Morgan Sleeper (msleeper@macalester.edu)
To appear in Journal of Celtic Linguistics (pre-print version)

Placename mutation variation in Wales and Patagonia

1. Introduction

Initial consonant mutation – a morphophonological process that affects certain word-initial consonants – is a characteristic feature of the Celtic languages. In Welsh – spoken in Wales and in the Chubút province of Argentine Patagonia – there are three distinct types of initial consonant mutations: soft mutation, nasal mutation, and aspirate mutation, listed in Table 1:

Table 1: Welsh Initial Consonant Mutations

Soft Mutation

Radical		Mutation
/p/	→	/b/
/t/	→	/d/
/k/	→	/g/
/b/	→	/f/
/d/	→	/ð/
/g/	→	∅
/m/	→	/f/
/ɬ/	→	/l/
/r̥/	→	/r/

Nasal Mutation

Radical		Mutation
/p/	→	/m̥/
/t/	→	/n̥/
/k/	→	/ŋ̊/
/b/	→	/m/
/d/	→	/n/
/g/	→	/ŋ/

Aspirate Mutation

Radical		Mutation
/p/	→	/f/
/t/	→	/θ/
/k/	→	/x/

Each of these three mutations apply in different morphological, phonological, and syntactic contexts. Nasal mutation, for instance, affects words following the locative particle *yn* 'in', the first-person singular possessive pronoun *fy* 'my', and the negative prefix *an-* 'un-'. When preceded by one of these triggering morphemes, word-initial /p/, /t/, /k/, /b/, /d/, or, /g/ will mutate into a homorganic nasal (produced in the same place of articulation) while retaining their original voicing specification (+VOICE or -VOICE), as in the following example:

(1)	**Bangor** ['baŋgɔr]		**ym**	**Mangor** ['maŋgɔr]		
	Bangor	→	in	Bangor		
	'Bangor'		'in Bangor'			(Hannahs 2013:127)

As a relatively cross-linguistically rare phenomenon on the interface between morphology, syntax, and phonology, mutations have long attracted the attention of linguists, and especially phonologists. Welsh mutations have been analyzed in phonological frameworks ranging from autosegmental phonology (Lieber 1987) and Optimality Theory (Gnanadesikan 1997) to pattern extraction (Hannahs 2013). While accounts of the alternations themselves are widespread, comparatively less is known about the specifics of how mutations vary in use.

The paradigm for nasal mutation described above, for example, is presented as standard in grammars of Welsh (King 2003:17-18; Hannahs 2013:126-127; Thomas 1984:220; Willis 1986:7), yet dialectal and socially-conditioned variation, as well as individual differences, are well-documented in speakers' use of the nasal mutation in Wales. King (2003:17) states that nasal mutation is “not consistently applied after [*yn* or *fy*] in many parts of Wales”, while Thomas (1984:214) notes that in South Glamorgan Welsh, which lacks phonemic voiceless nasals, both voiced and voiceless stops are replaced by their homorganic voiced nasals, and that different groups of speakers realize the mutation more or less frequently in expected environments (221-222). Further, Hannahs (2013:125) points out that lexical restrictions can block mutations entirely in certain words, Prys (2016:234) finds that placenames are significantly less likely to be mutated than non-placenames, and Ball & Müller (1992:205) note that placenames designating locations outside of Wales may be less likely to participate in mutation than Welsh placenames.

This variation in nasal mutation is especially important in terms of the phonological system of Welsh, since the voiceless nasals /m̥/, /n̥/, and /ŋ̊/ it prescriptively results in are only considered phonemes as a result of their inclusion in nasal mutation (Watkins 1993). Because of their frequency in natural discourse directly following *yn* 'in' (a canonical environment for nasal mutation), and their unique interactions with geography (Ball & Müller 1992:205), placenames provide an excellent vehicle for studying this nasal mutation variation in modern conversational Welsh.

Taking these observations as a starting point, the present study uses corpus data to examine the following research questions:

(1) How do speakers in both Wales and Patagonia mutate (or not mutate) placenames following the locative *yn* 'in' in modern, conversational Welsh?

(2) What factors influence mutation in this context?

Results find no significant difference in how Patagonian speakers and speakers from Wales mutate in this environment, but show that the radical initial consonant, speaker age, and type of placename – how geographically, linguistically, and culturally 'Welsh' the placename is – all affect mutation behaviour. Further, while nasal mutation is certainly present in the data, these results also show the growing importance of soft mutation in the prescriptively nasally mutating environment following *yn*.

2. Background

2.1 Patagonian Welsh

The Welsh-speaking presence in Argentine Patagonia stems from the 1865 founding of *Y Wladfa* ('the colony'), a settlement of Welsh speakers in the Chubút Valley, a desert area that the Argentine government hoped to populate with settlers to solidify their claim to the territory over Chile (Birt 2005). The settlement began with 160 colonists who left Wales specifically in order to preserve the Welsh language, religion, and cultural values away from English influence, and remained largely independent and Welsh-speaking until the second World War, with Welsh-language schools, churches, and local government maintaining the language's importance in daily life (Johnson 2009; Jones 1984; Birt 2005). After the end of the war, however, the Argentine government increased its influence in the area, and migration both from Europe and other Spanish-speaking areas of Argentina led to a diglossic situation, where Spanish became the language of daily public life, and Welsh was restricted to Welsh-speaking homes and social circles (Johnson 2009; Williams 1991).

The 1965 centenary of the colony, however, marked a turning point for the language in Patagonia, with renewed interest in Welsh both as a heritage language for descendants of the original settlers, and as an aspect of shared cultural heritage for all residents of modern Chubút. This shift, along with a rise in festivals, events, and contact and cultural visits from Wales, led to a major improvement in the status of the language which has continued through the 1990s and 2000s (Jones 1996; Johnson 2009). In addition, access to Welsh-medium education has increased steadily over the last 20 years, especially since the 1997 founding of Prosiect yr Iaith Gymraeg ('Welsh Language Project') – a cooperative grant agreement for promoting Welsh in Chubut managed by the British Council in collaboration with the Welsh Government, the Wales-Argentina Society, and Cardiff University, among other organizations (Arwel 2016). As of 2018, Ysgol Feithrin Gaiman, Ysgol yr Hendre in Trelew, and the recently opened (2015) Ysgol y Cwm provide Welsh-immersion nursery and primary schooling in Chubut, and Ysgol Gymraeg yr Andes in Trevelin and Ysgol Camwy in Gaiman offer successful after-school programs for older children and adult learners (Kiff 2013; Arwel 2016). Prosiect yr Iaith Gymraeg. Precise numbers of current speakers are difficult to obtain, but Ó Néill (2005:429) reports that in 2005 there were at least “several thousand” speakers in Chubut with “some knowledge of Welsh with varying degrees of fluency”.

While there has been considerable sociocultural interest in the Welsh spoken in Patagonia – from vitality and revitalization (Johnson 2009; Birt 2005) to linguistic identity (Trosset et al. 2007) and linguistic landscapes (Coupland & Garrett 2010) – fewer studies have focused on unique structural and/or phonetic features of Patagonian Welsh. In one of the first investigations of phonetic features of Patagonian Welsh, Jones (1984) provides a variationist account of Welsh as spoken in Gaiman, Chubut, and shows the varying productions of /x/ vs. /x̟/, /ʃ/ vs. /s/, /ə/ vs. /ɪ/, and /p^h t^h k^h/ vs. /p t k/ by different speaker groups, with each of the latter realizations theorized as influences from Spanish contact.

