Heritage Welsh and the Bilingual Continuum: Grammatical Gender Under Incomplete Acquisition[1]

Erin D. Boon

The term *heritage language* (HL) refers to the proficiency of a speaker with a particular language background. The label connotes a cultural connection between the speaker and the language, but, in the narrow sense used here, also indicates some level of functional proficiency in that language. The heritage speaker is both, and neither, a native and non-native speaker. The HL was the first language (L1) chronologically, but the acquisition process was curtailed during childhood. Consequently, that L1 never reached full native speaker proficiency. The result is an imbalanced bilingual whose second language (L2) has become her dominant language. This bilingual state is closely connected to the phenomenon of language shift, not only in the context of immigration, but also among young speakers of minority languages. In the Celtic countries, these speakers have been referred to most frequently as 'semi-speakers',[2] though I hope to make a case for a revision of that terminology here.

A description of the grammatical consequences of the heritage speaker profile, in which the HL is only partially acquired in childhood, is one of the principal objectives of the HL research agenda. Speaking broadly, the HL grammar is strong in phonology and weak in morphosyntax.[3] Because of the early exposure to the HL, heritage speakers generally have near-native command of the HL sound system, manifesting as native-like accent and intonation patterns. Heritage speakers of Spanish living in the Los Angeles, California, area exhibit a native-like mastery of the stress and intonation patterns of Spanish.[4] Heritage speakers of Korean, likewise, have been found to possess a clear advantage over adult learners in both the perception and production of Korean phonemes.[5] Fully native-like phonology is not always characteristic of the heritage speaker, however, as was found in a study of first and second generation Italian immigrants in San Francisco, California. For

---

1   This paper was presented in an earlier form at the 36th University of California Celtic Studies Conference, held at UCLA in March 2014. I want to thank those who were in attendance for their comments and suggestions.

2   Most notably in the work of Nancy Dorian. Cf. Nancy Dorian, 'The problem of the semi-speaker in language death', *International Journal of the Sociology of Language,* 12 (1977), 23–32; Nancy Dorian, 'Language shift in the community and individual: The phenomenon of the laggard semi-speaker', *International Journal of the Sociology of Language,* 25 (1980), 85–94.

3   'Morphosyntax' incorporates both the syntax of the language, including features such as word order and sentence structure, as well as morphology, which in Welsh includes verb-subject agreement, initial consonant mutations, and plural suffixes, for example.

4   Claudia Parodi, 'Stigmatized Spanish Inside the Classroom and Out: A Model of Language Teaching to Heritage Speakers,' in *Heritage Language Education: A New Field Emerging,* ed. by Donna M. Brinton, Olga Kagan, and Susan Bauckus (New York: Routledge, 2008), pp.199–214.

5   Janet S. Oh, Sun-Ah Jun, Leah M. Knightly and Terry Kit-fong Au, 'Holding on to childhood language memory', *Cognition*, 86, B53–B64.

some of these heritage speakers, the perception of geminate consonants in their HL, Lucchese, was lost.[6] While it is clear that early exposure to the HL confers some amount of phonological intuition to the heritage speaker, it is not always enough to ensure an entirely native-like outcome.

Morphosyntactic characteristics of the heritage grammar are similarly variable across speakers, and to a degree even greater than phonological characteristics. In a study comparing heritage speakers of Spanish with fully proficient native speakers and adult L2 learners, the heritage speakers varied in their performance on a grammaticality judgment task based on whether they spoke the HL as children or merely overheard it.[7] The participants listened to sixty-six sentences, half of which contained an error in negative marking, tense agreement, person agreement, gender agreement, number agreement, or indirect object marking. They were then asked to indicate whether the sentence was grammatically acceptable. Those heritage speakers who had actively used the language in childhood performed worse than native speakers, but better than typical late L2 learners and, importantly, better than heritage speakers who merely had passive exposure to Spanish as children. The 'childhood over-hearers,' whose only early exposure was in the form of listening to the HL spoken around them, though never producing it themselves, did not show any advantage over the adult L2 learners on this grammaticality judgment task. This is a particularly interesting finding given the fact that these over-hearers had been found previously to have a phonological advantage over adult L2 learners.[8] Between these two components of the heritage grammar, phonology and morphosyntax, it seems to be morphosyntax which is most affected by the interruption in exposure to the HL during the acquisition process. Measuring the variability in language proficiency among speakers with different language backgrounds is one of the principal components of the HL research agenda, but the acquisition of phonology seems to be more likely to reach a native-like outcome than does morphosyntax when the acquisition process is curtailed. This is remarkable when compared to the most common outcome in L2 acquisition, in which an adult learner may speak with impressively native-like grammar and vocabulary yet still give himself away with a non-native accent.

