent home language and community language affect the way young Welsh speakers vary their language according to the situation they're talking in (i.e. more formal language for more formal situations).

2. Why has my child been invited to take part?

Your child has been invited because they are known to use Welsh in a variety of different situations at school, some more formal and some more casual.

3. Does my child have to take part?

No, your child's participation in this research project is entirely voluntary and it is up to them to decide whether or not to take part. If they decide to take part, we will discuss the research project with your child and ask them to sign a consent form. If they decide not to take part, they do not have to explain their reasons and it will not affect their legal rights. It should also be noted that involvement in this research project will have no effect on their education or progression through a sixth form course.

Your child is free to withdraw their consent to participate in the research project at any time, without giving a reason, even after signing the consent form.

4. What will taking part involve?

Your child will be audio recorded in 3 situations; a chat with friends, a mock job/university interview and a follow up reflective interview. Your child will complete a prerecorded 'Careers module', which will provide them with opportunities to gain knowledge about the process of writing CVs and preparing for interviews. Your child will be invited to complete the workshops with a friend or in groups of 3, and will be asked to audio-record their discussions as they complete tasks and activities from the module. I ask that your child uses their mobile phone to record their discussions during each workshop, sending me their audio-recorded discussions using a safe and dedicated upload link which will be sent to them for the purpose. After completing the module, I will offer your child the opportunity to undertake a mock job/University interview with me, to give them one-to-one practice and support with what can be a stressful experience. They will then have a follow-up interview with me, as an opportunity for them to reflect on their performance in the mock interview and I will ask them about their hobbies and interests, and a chance for them to tell me a bit more about their lives. The audio-recordings in different settings will amount to approx. 1.5 hours in total. This will take place using a virtual learning environment (such as Google Meet), in a classroom at school. The school and your child will help me to minimize disruption to lessons by arranging recording times around break, lunch and free periods where possible – this will be decided between your child and the school. It is anticipated that peer-to-peer recordings for each workshop in the module will last around 15 minutes over the course of around 4 weeks. The mock interview following the module will take around 40 minutes to complete, and the follow-up feedback interview will take around 30 minutes to complete. Overall, I expect to be working with your child once a week over the course of 6-8 weeks. Your child will also complete a questionnaire on the topic of the Welsh language in their lives.

5. Will my child be paid for taking part?

No. You should understand that any data your child gives will be as a gift and they will not benefit financially in the future.

6. What are the possible benefits of taking part?

Your child will benefit from a tailored module of workshops to help them to prepare for their future careers, with a chance for a one-to-one mock interview and individual follow up reflective interview. On top of that, their contribution will help us understand how the Welsh language is used by young speakers in different registers of formality, providing essential insight into the vitality of the Welsh language.

7. What are the possible risks of taking part?

There are no foreseeable risks to your child taking part in this study.

8. Will my child's taking part in this research project be kept confidential?

Any personal information your child provides will be managed in accordance with data protection legislation. Your child will be given the option as to whether they consent to their data being shared with other researchers in the field or on an online repository. Please see 'What will happen to my child's Personal Data?' (below) for further information.

9. What will happen to my child's Personal Data?

Cardiff University is the Data Controller and is committed to respecting and protecting participants' personal data (name, date of birth, gender, signed consent form, audio recordings and any other information from which an individual could be identified) in accordance with your expectations and Data Protection legislation.

Personal data

The table below shows the different types of data that will be collected with your child's help, and well as the ways in which I will attempt to protect their anonymity:

<u>The data collected</u>	<u>Identifiable personal data</u>	<u>Pseudonymized personal data</u>
Audiorecorded data of interviews and casual conversation with peers	Audiorecordings	Pseudonymized transcripts (participant pseudonym given and other identifiable information – such as family member names etc. – removed)
Questionnaire data (name, gender, place of birth and places lived, primary school name, first language, national identity, parents' first language and occupation, Welsh use)	Original questionnaires containing participant names	Pseudonymized questionnaire data (participant pseudonym given)

Consent forms	Names, email addresses and signatures of participants	To be kept securely for as long as pseudonymized data to show consent was given. Following pseudonymization of above data, consent forms will not be able to be matched to individual participants.
---------------	---	---

All collected data will be pseudonymized as above (as well as the names of schools which they attend to further protect the identity of all participants in all publications).