More recent studies continue this focus on potential contact effects from Spanish. Agozzino (2006) profiles lexical contact features, including calques, loanwords, and hybrid Welsh-Spanish idioms unique to Patagonian Welsh. Bell (2015) examines phonemic vowel length in Patagonian Welsh, showing that Welsh-Spanish bilingual speakers continue to produce the phonemic vowel length contrast in Welsh, regardless of which language was acquired first (Welsh L1 or Spanish L1). Sleeper (2015), following up on one aspect of Jones' (1984) study, finds that Patagonian Welsh speakers produce the voiceless stops /p t k/ with significantly shorter VOT values than speakers from Wales.

2.2 Mutation variation

Given the relative typological rarity of consonant mutation as a linguistic system, variation in Welsh mutation realization has long been of interest to linguists. In one of the first forays into Welsh dialectology, for instance, Sweet (1884:432-435, cited in Ball & Müller 1992) points out that in Dyffryn Gwynant Welsh, feminine nouns do not appear to mutate after *un* 'one', and notes the dialect's non-standard inclusion of /m/, /n/, and /w/ into the aspirate mutation paradigm by adding an /h/ after each consonant.

Ball (1985) examines mutation – among other linguistic variables – in the context of stylistic variation, by looking at the realization of aspirate mutation and nasal mutation after *yn* in three different types of programming on Radio Cymru (BBC's Welsh-language radio network): a read news show, a spontaneous variety program, and a religious service that had a mix of the two modes of delivery. He found that while nasal mutation following *yn* remained consistently standard (apart from a single token realized with soft mutation), use of the aspirate mutation varied greatly by program, with the news show at 100% standard (i.e. showing 100% aspirate mutation, as opposed to soft mutation or a radical realization), the religious broadcast 80% standard, and the variety show only 31% standard.

More recently, Prys (2016) also explores mutation variation as a stylistic variable in speech on Radio Cymru, by comparing radio presenters' speech across programs of varying formalities to corpus data from natural vernacular speech. He finds that the aspirate mutation triggers *a* 'and' and *â/gyda* 'with/as' act as stylistic markers which speakers can manipulate for stance-taking purposes, with the use of prescriptively 'standard' mutation realizations (which diverge from community norms) being associated with formality and authority, and 'non-standard' realizations (more in line with community norms) with informality. Mutation variation following the nasal mutation trigger *fy* 'my' and the soft mutation triggers *ei* 'his', *am* 'for', and *o* 'from', however, was found to reflect speakers' social backgrounds rather than stylistic stratification. *Yn* 'in' interestingly represents a hybrid of the two, with its usage partially reflecting stylistic stratification, but complicated by both the special consideration of following placenames and the frequency of presenters using soft mutation instead of nasal mutation following *yn* (Prys 2016:384).

Howell (2010) profiles the entire mutation system of Pembrokeshire Welsh by speaker age and mutation type, and finds that younger speakers mutated less than either middle-aged or older speakers, and that nasal mutation and aspirate mutation – while in decline among all speakers – were particularly infrequent among younger speakers. She also finds that soft mutation is the 'strongest mutation', used more by speakers than either nasal mutation or aspirate mutation. This is a common theme in the literature on Welsh mutation, and in addition to its vitality relative to the other mutations, soft mutation is often seen to be displacing nasal and/or aspirate mutation in various contexts (cf. Thomas 1984).

While there have been no studies focusing specifically on mutation systems of Patagonian Welsh, Jones' pioneering (1984) survey of the Welsh in Gaiman, Patagonia mentions mutation. He notes that in casual speech (where speakers were “observed interacting with family, friends, and acquaintances and discussing a variety of topics from politics to motor racing, with the fieldworker and

others” (246)) no nasal mutation followed *yn* 'in', while the expected nasal mutation was found “sporadically” in formal speech (“interview situation questions and answers” (246)).

Nasal mutation in particular seems to be susceptible to a large amount of variation in production, and especially following the (potential) trigger *yn* 'in'. Of the various University of Wales theses on different varieties of Welsh cited in Ball & Müller (1992:238) which mention mutations following *yn*, three cite the standard nasal mutation as the usual realization (Evans 1930, Rees 1936, Griffiths 1974), two cite non-mutated forms (Davies 1955, Davies 1934), one reports exclusively soft mutation (Samuel 1970), and four report a mix of nasal mutation, soft mutation, and/or the radical unmutated form (Davies 1968, Bevan 1970, Thorne 1971, Roberts 1972). These studies are not primarily concerned with mutation, of course, and many only mention these tendencies in passing, without defining terms like 'often' or 'usual' (Ball & Müller 1992:239). Intriguingly, however, two of these studies (Evans 1930, Griffiths 1974) make specific mention of placenames, noting that while nasal mutation normally applies following *yn* in the variety under investigation, placenames did not participate in mutation. Given these varied findings, along with speaker intuitions and other observations in the literature that placenames may be subject to special conditions in mutation behaviour (cf. Ball & Müller 1992:205), they represent an important area for further study.

The current study aims to build on these findings by describing the mutation of placenames following *yn* 'in' in modern conversational Welsh in both Wales and Patagonia, in order to increase our understanding of Patagonian Welsh, of Welsh mutation paradigms in general, and of how geographic, linguistic, and cultural factors play into mutation variation in both Patagonia and Wales.

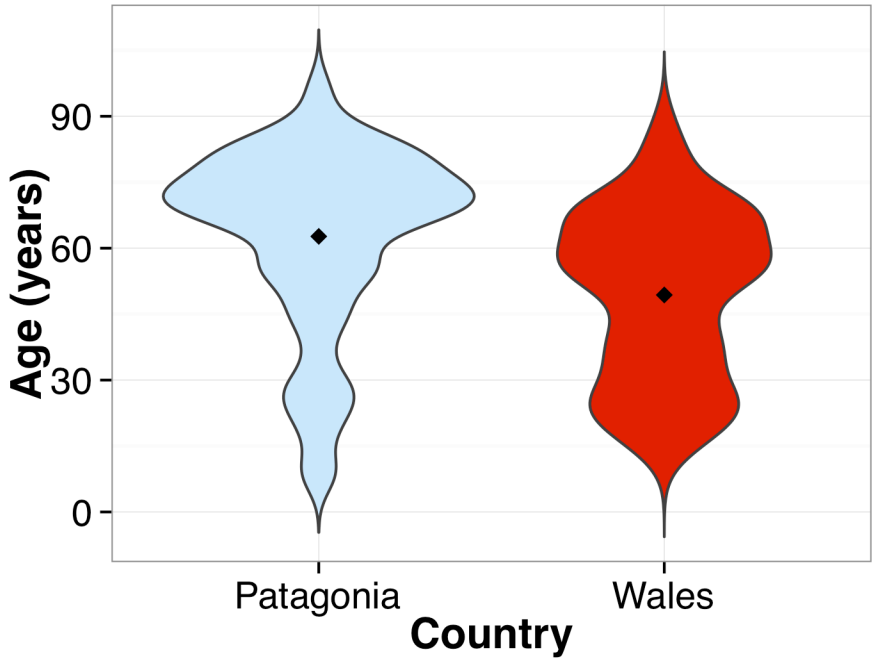
3. Data & Methods

3.1 Corpus data

The data for this study come from two corpora of conversational modern Welsh speech: the Wales data from the Bangor Siarad corpus, and the Patagonia data from the Bangor Patagonia corpus, both assembled by the ESRC Centre for Research on Bilingualism in Theory & Practice at Bangor University, Wales (Deuchar et al. 2014). Both corpora are licensed under the Free Software Foundation's General Public License and are freely available at http://www.bangortalk.org.uk. The corpora include both audio data and glossed transcriptions of conversations recorded in speakers' homes, workplaces, and schools, with 151 speakers and approximately 40 hours of conversation in the Bangor Siarad corpus, and 94 speakers and 21 hours of conversation in the Bangor Patagonia corpus. The Siarad corpus represents a geographical spread of speakers from across Wales, including participants brought up in Northwest, Northeast, Southwest, Southeast, and Mid Wales, along with speakers raised abroad and in more than one of these regions. The Patagonia corpus likewise represents speakers brought up in both the East and West of Chubut, among other origins. Age distributions of the speakers by country are shown in Figure 1 below, which illustrates number of speakers by age with relative width and marks the mean of each group (Wales = 49.4; Patagonia = 62.7)[1]:

1 Note that these data represent speakers who have tokens included in the current study, rather than all speakers in the corpora.