Closely tied to considerations of imbalanced bilingualism is the nature of bilingualism itself. At what point do we consider an individual to be bilingual? The current theoretical model represents bilingual speakers along a continuum of proficiencies which is quite inclusive. The perfectly balanced bilingual is a purely hypothetical construct. Any imbalance, therefore, in a bilingual's language system must not be seen as a factor which limits her inclusion along this continuum. In describing the variable Spanish proficiencies of a bilingual community in East Los Angeles, California, Carmen Silva-Corvalán defines this 'bilingual continuum.'

6   Chiara Celata and Jessica Cancila, 'Phonological attrition and the perception of geminate consonants in the Lucchese community of San Francisco (CA)', *International Journal of Bilingualism,* 14.2 (2010), 185–209.

7   Terry Kit-fong Au, Janet S. Oh, Leah M. Knightly, Sun-Ah Jun, and Laura F. Romo, 'Salvaging a childhood language', *Journal of Memory and Language*, 58 (2008), 998–1011.

8   Terry Kit-fong Au, Leah M. Knightly, Sun-Ah Jun, and Janet S. Oh. 'Overhearing a language during childhood,' *Psychological Science*, 13 (2002), 238–243.

> The complexity of this community accounts for the existence of
> what I call a bilingual continuum, similar to a creole continuum in
> that one may identify a series of lects ranging from full-fledged to
> emblematic Spanish and, vice versa, from full-fledged to emblematic
> English depending on whether the bilingual is more or less dominant
> in Spanish or English. Even further, these individual lects do not
> correspond to fixed dichotomies of the type 'compound-coordinate,'
> or 'balanced-imbalanced.' Rather, at the individual level they
> represent a wide range of dynamic levels of proficiency in the
> subordinate language, i.e., it is in principle possible for an individual
> to move or be moving toward one end or the other of the continuum
> at any given synchronic stage of his life. [9]

Importantly, the speaker is still considered to be bilingual when her weaker
language abilities are very rudimentary. While this statement does not necessarily
represent the practical reality for the speaker, as, of course, effective communication
may not be achievable with such rudimentary proficiency, the continuum model does
avoid the theoretical complications of setting a standard of language proficiency
which must be reached in order for a speaker to qualify as 'bilingual.' That
standard, inevitably, would be arbitrary. This trend toward greater inclusion under
the term 'bilingual' is part of what I am here considering in the context of heritage
speakers. Differences between speakers, either in their language experiences or
resulting abilities, do not preclude inclusion on the bilingual continuum. To limit
the type and level of proficiency necessary to be considered a bilingual speaker
would privilege certain kinds of language experience over others, and insufficiently
represent the profile of the entire language community, particularly in the case of a
language such as Welsh.

The dynamism of bilingualism as it is currently described – i.e., the emergent
properties of a mental system where two languages are in contact and where
the speaker may vary in how much she uses one or the other over time – clearly
leaves room for the inclusion of the heritage speaker. The dominance of her two
languages effectively swapped places at some point in childhood, where the L1,
heritage Welsh, switched from being the dominant to the weaker language,[10]
and English acquisition proceeded to full proficiency. Both languages persist,
though to varying degrees. We must be allowed to categorize heritage speakers as
bilinguals because they clearly fit onto the bilingual continuum, and therefore also
account for them in any figures that are supposed to represent the total number of

---

9   Carmen Silva-Corvalán, 'Spanish language attrition in a contact situation with English',
    in *First Language Attrition,* ed. by Herbert W. Seliger and Robert M. Vago (Cambridge:
    Cambridge University Press), pp. 151–171 (p.151).

10   A bilingual may also be considered a heritage Welsh speaker if she acquired both English and
    Welsh *simultaneously* during early childhood (i.e., she has two L1s), though perhaps as an
    adult she does not recall which was the stronger language at that time. If her acquisition of
    English outpaced that of Welsh at some point in childhood, Welsh is still considered the HL in
    that situation.

Welsh speakers. Survey data already include second language learners as Welsh speakers. In fact, it is a well-known and controversial issue that school-age learners are included in counts of Welsh speakers.[11] As long as the data that are gathered effectively categorize the abilities of these respondents, there is no reason to discourage such positive responses to the Welsh question. The same can be said for heritage speakers. The goal in assessing the language abilities of a population is to gather information about every individual's proficiency, however slight, rather than to discourage the inclusion of those who are less confident in their fluency. Acknowledging that heritage speakers are Welsh-English bilinguals as well would only lead to greater accuracy in the account of Welsh language abilities in the population. This is certainly not a revolutionary suggestion inside Wales, where variability among speakers is encountered daily, but heritage Welsh speakers are not only to be found in Wales.