Following efforts to pseudonymize all personal data as well as possible, the risk of identifying your child from pseudonymized data is remote, according to the Information Commissioner's Office's test of whether it is 'reasonably likely' that a participant can be recognised. There may, however, still be a risk of identifying your child (particularly in smaller, more tight-knit communities), therefore your child will be made aware of this small risk when signing the consent form. The signed consent forms will be retained indefinitely and may be accessed by members of the research project team and, where necessary, by members of the University's governance and audit teams or by regulatory authorities, who will maintain the confidentiality of the data provided. This is to ensure that we did gain consent from our participants.

The researcher recognizes that according to GDPR, even pseudonymized information is still classed as personal data. The storage of personal data will be discussed below.

Data storage

The two categories below have been created in order to show how I will store your child's data safely and how I will keep their data safe and protect their anonymity:

Pseudonymized data

Pseudonymized transcripts and questionnaire data will be held indefinitely by the research team, ensuring the material is preserved and made available for further sociolinguistic research. pseudonymized data will be stored on password-protected computer network and external hard drives.

Identifiable personal data

Identifiable personal data will be held indefinitely by the research team, ensuring the material is preserved and made available for further sociolinguistic research.

Your child will also be asked whether they are happy to consent to their data being shared (with their names removed from audio recordings) more widely under certain conditions (a) with other researchers in the field and (b) placed in an online repository with the ESRC (the funding body), as long as anyone accessing the data signs a release form stating that they will use my data in accordance with the consent form signed by your child. Your child's decisions about the use of their data will be respected.

Data sharing

All personal data will be kept by the researcher and project team unless your child agrees to have their data shared with other researchers or an online repository by choosing one option from the consent form.

Publications and conferences

Participants will be asked to provide consent for their transcript and audiorecorded data (only isolated words and phrases, and no personal details which could identify them) to be used in future publications and conferences.

Further information about Data Protection, including:

- Your child's rights
- the legal basis under which Cardiff University processes your child's personal data for research
- Cardiff University's Data Protection Policy
- how to contact the Cardiff University Data Protection Officer
- how to contact the Information Commissioner's Office

may be found at <https://www.cardiff.ac.uk/public-information/policies-and-procedures/data-protection>

10. What happens to the data at the end of the research project?

The pseudonymized data may be shared within the University and/or shared outside of the University (e.g. in publications, presentations).

11. What will happen to the results of the research project?

The results will be used for a University assessment and are likely to be published in academic journals and presented at conferences. Your child will not be identifiable in any report, publication or presentation. Where verbatim quotes from your child are used, pseudonyms will be employed.

12. What if there is a problem?

If you wish to complain, or have grounds for concerns about any aspect of the manner in which you or your child have been approached or treated during the course of this research, please contact the Chief Investigator/Supervisor, Dr Mercedes Durham (DurhamM@cardiff.ac.uk). If your complaint is not managed to your satisfaction, please contact the ENCAP Research Ethics Officer, Dr Sara Pons-Sanz (pons-sanzs@cardiff.ac.uk). Complaints will be dealt with within a month of being raised, and if your child wishes to remove their data from the study, the researcher will need a month to process this and delete the data.

If your child is harmed by taking part in this research project, there are no special compensation arrangements. If your child is harmed due to someone's negligence, you may have grounds for legal action, but you may have to pay for it.