Figure 1: Speaker age distribution by country



Naturalistic corpus data are a particularly good fit for studying Welsh mutation variation. As Ball & Müller (1992:242) point out, studying mutation requires a relatively large amount of data, so that enough mutation triggers are present in order to observe potential mutations, but wordlists, elicitation, and other controlled tasks either involve written language – where the orthographic representation of mutations in Welsh compromises results aimed at spoken language – or significantly heightened formality and attention paid to speech (cf. Labov 1981:3, Eckert 2001:122). Since 'standard' paradigms of mutations are taught in Welsh (and Patagonian) schools, and mutations are a particularly salient feature of Welsh production (Ball & Müller 1992:264), laboratory settings may make speakers hyper-aware of their mutation behaviour and more likely to produce mutations they perceive as 'standard'.

The conversations in the Patagonia and Siarad corpora, on the other hand, in which participants recorded themselves without a researcher present (in most cases), represent a context where speakers

would presumably be less attentive to their speech than recording a wordlist in a laboratory setting, or even a traditional sociolinguistic interview. In the naturalistic speech of these corpus recordings, effects of speakers' awareness of being observed would be reduced due to both the longer duration of the recordings and the less formal nature of the conversations.

Tokens were located within text transcriptions of the data, by searching for all instances of the locative particle *yn* (or its variants *ym* or *yng,* via their shared gloss 'in.PREP') where the following word was a placename. From there, tokens were narrowed to placenames with mutable initial radical consonants (i.e. /p t k b d g/). Tokens like *yn Awstralia* 'in Australia' and *yn Ffrainc* 'in France', for instance, were present in the data but not included in analysis, because nasal mutation does not apply to their initial segments. Examples of tokens in their larger context are presented below, with placenames and their (potential) mutation triggers in bold:

Nasal mutation:

(2)

a	pan	oedd	o	yn	aros	hefo		
and	when	be.3S.IMPERF	3SM	PRT	wait.INFIN	with		
teulu	fy	mam	**ym**	**Metws-y-Coed**	…			
family.M.SG	1S.POSS	mother.SG	**in.PREP**	**Betws-y-Coed**	…			

"and when he stayed with my mother's family **in Betws-y-Coed** …"
(Siarad, Fusser4:205)[2]

(3)

dan	ni	yn	byw	**yn**	**Nhrelew**
be.1P.PRES	1P	PRT	live.INFIN	**in.PREP**	**Trelew**

"we live **in Trelew**"
(Patagonia, Patagonia10:178)

2 Examples are cited by corpus ('Siarad' (Wales) or 'Patagonia'), corpus file name, and line number.

(4)

mae	gynni	fancy	mynd	i	fyw
be.PRES	with.her	fancy	go.INFIN	to	live.INFIN+SM

ym	**Mharis**	am	gyfnod
in.PREP	**Paris**	for	period.M.SG+SM

"she fancies going to live **in Paris** for a while"
(Siarad, Stammers4:781)

Soft mutation:

(5)

dyna	ein	hanes	ni	fan	hyn
that.is	1P.POSS	history.M.SG	1P	place.SG+SM	this

yn	**Batagonia**
in.PREP	**Patagonia**

"that's our history here **in Patagonia**"
(Patagonia, Patagonia3:525)

(6)

mae	yna	ddwy	ffair	**yn**	**Griccieth**
be.3S.PRES	there	two.F+SM	fair.F.SG	**in.PREP**	**Criccieth**

"there are two fairs **in Criccieth**"
(Siarad, Davies10:805)

(7)

fi	ddarllen	o	mewn	llyfr	yn	llyfrgell
1S+SM	read.INFIN+SM	3SM	in.PREP	book.M.SG	in.PREP	library.M.SG

ddoe	**yn**	**Dreffynnon**
yesterday	**in.PREP**	**Treffynnon**(='Holywell')

"I read it in a book in the library yesterday, **in Holywell**"
(Siarad, Stammers8:227)

Non-mutated (radical) forms:

(8)

be	**yn**	**California**?
what.INT	**in.PREP**	**California**

"what, **in California**?"
(Siarad, Robert8:70)

(9)	mae	Siop Griffiths	wedi	symud	**yn**	**Porthmadog**	yndo
	be.3S.PRES	Siop Griffiths	after	move.INFIN	**in.PREP**	**Porthmadog**	Q.PST

"Siop Griffiths has moved **in Porthmadog**, hasn't it"
(Siarad, Davies11:64)

(10)	a	dach	chi	yn	mynd	i	gael
	and	be.2P.PRES	2P	PRT	go.INFIN	to.PREP	get.INFIN

cwrdd	yn	Gymraeg	**yn**	**Bryncrwn**?
meet.INFIN	in.PREP	Welsh.F.SG+SM	**in.PREP**	**Bryncrwn**

"and are you going to the Welsh service **in Bryncrwn**?"
(Patagonia, Patagonia14:215)

It's worth noting that although this coding relies on the orthographic representation of mutations present in the corpus – necessarily imposing categorical interpretations on a potentially phonetically gradient phenomenon – the native-speaker transcribers are not bound by the categories of the 'standard' paradigms in transcribing the mutations. In example (11) below, for instance, *yn Ngymru* 'in Wales' is transcribed with initial <ng>, indicating a voiced /ŋ/ onset. Voiced /ŋ/ is the prescriptively standard nasal mutation reflex of /g/, but not of /k/ – the initial segment of the radical form of *Cymru* 'Wales':

(11)	ond	ddoth	y	ffilm	yna
	but	come.3s.pst+SM	the	film.f.sg	there

allan	**yng**	**Ngymru**
out	**in.prep**	**Cymru(='Wales')**

"but that film came out **in Wales**"
(Patagonia, Patagonia22:807)

For placenames with radical-initial /k/ like *Cymru* 'Wales', the 'standard' mutation paradigms would allow for initial <g> /g/ in soft mutation (which itself would represent a non-'standard' mutation in this prescriptively nasally-mutating context) or initial <ngh> /ŋ̊/ in nasal mutation. The initial <ng> /ŋ/ indicated above shows that the transcribers are sensitive to the (potentially non-'standard') phonetic

productions of these mutations, and that while there may be gradience underlying these categorical distinctions, the tokens are classified in categories that are salient and meaningful to speakers.[3]

In total, 706 tokens were analyzed: 419 from Wales and 287 from Patagonia. These tokens represent every instance of a placename with mutable initial radical consonants following *yn* 'in' present in the two corpora. The data were drawn from drawn from a total of 187 speakers: 116 unique speakers from Wales and 71 unique speakers from Patagonia. Once isolated, each token was coded for the following features:

Table 2: Features coded

Feature	**Description**	**Levels**
SpeakerID	Numeric identifier	1-216
Country	Variety of Welsh spoken, determined by corpus	Patagonia Wales
Sex	Gender of speaker	Female Male
AgeYrs	Age of speaker in years	–
RadWord	Radical (non-mutated) version of word	–
RadCon	Radical (non-mutated) initial consonant of word	/p/ /t/ /k/ /b/ /d/ /g/
PlaceNameType	Type of placename	Welsh OutsideWelsh OutsideNonWelsh YWladfa
Realization	Realization of initial consonant mutation	Rad NM SM

3 Intriguingly, the instability of both /h/ and the voiceless nasals in Patagonian Welsh has been noted in Jones (1984:250), which could account for this particular realization; however, this token is the only example in the data of radical /p/ /t/ or /k/ realized with a voiced nasal.

For PlaceNameType, four different categories of placenames were distinguished: names of places in Wales (Welsh), places outside of Wales with specifically Welsh-language names (OutsideWelsh), places outside of Wales without Welsh-language names (OutsideNonWelsh), and places in the Wladfa, or Welsh-speaking Patagonia (YWladfa). Examples of each type from the data are provided in Table 3:

Table 3: Examples of PlaceNameType levels

Welsh	*Caerdydd* *Porthmadog* *Caergybi*	'Cardiff' 'Porthmadog' 'Holyhead'
OutsideWelsh	*De Affrica* *Brwsel* *Caeredin*	'South Africa' 'Brussels' 'Edinburgh'
OutisdeNonWelsh	*Toronto* *Dehli* *Papua New Guinea*	'Toronto' 'Dehli' 'Papua New Guinea'
YWladfa	*Trelew* *Porth Madryn* *Gaiman*	'Trelew' 'Porth Madryn' 'Gaiman'

These categories were decided upon following Ball & Müller's (1992:205) observation that “place-names from outside Wales generally only mutate if there is either a Welsh version of the name, or where the name is considered to be common enough to be brought into the system”, with YWladfa added as a separate category because of the Welsh-language origins of many of its placenames and its potentially different geographic importance for Patagonian speakers. The YWladfa category includes placenames of both Welsh-language origin (*Trelew*, *Porth Madryn*) and also of Tehuelche (*Gaiman*) and Spanish origin (*Patagonia*); this is in order to parallel the Welsh category (which includes placenames in Wales of non-Welsh origin, i.e. *Port Talbot*) in its geographical rather than strictly etymological basis, following Ball & Müller (1992:205).

The variable Realization indicates which mutation strategy was employed by the speaker when the placename was uttered following *yn*: Rad for no mutation (radical form), NM for nasal mutation,

and SM for soft mutation. A fourth possible category – AM for aspirate mutation – was originally included, but no tokens in the data made use of aspirate mutation.

3.2 Statistical analysis

Once coded, the data were analyzed in R (R Core Team 2015) using a two-step approach of a conditional inference tree and random forests, following Bernaisch et al. (2014). First, in order to show the variation present in these specific data, a conditional inference tree was generated from the dataset using the function *ctree* (Hothorn et al. 2006a) from the *party* package (Hothorn et al. 2006b). Conditional inference trees work by recursively partitioning the data; they attempt to best categorize different outcomes of a dependent variable (in this case, Realization) in terms of multiple, recursive binary splits of independent variables (here, Country, Sex, AgeYrs, RadCon, and PlaceNameType). This process continues to split the data until adding more divisions would no longer result in a significant increase in classification accuracy, and the end result is a tree which can be read like a decision-making flowchart, from top to bottom.

Since conditional inference trees are sometimes unstable, in that small changes in the data can lead to markedly different trees, and in order to insure both robustness and generalizability of the classification, a random forest was also generated from the same data, using the *cforest* (Strobl et al. 2007) implementation in the *party* package (Hothorn et al. 2006b). Random forests are based on classification trees, but – as the name suggests – their results come from an amalgamation of an extremely large number of such trees, and they expand on the base process by introducing two layers of randomness. First, the many trees in a random forest are created using differently bootstrapped samples of the data. Second, at each split, each tree is only able to select from a randomly-chosen subset of the predictors. These added layers of randomness help both to uncover the importance of variables or interactions that might otherwise go unnoticed, and to protect against overfitting (Deshors & Gries

2016). In fitting a random forest, the user is able to specify numbers for both of these randomness-inducing parameters – here, 1,500 trees were grown, and 3 predictors were sampled at each split. Since the results of a random forest are necessarily difficult to visualize, variable importance measures (computed using the *party* package (Hothorn et al. 2006b)) are used as a representative proxy.

Both the conditional inference tree and the random forest were given the same model structure as input, with Realization as the dependent variable and Country, Sex, AgeYrs, RadCon, and PlaceNameType as independent variables.

4. Statistical Results

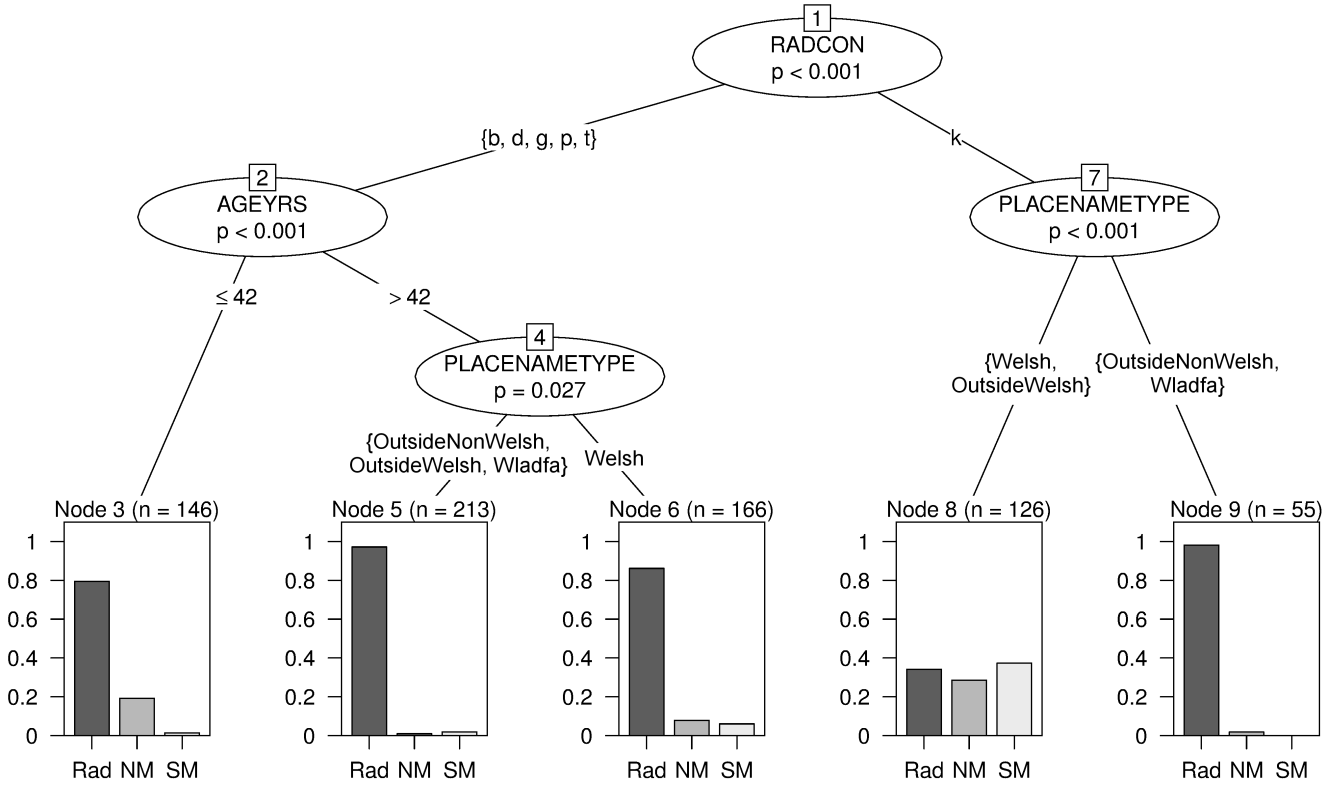
This section presents the statistical results of both the conditional inference tree and the random forest, which will be examined in detail in in terms of the research questions – (1) How do speakers in both Wales and Patagonia mutate (or do not mutate) placenames following the locative *yn* 'in' in modern, conversational Welsh?, and (2) What factors influence mutation in this context? – in section 5.