13. Who is organising and funding this research project?

The research is organised by Katharine Young at the School of English Communication and Philosophy and the School of Welsh in Cardiff University. The research is currently jointly funded by the Economic and Social Research Council and Welsh Government. This PhD is part of a wider project run by supervisors Dr Mercedes Durham and Dr Jonathan Morris (who

along with Katharine Young form the ‘project team’), funded by the Economic and Social Research Council (ESRC).

14. Who has reviewed this research project?

This research project has been reviewed and given a favourable opinion by ENCAP’s Research Ethics Committee.

15. Further information and contact details

Should you have any questions relating to this research project, you may contact us during normal working hours:

Katharine Young

youngks2@cardiff.ac.uk

Room 1.57

John Percival Building

Colum Drive

CF10 3EG

Thank you for considering participating in this research project. If your child decides to participate, they will be given a copy of the Participant Information Sheet and a signed consent form to keep for their records.



11.6 Information letter for participants

THE ACQUISITION OF SOCIOLINGUISTIC COMPETENCE IN A WELSH IMMERSION CONTEXT

You are being invited to take part in a research project. Before you decide whether or not to take part, it is important for you to understand why the research is being undertaken and what it will involve. Please take time to read the following information carefully, and discuss the research with your parent/guardian if you wish to. If you agree to participate, you will be asked to sign a consent form.

Thank you for reading this.

1. What is the purpose of this research project?

The purpose of this project is to understand to what extent home language and community language affect the way young Welsh speakers vary their language according to the situation they're talking in (i.e. more formal language for more formal situations).

2. Why have I been invited to take part?

You have been invited because you are known to use Welsh in a variety of different situations at school, some more formal and some more casual.

3. Do I have to take part?

No, your participation in this research project is entirely voluntary and it is up to you to decide whether or not to take part. If you decide to take part, we will discuss the research project with you and ask you to sign a consent form. If you decide not to take part, you do not have to explain your reasons and it will not affect your legal rights. It should also be noted that involvement in this research project will have no effect on your education or progression through a sixth form course.

You are free to withdraw your consent to participate in the research project at any time, without giving a reason, even after signing the consent form.

4. What will taking part involve?

You will be audio recorded in 3 situations; chats with friends, a mock job/university interview and a reflective interview to talk about your hobbies and interests. You will be invited to complete a prerecorded 'Careers module' designed to help prepare you for the world of work and university applications. You can complete each workshop in the module with a friend, and I will ask you to audio-record your discussions as you complete tasks and activities from the module. You will use your phone to record your discussions, and then send your discussions to me through a safe and direct link that I will send you. After you've finished all the workshops in the module, you'll be offered a mock job/University interview with me to help you prepare for the interview experience. I'll then ask you to come to a follow up interview, as an opportunity for you to reflect on your performance in the mock interview and I'll also ask you some questions about hobbies and interests. The audio-recordings in different settings will amount to approx. 1.5 hours in total. This will take place using a virtual learning environment (such as Google Meet) and you can talk to your school about arranging to use a classroom or quiet space to do this. I hope to work with you and your school to minimize disruption to lessons by arranging recording times around break, lunch and free periods where possible. It is anticipated that the

module is made up of a series of workshops which will take around 45 minutes to complete, once a week. Your chats with friends are expected to last 15 minutes for each workshop. The mock interview, after you've finished all the workshops will last around 40 minutes, and the follow up reflective interview will last around 30 minutes. I expect to work with you (once a week) over the course of 6-8 weeks. You will also complete a questionnaire on the topic of the Welsh language in your life.

5. Will I be paid for taking part?

No. You should understand that any data you give will be as a gift and they will not benefit financially in the future.

6. What are the possible benefits of taking part?

You will benefit from one-to-one support in preparing for the world of work/university, as well as a tailor-made series of workshops to help improve your knowledge of CV-building and application-writing. On top of that, your contribution will help us understand how the Welsh language is used by young speakers in different registers of formality, providing essential insight into the vitality of the Welsh language.

7. What are the possible risks of taking part?

There are no foreseeable risks to you taking part in this study.

8. Will my taking part in this research project be kept confidential?

Any personal information you provide will be managed in accordance with data protection legislation. You will decide whether you are happy for your data to be shared with other researchers in the field or for it to be placed into an online data repository. Please see 'What will happen to my Personal Data?' (below) for further information.

9. What will happen to my Personal Data?

Cardiff University is the Data Controller and is committed to respecting and protecting participants' personal data (name, date of birth, gender, signed consent form, audio recordings and any other information from which an individual could be identified) in accordance with your expectations and Data Protection legislation.

Personal data

The table below shows the different types of data we'll be working together to produce, and the measures I will take to help protect your anonymity during the course of the project:

<u>The data collected</u>	<u>Identifiable personal data</u>	<u>Pseudonymized personal data</u>
Audiorecorded data of interviews and casual conversation with peers	Audiorecordings	Pseudonymized transcripts (participant pseudonym given and other identifiable information – such as family member names etc. – removed)
Questionnaire data (name, gender, place of birth and places lived, primary school	Original questionnaires containing participant names	Pseudonymized questionnaire data

name, first language, national identity, parents' first language and occupation, Welsh use)		(participant pseudonym given)
Consent forms	Names, email addresses and signatures of participants	To be kept securely for as long as pseudonymized data to show consent was given. Following pseudonymization of above data, consent forms will not be able to be matched to individual participants.

All collected data will be pseudonymized as above (as well as the names of schools which you attend to further protect the identity of all participants in all publications).

Following efforts to pseudonymize all personal data as well as possible, the risk of identifying you from pseudonymized data is remote, according to the Information Commissioner's Office's test of whether it is 'reasonably likely' that a participant can be recognised. There may, however, still be a risk of identifying you (particularly in smaller, more tight-knit communities), therefore you should be aware that this small risk exists. The signed consent forms will be retained indefinitely and may be accessed by members of the research project team and, where necessary, by members of the University's governance and audit teams or by regulatory authorities, who will maintain the confidentiality of the data provided. This is to ensure that we did gain consent from you.

The researcher recognizes that according to GDPR, even pseudonymized information is still classed as personal data. The storage of personal data will be discussed below.

Data storage

The two categories below have been created in order to show how I will store your data safely and how I will keep your data safe and protect your anonymity:

Pseudonymized data

Pseudonymized transcripts and questionnaire data will be held indefinitely by the research team, ensuring the material is preserved and made available for further sociolinguistic research. pseudonymized data will be stored on password-protected computer network and external hard drives.

Identifiable personal data

Identifiable personal data will be held indefinitely by the research team, ensuring the material is preserved and made available for further sociolinguistic research.

In the consent form, you will be asked whether you are happy to consent to your data being shared more widely under certain conditions (a) with other researchers in the field and (b) placed in an online repository with the ESRC (the funding body), as long as anyone accessing the data signs a release form stating that they will use my data in accordance with the consent form signed by you. Your decisions about the use of your data will be respected.

Data sharing

All personal data will be kept by the researcher and project team unless you agree to have your data shared with other researchers or an online repository by choosing one option from the consent form

Publications and conferences

You will be asked to provide consent for your transcript and audiorecorded data (only isolated words and phrases, and no personal details which could identify you) to be used in future publications and conferences.

Further information about Data Protection, including:

- your rights
- the legal basis under which Cardiff University processes your personal data for research
- Cardiff University's Data Protection Policy
- how to contact the Cardiff University Data Protection Officer
- how to contact the Information Commissioner's Office

may be found at <https://www.cardiff.ac.uk/public-information/policies-and-procedures/data-protection>

10. What happens to the data at the end of the research project?

The pseudonymised data may be shared within the University and/or shared outside of the University (e.g. in publications, presentations).