The results of the conditional inference tree are shown in Figure 2. To interpret the results, the diagram should be read from the top node (1) down. At each node, there is a left branch and a right branch, depending on the assumed value of the predictor named in the node: from node 1, for instance, if the RadCon (radical consonant) is any of /b d g p t/, the path continues left to node 2, but if the RadCon is a /k/, it continues right to node 7. This continues until a terminal node (at the bottom of the chart) is reached, which shows (a) the node number and how many tokens characterized by this particular combination of predictors are present in the data, and (b) a bar graph of observed percentages within that node of the three mutation Realization outcomes – Rad (radical/non-mutated), NM (nasal mutation), and SM (soft mutation).

As an example, the rightmost terminal node (9) shows that there are 55 tokens in the data with /k/ as an initial radical consonant (as indicated by the path taken from node 1 to node 7) and

PlaceNameTypes of either OutsideNonWelsh or YWladfa (node 7 to node 9), and that well over 90% of these tokens are unmutated (Rad), with only a very small percentage undergoing nasal mutation (NM).

Figure 2: Conditional inference tree



The results of the random forest, given in the form of variable importance measures, are as follows:

Table 4: Random forest variable importance, highest to lowest

RadCon	>>	**AgeYrs**	>	**PlaceNameType**	>>	**Sex**	>	**Country**
0.05048		0.02539		0.02318		0.00371		0.00326

These measures show how important each of these predictors are to the classification, with RadCon being the most important, AgeYrs and PlaceNameType each about half as important to the overall classification, and Sex and Country significantly less important.

Importantly, the results of the random forest validate the structure of the classification tree; the three most important variables in the random forest (RadCon, AgeYrs, and PlaceNameType, in

descending order) are also the three most important variables in the classification tree in the same order (with RadCon in the highest split in the tree, followed by AgeYrs and PlaceNameType on the same level), and together account for all splits seen in the tree. This means that the radical consonant is the most important factor in predicting nasal mutation variation, followed by speaker age and type of place name.

The predictive accuracy of the random forest (84%) is also robust; when compared to a baseline where predictions are made randomly but in proportion to the frequency of the levels of the dependent variable (66%), it represents a significant improvement ($P_{binomial\ test} < 10^{-25}$). This relatively small but significant effect – along with the fact that it already represents predictive rather than classificatory power – allows us to use the random forest as support for the structure of the classificatory tree.

Combined, the results of the conditional inference tree and the random forest reveal several interesting patterns that will be discussed in detail in section 5. In terms of general distribution, three points stand out: first, mutation is overall very rare in the data; second, initial /k/ patterns differently to the other initial consonants; and finally, speakers from Wales and Patagonia behave largely the same.

5. Discussion

First, and perhaps most strikingly, the results of the classification tree and random forest taken together show that there is no significant difference between how Patagonian speakers and speakers from Wales mutate placenames after *yn* in these data. The variable Country does not factor into the classification tree as a split, and it has the lowest amount of variable importance among all variables in the random forest model (0.00326), meaning that, on the whole, Welsh and Patagonian speakers both mutate placenames in the same way following *yn*.

In both Wales and Patagonia, the most important variables to emerge in predicting placename mutation behaviour following *yn* were RadCon, AgeYrs, and PlaceNameType. The initial radical

consonant (RadCon; variable importance = 0.05048) of the placename represents the highest split in the classification tree, with /k/ patterning differently than the rest of the mutable initial consonants (/p t b d g/) in terms of the other two variables: AgeYrs and PlaceNameType.

A speaker's age in years (AgeYrs; variable importance = 0.02539) accounts for one of the next splits in the classificatory tree, in node (2), where age makes a difference in mutating non-/k/-initial placenames: speakers aged 42 or younger are more likely to mutate than speakers over the age of 42[4]. This result is interesting because it seems to run contrary to what might be expected – for a minoritized language, we might normally expect younger speakers to use complex features like mutation less than older speakers – but one explanation for the result seen here could be due to language attitudes. Mutation is a particularly salient feature of Welsh, one immediately associated with the language and which speakers are keenly aware of (Ball & Müller 1992:264), and since the language rights and Welsh-language popular cultural movements of the 1960s, young speakers' attitudes towards Welsh in Wales have been increasingly positive (Llewellyn 2000, Williams 2009). Use of more mutations by younger speakers in Wales, then, could be part of marking a 'Welsh' identity and a positive alignment with this salient aspect of the language (Phillips 2007:179). In Patagonia, while attitudes towards the language also saw improvements in the 1960s following the 1965 centennial celebrations (Jones 1996; Johnson 2009), a marked increased in access to Welsh-medium education in the last several decades – especially since the founding of Prosiect yr Iaith Gymraeg in 1997 – could also be an influential factor here. Further, the discrepancy between this and previous studies on Welsh mutation with regards to younger speakers' mutation habits (i.e. Howell 2010) could be due to the focus here on placenames and in the particular environment of following *yn* – given that at least two studies have previously reported

4 Note that the variable AgeYrs was coded continuously, and the binary division here was selected as optimal by the conditional inference tree. Of the 116 speakers from Wales in the data, 56 were aged over 42 (resulting in 259 tokens), and 60 were aged 42 or younger (160 tokens). Of the 71 speakers from Patagonia, 55 were aged over 42 (resulting in 238 tokens), and 16 were 42 or younger (49 tokens).

Welsh varieties where placenames did not mutate following *yn* while other nouns did (Evans 1930, Griffiths 1974), mutation of placenames in this environment could actually be an innovation by younger speakers in some varieties.

Along with speaker age, the geographic and linguistic properties of the placename itself (PlaceNameType; variable importance = 0.02318) affect how speakers mutate. For placenames with initial radical /k/, placenames that were either Welsh or OutsideWelsh (i.e. located outside of Wales but with a specific Welsh-language name) were more likely to mutate than other placenames, confirming speaker intuitions about the 'Welshness' of placenames being important for mutation. In fact, these initial radical /k/ placenames were more likely to be mutated – either through nasal mutation or soft mutation – than not, representing the only node (8) in the classification tree where mutation was the most common outcome. Interestingly, placenames in the YWladfa category – places in Welsh-speaking Patagonia – here patterned with OutsideNonWelsh places, rather than Welsh and OutsideWelsh. This is perhaps surprising, especially given the Welsh-language origin of many placenames in Y Wladfa (like Trelew, Porth Madryn, and Dolafon), though it could also be a function of the relatively small number of /k/-initial placenames in the YWladfa category (n=10/165).

Similarly, PlaceNameType also had an effect on placenames with all other initial radical consonants (/p t b d g/) for speakers over the age of 42: here, placenames in the Welsh category were more likely to be mutated than OutsideWelsh, OutsideNonWelsh, or YWladfa placenames. Again, the fact that YWladfa placenames patterned with the less geographically and/or linguistically Welsh placenames (here OutsideWelsh and OutsideNonWelsh) is striking, and here there are significantly more tokens for non-/k/-initial YWladfa placenames (n=155/165). This means that YWladfa placenames seem to be conceptualized in terms of their geography relative to a Welsh reference point (i.e. 'outside of Wales'), rather than their linguistic features – where we might otherwise expect to see

the mostly Welsh-origin names in YWladfa (n=124/165) causing the category to be mutated like Welsh placenames – or sociocultural importance – where we might expect to see Patagonian speakers mutate them analogously to Welsh speakers with Welsh placenames. It is important to note that because of differences in data distributions between the two corpora, this particular result in terms of Wladfa placenames can only speak to the mutation behaviour of Patagonian speakers, rather than speakers from both countries; of the 419 tokens analyzed from the Siarad corpus, only five were Wladfa placenames[5]. The distribution can be seen in Table 5:

Table 5: Distribution of placename types by corpus

	Welsh	**YWladfa**	**OutsideWelsh**	**OutsideNonWelsh**	**Total**
Wales	82% (345)	1% (5)	3% (11)	14% (58)	419
Patagonia	9% (26)	56% (160)	0% (0)	35% (101)	287

Similarly, as no OutsideWelsh placenames appeared in the Patagonia corpus, these results cannot speak to the mutation behaviour of Patagonian speakers for OutsideWelsh placenames. What the results do show, however, is that for Welsh speakers from both Wales and Patagonia, Wales serves as a relative geographic reference that partially influences whether and how placenames mutate following *yn*, and that this is not the case with the Wladfa for (at least) Patagonian speakers.