11. What will happen to the results of the research project?

The results will be used for a University assessment and are likely to be published in academic journals and presented at conferences. You will not be identifiable in any report, publication or presentation. Where verbatim quotes from your transcript are used, pseudonyms will be employed.

12. What if there is a problem?

If you wish to complain, or have grounds for concerns about any aspect of the manner in which you have been approached or treated during the course of this research, please contact the Chief Investigator/Supervisor, Dr Mercedes Durham (DurhamM@cardiff.ac.uk). If your complaint is not managed to your satisfaction, please contact the ENCAP Research Ethics Officer, Dr Sara Pons-Sanz (pons-sanzs@cardiff.ac.uk). Complaints regarding any aspect of data collection will be dealt with within a month of being raised, and if you wish to withdraw your data, contact Katharine Young on youngks2@cardiff.ac.uk and your data will be deleted within one month.

If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, you may have grounds for legal action, but you may have to pay for it.

13. Who is organising and funding this research project?

The research is organised by Katharine Young at the School of English Communication and Philosophy and the School of Welsh in Cardiff University. The research is currently jointly funded by the Economic and Social Research Council and Welsh Government. This PhD is part of a wider project run by supervisors Dr Mercedes Durham and Dr Jonathan Morris (who

along with Katharine Young form the ‘project team’), funded by the Economic and Social Research Council (ESRC).

14. Who has reviewed this research project?

This research project has been reviewed and given a favourable opinion by ENCAP’s Research Ethics Committee.

15. Further information and contact details

Should you have any questions relating to this research project, you may contact me during normal working hours:

Katharine Young

youngks2@cardiff.ac.uk

Room 1.57

John Percival Building

Colum Drive

CF10 3EG

Thank you for considering participating in this research project. If you decide to participate, you will be given a copy of the Participant Information Sheet and a signed consent form to keep for your records.



11.7 Participant consent form (for students)

Title of research project: *The acquisition of sociolinguistic competence in a Welsh immersion context*

SREC reference: [Insert reference number from approval letter, once approved]

Name of Chief/Principal Investigator: *Katharine Young*

**Please
initial box**

I confirm that I have read the information sheet dated 22/07/20 version 3 for the above research project.	
I confirm that I have understood the information sheet dated 22/07/20 version 3 for the above research project and that I have had the opportunity to ask questions and that these have been answered satisfactorily.	
I understand that my participation is voluntary and I am free to withdraw any of my data at any time without giving a reason and without any adverse consequences.	
I understand that data collected during the research project may be looked at by individuals from Cardiff University or from regulatory authorities, where it is relevant to my taking part in the research project. I give permission for these individuals to have access to my personal data.	
I consent to the processing of my personal information (name, email address, gender, ethnicity, age, place of birth, home language, parent first language, information on Welsh use outside of school, recordings of speech and consent form) for the purposes explained to me. I understand that such information will be held in accordance with all applicable data protection legislation and in strict confidence, unless disclosure is required by law or professional obligation.	
I understand that my pseudonymized data will be held indefinitely by the research team and that it will be securely stored on password protected computers and hard drives.	
Please consent to one of the following (a, b or c):	
a) I consent to only the researcher and project team holding my data; or	

b) I consent to the researcher and project team holding my data (as stated in a) above) and sharing it with other researchers as long as they sign a release form stating that they will use my data in accordance with the current consent form; or	
c) I consent to the researcher holding my data, sharing it with other researchers (as stated in a) and b) above) and placing my data within 5 years of the completion of the project in a secure online data repository where it can be accessed by other researchers.	
I consent to being audio recorded for the purposes of the research project and I understand how the audio recording will be used in the research.	
I understand that pseudonymized excerpts and/or verbatim quotes from transcripts of my audio recordings and questionnaire may be used as part of the research publication.	
I understand that excerpts of isolated linguistic features from my audio recordings may be used in conferences, but that any accompanying personal data will be pseudonymized in order to protect my identity.	
I understand how the findings and results of the research project will be written up and published / publicised.	
I agree to take part in this research project.	

Name of participant (print)

Date

Signature

Hwb email address of participant

THANK YOU FOR PARTICIPATING IN OUR RESEARCH

YOU WILL BE GIVEN A COPY OF THIS CONSENT FORM TO KEEP

11.8 Example interview questions

You've applied to be a Maths or English tutor. Which subject would you prefer to tutor? Why?

What is your previous experience in the field?

What achievement has made you proudest of yourself?

Tell me about the last book you read.

Can you tell me about a time when you worked well with others?

Can you tell me about a time when you managed deadlines well?

What would you like to achieve by the end of your first week at the centre?

What motivates you?

Why do you want to work here?

Tell me about a time you overcame a difficult situation at work/school

What interests you about this role?

What are your greatest strengths and weaknesses?

What is your dream job?

What qualities do you think make a good leader?

What qualities do you think are needed in this role?

Why do you feel you are the right person for this position?

Where do you see yourself in five years? 10 years?

What do you want to accomplish in the first 30 days in this role?

What are you looking forward to most about coming to work? What is your least favourite?

How do you keep yourself organised at work?

Who are your role models, and why?

What are your hobbies and interests?

What are you studying at A-Level? Why did you choose that subject?

What is your proudest achievement?

11.9 Sociolinguistic interview questions

Modules:

Family and friends (demographic – based on the pre-interview questionnaire)

Firstly how long have you known your best friend (as stated on the questionnaire)?

How did you become friends?

Do you know their family? Tell me about them.

What do you like about them?

Tell me a bit about your family, where are they from?

Why did they move to where you live now?

Community

What was it like growing up in _____?

Do you have a favourite memory as a child? (birthday or Christmas?)

Has anything changed in your area since when you were a child?

How is where you grew up different to where you live now?

What was moving house like?

Do you have next door neighbours?

What do you know about them?

You live in _____, what do you like about your area?

What don't you like?

Do you use local amenities (parks, leisure centres, cinemas, restaurants, bars, friends' houses etc.)?

What do you do there?

Culture

In your opinion is your area different from other places in Wales?

What's your favourite place to go locally or elsewhere?

What do you do there?

Do you go to the cinema/theatre with school or with family and friends?

What did you last see?

What's your favourite type of film/show?

Have you ever been to any music gigs or concerts?

Do you go often?

What's the best gig you've been to and why?

Who's your favourite artist?

Do you or your family support any sports teams locally?

When was the last time you went to see them?

What happened in the game?

School

What's the best thing about going to this school?

Is there a difference between this school and (another local school)?

Why are they different?

What's the worst thing?

When I was in sixth form people complained a lot about canteen food, what do you think about it?

What subjects do you take for A/AS Level?

What's your favourite / least favourite subject and why?

Do you like your teachers?

Have you ever been told off at school (or elsewhere)?

Are you allowed to leave the school site at lunch/break/free periods?

What do you do usually?

Were you ever given detention?

What happened?

Have you been on many school trips since you started secondary school / sixth form?

Where did you go?

Did anything interesting happen?

Do you remember a particularly memorable event from being at school (funny, terrible) from school (primary, secondary or sixth form)?

What do you remember about your first day at sixth form/secondary school?

Friends and relatives

Are you still friends with some people from your first day at secondary school?

If not, why?

Have you known any of your friends since primary school or earlier?

Does anyone you know of fall out with their friends in sixth form?

What do they fall out over?

Films tend to show stereotypical cliques of friendship groups in high school, do you think that represents life as you know it at your sixth form?

Which clique would you say you belong to?

Can you name some of the cliques you think exist?

Fear (danger of death)

Do you have any phobias?

When was the last time you saw a spider/clown/snake...?

How did you feel?

Do you tend to watch horror films?

What was the last horror film you saw?