As a result, the vast majority of tokens of YWladfa placenames (n=154/165) in the data are unmutated radical forms, like those in the following examples:

5 All five YWladfa tokens from the Siarad corpus were unmutated.

Radical-/g/-initial YWladfa placename, unmutated:

(12)	**yn**	**Gaiman**	ti	yn	mynd	i	yr	deintydd?
	in.PREP	**Gaiman**	2S	PRT	go.INFIN	to	the	dentist.M.SG

"do you go to the dentist **in Gaiman**?"
(Patagonia, Patagonia37:1109)

Radical-/k/-initial YWladfa placename, unmutated:

(13)	**yn**	**Cwm Hyfryd**
	in.PREP	**Cwm Hyfryd**

"**in Cwm Hyfryd**"
(Patagonia, Patagonia22:545)

Radical-/p/-initial YWladfa placename, unmutated:

(14)	mae	yna	un	**yn**	**Porth Madryn**
	be.3S.PRES	there	one	**in.PREP**	**Porth Madryn**

"there's one **in Porth Madryn**"
(Patagonia, Patagonia4:396)

In addition to the effects of radical consonant, speaker age, and placename type on mutation, one salient aspect of Welsh mutation variation that has emerged from these results is the importance of soft mutation in the paradigm following *yn*. The use of soft mutation as an alternative to nasal mutation has been previously mentioned in the literature, but its extent in these data is striking: for placenames with initial radical /k/, soft mutation is a more common realization than nasal mutation, and for Welsh and OutsideWelsh placenames with initial radical /k/, soft mutation – as in examples (15) and (16) below – is actually the most common realization, even moreso than radical (non-mutated) realizations:

Radical-/k/-initial Welsh placename, soft mutation

(15)

mae		gyn	ti	er	gored	bach
be.3S.PRES		with	2S	er	weir	small
nice	iawn	**yn**		**Glynnog**	de	
nice	very	**in.PREP**		**Clynnog**	be.IM+SM	

"you've got a really nice little weir **in Clynnog**, right"
(Siarad, Fusser9:109)

OutsideWelsh radical-/k/-initial placename, soft mutation

(16)

oh	ryw	boutique	**yn**	**Gaer**
oh	some+SM	boutique	**in.PREP**	**Caer(='Chester')**
dw	ddim	yn	gwybod	
be.1S.PRES	not+SM	in.PREP	know.INFIN	

"oh some boutique **in Chester**, I don't know"
(Siarad, Fusser30:434)

There could be articulatory reasons for this preference. Since, in the case of /k/, soft mutation involves the addition of voicing, and these placenames all follow the fully-voiced locative particle *yn* /ən/, soft mutation here represents the sustaining of intervocalic voicing (as in *yn Gaerdydd* /ən gairˈdiːð/ 'in Cardiff') – articulatorily easier than either the radical version (y*n Caerdydd* /ən kairˈdiːð/ 'in Cardiff'), with a voiceless /k/ between a voiced nasal and a voiced vowel, or the nasal mutation (y*ng Nghaerdydd* /əŋ ŋ̊airˈdiːð/ 'in Cardiff'), with a voiceless /ŋ̊/ in the same environment (cf. Westbury & Keating 1986). Of course, this would also hold true for placenames beginning with /p/ and /t/, and while it is possible that the longer articulation duration of velar stops relative to alveolar and bilabial equivalents (cf. Lisker & Abramson 1964) plays some part in distinguishing /k/ from the other two Welsh voiceless stops, the nature of the effect seen here could also be due to particularities with the current corpus data.

The preponderance of soft mutation may be due at least in part, for instance, to the fact that some specific placenames (most of them /k/-initial) seem especially prone to soft mutation.

Caernarfon, for instance, is realized with soft mutation as *yn Gaernarfon* 20 out of the 23 times it appears in the data, while *Caerdydd* 'Cardiff' is realized as *yn Gaerdydd* 16 out of 44 times:

Caernarfon, soft mutation

(17)

ar	y	bike	**yn**	**Gaernarfon**
on	the	bike	**in.PREP**	**Caernarfon**

"on the bike **in Caernarfon**"
(Siarad, Davies3:48)

Caerdydd ('Cardiff'), soft mutation

(18)

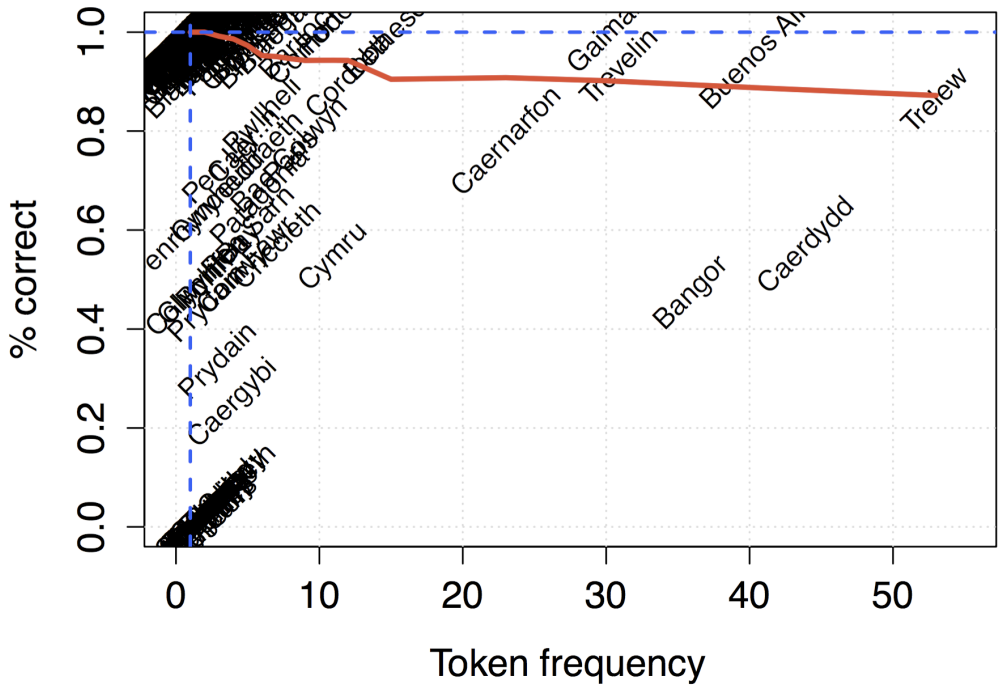
wedyn	wnaeth	y	genod	**yn**	**Gaerdydd**
afterwards	do.3S.PST+SM	the	girls	**in.PREP**	**Caerdydd(='Cardiff')**

prynu	yr	un	away	…
buy.INFIN	the	one	away	…

"then the girls **in Cardiff** bought the away [rugby shirt] … "
(Siarad, Robert4:780)

These two localities are also part of a small group of placenames that seem to stand out in terms of the random forest model. Figure 3 shows individual radical placenames (RadWord) in terms of their percentage of correct predictions in the model on the y-axis by token frequency on the x-axis, with the dashed blue lines representing the median values for each axis and the red line a locally-weighted smoother summarizing the correlation between the two plotted variables. Four placenames – *Cymru* 'Wales', *Caernarfon*, *Bangor*, and *Caerdydd* 'Cardiff' – stand out as having relatively high token frequencies, but relatively low prediction accuracy in the model:

Figure 3: Placenames by token frequency and percentage correct predictions



For each of these placenames, speakers mutate more than the model would predict. Mutation is actually a relatively rare outcome in most of the data, and each of these placenames is produced in mutated forms (whether through nasal or soft mutation) more than their radical forms, as shown in Table 6:

Table 6: Outlying placenames by mutation realization

	Rad	**NM**	**SM**	**Total**
***Cymru* 'Wales'**	3	7	1	11
Caernarfon	3	0	20	23
Bangor	17	14	5	36
***Caerdydd* 'Cardiff'**	7	21	16	44

Importantly, these placenames all have something in common: they each name places of significant cultural importance for Welsh speakers. Bangor and Caernarfon are the first and second most populous towns, respectively, in Gwynedd – the area of Wales with the single highest percentage of Welsh speakers (Welsh Language Commissioner 2011a). Bangor is home to Prifysgol Bangor / Bangor University, the university with the most students studying through the medium of Welsh in the country

(Prifysgol Bangor University 2011), and Caernarfon is a hub of Welsh-speaking culture (Williams 2009), with the highest percentage (85.6%) of Welsh-speakers in any locality in Wales (Welsh Language Commissioner 2011b). *Caerdydd* 'Cardiff' is, of course, the capital of and largest city in Wales, and is home to a growing body of Welsh-language media, including the Welsh-language television channel S4C, as well as important venues for Welsh-language musical performances like Clwb Ifor Bach. *Cymru* 'Wales', meanwhile, is iconic in its representation of Welsh-speaking Wales, and is familiar as a signifier of national pride even to non-speakers.

The fact that these four placenames mutate more than the model would expect – even within /k/-initial Welsh placenames, part of the most mutated group – shows that just as geography matters in mutation, so does culture. It also shows that certain placenames may be lexically specified for certain mutation behaviours. Caernarfon, for instance, appeared with soft mutation as *yn Gaernarfon* 20/23 times, and never underwent nasal mutation; given these data, soft mutation – as in (20) below – seems to be the 'default' realization of Caernarfon following *yn*[6]:

(20)	yeah	ges	i	erioed	ddrink
	yeah	get.1S.PST+SM	1S	never	drink+SM
	am	ddim	**yn**	**Gaernarfon**	
	for	nothing.M.SG+SM	**in.PREP**	**Caernarfon**	

"yeah, I never got a drink for free **in Caernarfon**"
(Siarad, Fusser10:505)

6. Conclusions

The results of this study show, first, that there is no statistically significant difference in the way Patagonian speakers and speakers from Wales mutate placenames following the locative *yn* 'in' in these data. This result adds not only to our understanding of understudied Patagonian Welsh as a variety, but

6 Prys (2016:284) points out that *Y Gaernarfon* is an attested historical form of *Caernarfon*, which may play a role in the high frequency of its soft mutated form following *yn* in the data.

also of how Welsh is and is not affected by contact with Spanish and the sociocultural context of the Welsh Patagonia; while on the one hand Patagonian speakers show shorter VOT values (Sleeper 2015), on the other, they retain the same phonemic vowel length contrasts (Bell 2015) and the same mutation paradigm for placenames following *yn* (as per the current study) as speakers from Wales.

For speakers from both countries, we find the following: that younger speakers (≤ 42) are more likely to mutate non-/k/-initial placenames than older (> 42) speakers; and that geographic and cultural location matters – placenames that are more geographically or culturally 'Welsh' are more likely to be mutated, and even within Wales, names of places with special Welsh-language cultural significance – like *Cymru* 'Wales', *Caernarfon*, *Bangor*, and *Caerdydd* 'Cardiff' – undergo mutation more often than predicted by the random forest model based on the rest of the data.

Importantly, for Patagonian speakers, placenames of the Wladfa – including those with Welsh-language origins like Trelew and Porth Madryn – patterned the same way as geographically and/or linguistically 'less Welsh' placenames. This suggests that while Wales serves as a geographic and cultural locus for both groups of speakers in terms of mutation, speakers from the Wladfa do not seem to treat Wladfa placenames in the same way speakers from Wales treat Welsh placenames for mutation purposes. These results point towards the importance of further research – both linguistic and ethnographic – on language and place in modern Welsh Patagonia.

In addition to the importance of placename type, initial radical consonant, and age of the speaker in Welsh mutation behaviour, these results also highlight the role of soft mutation in the nominally nasally mutating environment following *yn*. King (2003:17) and Thomas (1984:214) both mention that soft mutation may be displacing nasal mutation in various contexts, and these results provides strong support for the idea that soft mutation is becoming an important part of the mutation paradigm following the locative *yn* in the spoken language, particularly with initial radical /k/. For placenames beginning with radical /k/, realizations with soft mutation are more common than the

prescribed nasal mutation, and for Welsh placenames (and, for speakers from Wales, places situated geographically outside of Wales but with specifically Welsh-language names), soft mutation is actually the single most common realization.

The implications of this development on Welsh phonology are important. Unlike the other typologically rare phonemes in Welsh (/ɬ/ and /r̥/), the voiceless nasals /m̥ n̥ ŋ̊/ are considered phonemic only by virtue of their inclusion in nasal mutation, and if nasal mutation eventually ceases to be applied, they will cease to create phonemically meaningful contrasts in the language (Watkins 1993). Of course, the locative *yn* is not the only trigger of nasal mutation in Welsh, and while the present study has focused on placenames, future research should also use conversational corpus data to explore mutation variation for words following the other two nasal mutation triggers: the first person possessive pronoun *fy* 'my' and the negative prefix *an-* 'un-'.

Finally, since the results of this study show that geographic and cultural location of placenames affect their participation in mutation following *yn*, future research should also examine whether this remains true in other mutable contexts, especially following *i* 'to', which is said to trigger soft mutation (i.e. *i Gaernarfon* 'to Caernarfon'). Since soft mutation has elsewhere been shown to be both more vital than other mutations and also expanding into new environments (Howell 2010, Thomas 1984, Prys 2016) – and especially given its importance in the results of the current study – soft mutation environments present a natural next step in which to further examine the importance of geographic, linguistic, and cultural factors in placename mutation.

References:

Agozzino, Maria Teresa. 2006. Transplanted traditions: An assessment of Welsh lore and language in Argentina. *e-Keltoi Journal of Interdisciplinary Celtic Studies* 1, 39–67.

Arwel, Rhisiart. 2016. *Welsh Language Project in Chubut: Annual Report 2016*. Cardiff: British Council.

Ball, Martin J. 1985. Radio Cymru: Programme style and linguistic variation. *Journal of Multilingual and Multicultural Development*, 6(2), 157-163.

Ball, Martin J. & Nicole Müller. 1992. *Mutation in Welsh*. New York: Routledge.

Birt, Paul. 2005. The Welsh Langauge in Chubut Province, Argentina. In Diarmuid Ó Néill (ed.), *Rebuilding the Celtic Languages: Reversing Language Shift in the Celtic Countries.* Talybont: Y Lolfa.

Bell, Elise. 2015. *Phonemic vowel length in Welsh-Spanish bilingual speech.* University of Arizona: qualifying paper.

Bernaisch, Tobias, Stefan Th. Gries, and Joybrato Mukherjee. 2014. The dative alternation in South Asian English(es). *English World-Wide, 35*(1), 7-31.

Bevan, J. 1970. *Astudiaeth Seinyddol o Gymraeg Llafar Coety Walia a Rhuthun ym Mro Morgannwg*. University of Wales: unpublished MA Thesis.