Can you describe a scene from your favourite horror film?

Hobbies (and games)

What do you do in your spare time outside of school?

Are you a member in a club?

Think of a particular weekend, holiday or school inset recently, what did you and your friends do?

Do you and your friends ever go to town?

What do you do there?

What happened last time you went to town?

Think back, what differences are there between your hobbies now and your hobbies 5 years ago?

Why did you stop doing some hobbies?

Why did you take up _____?

When you were younger, what type of things would you do with your friends?

Did you play games at primary/secondary school?

What kinds of games?

Do you have an Xbox or PS4?

What games do you play?

Tell me about that game?

Do you use social media? Which ones?

What do you put on TikTok/Snapchat/Insta?

What do you like about TikTok/Snapchat/Insta?

Do you think social media is a good thing?

Have you heard about anyone having problems in the past with social media?

Have you ever done an escape room? What happened in it, what did you have to do?

Did you manage to escape?

Were any of your friends/family better than each other in the escape room? Why?

What television and film do you watch?

Do you have Netflix/Prime/NowTV?

What's your favourite program at the moment?

Describe the series?

Do you subscribe to any YouTube channels?

What kind of channels are they?

Which channel do you watch most?

Tell me what happens in an episode?

Is there any difference between things you do with friends from school and friends from the community who go to different schools?

Why?

Are those friends different or similar?

Future

When you were younger, what did you want to be when you grew up? Do you still want to do that now?

What do you think you will do after sixth form?

What's your dream job?

Work

Do you have a job or have you ever had a part time job?

What did you do in a regular shift?

How did you get the job?

How often do/did you work?

Language

Do you speak Welsh at work? (if they have a part time job)

If you do, do you speak differently at work to what you would with your friends?

Are you more comfortable speaking Welsh with friends or in other situations (work, family, school, teachers, interviews etc.)

Why?

Do you think your Welsh changes depending on who you're talking to?

Do you think your English changes depending on who you're talking to?

What kind of Welsh does your (name subject) teacher talk?

Is it different to yours?

How?

Do you think you have a local accent?

Can you describe what a local accent sounds like in Welsh and English?

Is there anything unique about the way you speak Welsh or English?

Do you use specific words or pronunciations which are different to your friends? Why do you think?

11.10 Questionnaire

This questionnaire will help me to find out more about you before I interview you. Please ask the interviewer if you are unsure about anything. Where there isn't a space to answer, circle the relevant options.

1. Name:

2. Gender: Male Female Other Prefer not to say

3. Where were you born? _____

4. Where do you live? _____

5. Have you ever lived anywhere other than the place you specified in Q4.?

Yes No

If you stated 'yes', answer Q6.

6. List the other places you've lived:

Other places you've lived	From (date)	To (date)

7. Which primary school did you go to? _____

8. What's your first language? Welsh English Other

If you stated 'other', please specify:

9. Describe your national identity:

Welsh British Welsh and British Other

If you circled 'other', please specify:

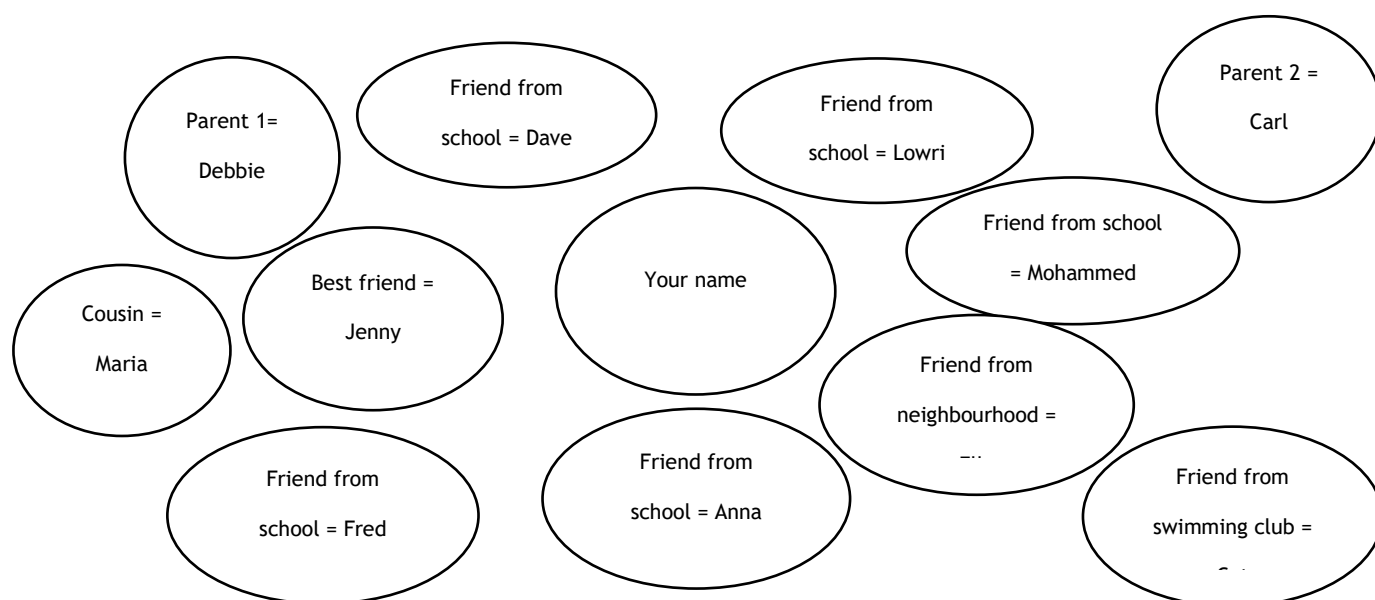
10. Fill in this table about your parents/guardians:

Parent (e.g Mum, Stepdad)	From where?	What language(s) do you use with them?	What is their job?

11. How often do you speak Welsh outside of school?

Every day Every week Every month Never

12. Starting with yourself in the middle, draw a sociogram on the next page, noting the names of your closest relatives and friends (and their relationship to you) surrounding it. (see the example below):



Colour each surrounding bubble according to the language you use with them:

Yellow for people who you speak mostly Welsh with.

Orange for people who can't speak Welsh

Bright red for people who you use Welsh and English with a roughly even split between the two languages.

Dark red for people who you speak mostly English with, even though this person can speak Welsh

11.11 Parent jobs by area

Cardiff	Number
Civil servant	2
Teacher	2
Welsh/English translator	2
Patent examiner	1
Solicitor	1
Accountant	1
Book Publisher	1
Lecturer	1
Location warden	1
Early retirement	1
Designer	1
Social worker	1
Doctor (GP)	1
Gwynedd	
Care worker	3
Builder	1
Farmer	1
Farmer/mechanic	1
Lorry driver	1
Paramedic	1
Engineer	1
Teacher	1
Administrator	1
Works in a kitchen	1
Nurse	1

11.12 Intensifiers identified in the literature

56 intensifiers identified in the literature and searched for			
So	Extremely	Fairly	Arswysur
Very	Incredibly	Quite	Aruthrol
Really	Completely	Braidd	Cythreuliedig
Bloody	Hollol	Cweit	Diawledig
Mor	Llwyr	Eithaf	Difrifol
Iawn	Hynod	Ychydig	Dros ben
Rili	Eithriadol	bach	Dychrynlllyd
Blydi	Ofnadwy	Weddol	Eithriadol
Wir	Cwbl	Reit	Gwirioneddol
Go	Llawn	Cymharol	Neilltuol
Tra	Holl	Digon	Pur
Mega	Perffaith	Andros	Lled
Totally	Cryn	Uffernol	Rhy
Perfectly	Pretty	Arbennig	