Davies, Evan J. 1955. *Astudiaeth Gymharol o Dafodieithoedd Dihewyd a Llandygwydd*. University of Wales: unpublished MA Thesis.

Davies, J.J. Glanmor. 1934. *Astudiaeth o'r Gymraeg llafar Ardal Ceinewydd.* University of Wales: unpublished MA Thesis.

Davies, L. 1968. *Astudiaeth seinyddol gan gynnwys geirfa o dafodiaith Merthyr Tudful a'r Cylch*. University of Wales: unpublished MA Thesis.

Deshors, Sandra C. and Stefan Th. Gries. 2016. Profiling verb complementation constructions across New Englishes. *International Journal of Corpus Linguistics*, 21(2), 192-218.

Deuchar, Margaret, P. Peredur Davies, Jon Russell Herring, M. Carmen Parafita Couto & D. Carter. 2014. Building bilingual corpora. In Enlli Môn Thomas & Ineke Mennen (eds.), *Advances in the Study of Bilingualism.* Bristol: Multilingual Matters.

Eckert, Penelope. 2001. Style and Social Meaning. In Penelope Eckert & John R. Rickford (eds.), *Style and sociolinguistic variation*. Cambridge University Press, 2001.

ESRC Centre for Research on Bilingualism in Theory & Practice. *Bangor Patagonia Corpus*. http://bangortalk.org.uk/. Bangor: University of Wales.

ESRC Centre for Research on Bilingualism in Theory & Practice. *Bangor Siarad Corpus*. http://bangortalk.org.uk/. Bangor: University of Wales.

Evans, D.G. 1930. *Tafodiaith Cwmtawe*. University of Wales: unpublished MA Thesis.

Gnanadesikan, Amalia E. 1997. Phonology with ternary scales. Ph.D. dissertation, University of Massachusetts.

Gries, Stefan Th. 2013. *Statistics For Linguistics with R.* Berlin: De Gruyter Mouton.

Griffiths, D.W. 1974. *Astudiaeth eirfaol o Gymraeg llafar Llanfair Caereinion*. University of Wales: unpublished MA Thesis.

Hannahs, S.J. 2013. *The Phonology of Welsh*. London: Oxford University Press.

Hothorn, Torsten, Kurt Hornik, C. Strobl, and A. Zeileis. 2006a. "Party: a laboratory for recursive part(y)itioning (2006)." *R package version 0.9-999.* http://CRAN.R-project.org/package=party.

Hothorn, Torsten, Kurt Hornik and Achim Zeileis. 2006b. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15(3), 651–674.

Howell, Sara. 2010. *A Sociolinguistic investigation into the use of mutations in Pembrokeshire Welsh.* Cardiff Metropolitan University: PhD thesis.

Johnson, Ian. 2009. How green is their valley? Subjective vitality of Welsh language and culture in the Chubut Province, Argentina. *International Journal of the Sociology of Language*. 195, 141-171.

Jones, Robert O. 1984. Change and Variation in the Welsh of Gaiman, Chubut. In Martin J. Ball and Glyn E. Jones (eds.), *Welsh Phonology*. Cardiff: University of Wales Press. 237-261.

Jones, Robert O. 1996. *A report on the Welsh language in Argentina's Chubut province 1996: Adroddiad ar yr iaith Gymraeg yn nhalatih Chubut, yr Ariannin 1996*. Cardiff: British Council.

Jones, Robert O. 1998. The Welsh language in Patagonia. In Geraint Jenkins (ed.), *A social history of the Welsh language: Language and community*. Cardiff: University of Wales Press. 287-316.

Kiff, Gareth. 2013. Welsh Language Project in Chubut: Annual Report 2012. Cardiff: British Council.

King, Gareth. 2003. *Modern Welsh: A Comprehensive Grammar*. New York: Routledge.

Labov, William. 1981. *Field methods of the project on linguistic change and variation*. Southwest Educational Development Laboratory.

Lieber, Rochelle. 1987. *An Integrated Theory of Autosegmental Processes*. Albany: SUNY Press.

Lisker, Leigh & Arthur S. Abramson. 1964. Cross-language study of voicing in initial stops: acoustical measurements. *Word* 20, 384-422.

Llewellyn, Meic. 2000. Popular Music in the Welsh Language and the Affirmation of Youth Identities. *Popular Music*, 19(3), 319-339.

Morris-Jones, John. 1913. *A Welsh Grammar, Historical and Comparative.* Oxford: Clarendon Press.

Ó Néill, Diarmuid. 2005. What is the Way Forward? In Diarmuid Ó Néill (ed.), *Rebuilding the Celtic languages: Reversing language shift in the Celtic countries*, 424–434. Talybont: Y Lolfa.

Phillips, John D. 2007. Rare and endangered linguistic subsystems in Celtic and Welsh. *Linguistica e Filologia*, 24. 153-182.

Prifysgol Bangor University, 2012. "Support for Welsh-speaking students and those studying through the medium of Welsh". Prifysgol Bangor University, Bangor, Wales. https://www.bangor.ac.uk/canolfanbedwyr/pdf/supporting_students_12-13.pdf Accessed 02.03.2017.

Prys, Myfyr. 2016. *Stye in the vernacular and on the radio: code-switching and mutation as stylistic and social markers in Welsh*. Unpublished PhD thesis: Bangor University.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rees, R.O. 1936. *Gramadeg tafodiaith Dyffryn Aman*. University of Wales: unpublished MA Thesis.

Roberts, A.E. 1972. *Geirfa a ffurfiau Cymraeg llafar cylch Pwllheli*. University of Wales: unpublished MA Thesis.

Samuel, O. 1970. *Astudiaeth o Gymraeg cylch Y Rhigos*. University of Wales: unpublished MA Thesis.

Sleeper, Morgan. 2015. Contact effects on voice-onset time (VOT) in Patagonian Welsh. University of California, Santa Barbara: MA Thesis.

Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis and Torsten Hothorn. 2007. Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC Bioinformatics*, 8(25). URL http://www.biomedcentral.com/1471-2105/8/25.

Sweet, H. 1884. Spoken North Welsh. *Transactions of the Philological Society*, 1882-4, 409-84.

Thomas, Peter Wynn. 1984. Variation in South Glamorgan Consonant Variation. In Martin J. Ball & Glyn E. Jones (eds.), *Welsh Phonology*. Cardiff: Cardiff University Press. 208-236.

Thorne, D. 1971. *Astudiaeth seinyddol a morffolegol o dafodiaith Llangennech gan gynnwys geirfa*. University of Wales: unpublished MA Thesis.

Watkins, T. Arwyn. 1993. Welsh. In Martin J. Ball & James Fife (eds.), *The Celtic Languages*. London: Routledge. 298-348.

Westbury, John R. & Patricia A. Keating. On the Naturalness of Stop Consonant Voicing. *Journal of Linguistics*, 22(1), 145-166.

Welsh Language Commissioner, 2011a. "2011 Census results by Local Authority". http://www.comisiynyddygymraeg.cymru/English/Policy,%20research%20and%20data/Census%20Data/Pages/2011Censusresultsbylocalauthority.aspx. Accessed 02.03.2017.

Welsh Language Commissioner, 2011b. "2011 Census results by Community". http://www.comisiynyddygymraeg.cymru/English/Policy,%20research%20and%20data/Census%20Data/Pages/2011CensusresultsbyCommunity.aspx. Accessed 02.03.2017.

Wickham, Hadley. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Williams, Glyn. 1991. *The Welsh in Patagonia: The State and the Ethnic Community.* Cardiff: University of Wales Press.

Williams, Eddie. 2009. Language attitudes and identity in a North Wales town: "Something different about Caernarfon"?. *International Journal of the Sociology of* Language, 2009(195), 63-91.

Willis, Penny. 1986. *The Initial Consonant Mutations in Welsh and Breton*. Bloomington: Indiana University Linguistics Club.