Information about the prevalence of the Welsh language is only collected on Census forms that are distributed in Wales. In England, the language question asks only about the respondent's primary language and not specifically about Welsh-English bilingualism. This is particularly problematic for an account of heritage speakers because England is where the majority of Welsh-speaking emigrants live. The Office of the Welsh Language Commissioner estimates that 110,000 speakers of Welsh actually live in England, and that figure almost certainly does not include the least proficient of heritage speakers. Along with a further 1,000 speakers in Scotland and Northern Ireland, there are roughly 690,000 Welsh speakers in the UK, though 17% of them live outside of Wales.[12] The exclusion of these speakers in regular Census counts is of particular significance to the heritage language issue because, outside of Wales, Welsh speakers are more likely to be heritage speakers than adult learners or, I would argue, fully proficient native speakers. The full spectrum of Welsh proficiency should be accounted for if the goal is an accurate reckoning of the prevalence of Welsh today – all speakers on the continuum, both in Wales and elsewhere in the UK.

The implications of the heritage speaker profile and its particular grammatical characteristics extend into the realm of language pedagogy as well. HL abilities in the foreign language classroom are both an advantage and a disadvantage to the student. If the heritage speaker is placed in a beginner's level class, her previous experience with the language can be very helpful, of course. These abilities are a serious disadvantage to the speaker, however, if they result in the student being assessed at a higher proficiency level than she truly is. The heritage speaker skill set cannot be measured in the same way as that of a second language learner because of the nature of her exposure to the language. Their respective imperfect grammars may be equally non-native, but where the adult learner excels at written and metalinguistic tasks, the heritage speaker struggles, and where the heritage speaker

11   Cf. Delyth Morris, 'Young people and their use of the Welsh language', in *Welsh in the Twenty-First Century,* ed. by Delyth Morris (Cardiff: University of Wales Press), pp. 80–98.
12   Hywel M. Jones, 'A Statistical Overview of the Welsh Language', in *Comisiynydd y Gymraeg/ Welsh Language Commissioner* (n.d.), <http://www.comisiynyddygymraeg.org/English/ Assistance/Dataandstatisitcs/Pages/StatisticaloverviewoftheWelshlanguage.aspx> [accessed 5 July 2014].

is comfortable with casual conversation and fluid speech in informal registers, the adult learner speaks with marked grammar and vocabulary. While the two types of student may thrive in a shared classroom, it is undeniable that they have different learning needs. This paper will describe in detail one of the characteristics of heritage Welsh which may require dedicated focus in the classroom, a reduced system of grammatical gender, following from the typical HL weakness in morphosyntax.

Beyond the broad contrast of phonological strengths and morphosyntactic weaknesses, careful investigations into heritage Welsh can show precisely how the heritage speaker grammar diverges from native speaker norms. One might question whether there is a single heritage Welsh speaker grammatical profile, given the variability in the personal backgrounds of heritage Welsh speakers, but this is one of the questions that crosslinguistic HL research has put to rest. The description of a heritage language, whether it be heritage Spanish, heritage Korean, or heritage Welsh, is a valid and operable construct because of the nature of human language in general. The reduced L1 acquisition process common within each group of heritage speakers will inevitably lead to the same HL outcomes, varying only with the amount and quality of exposure to the language received during childhood. The components of a grammar emerge in the developing system in predictable patterns, where the likelihood of some features can be determined based on the presence of others.[13]

A study of heritage Russian[14] found that the universal accessibility hierarchy for the formation of relative clauses, already known to accurately generalize the pattern of relative clause formation found crosslinguistically, also applies to the likelihood of the various kinds of relative clauses appearing in HL. The Accessibility Hierarchy captures the crosslinguistic generalization that if a language is able to relativize at a given position on the hierarchy then it will be able to at every higher position as well. For example, if a language only allows one position to relativize, it will be the subject; if it allows the indirect object to relativize, then it will also allow the direct object and subject.[15] Polinsky found that the adult heritage Russian speakers of her study struggled to comprehend object relative clauses more than they struggled with subject relative clauses, despite the fact that the dominant language of the heritage speakers was English, a language which allows object relative clauses. The conclusion to draw from this research is that heritage speakers do not always transfer linguistic information from their dominant language when

---

13   On first and second language acquisition. Cf. Eun-Young Kwon, 'The "Natural Order" of Morpheme Acquisition: A historical survey and discussion of three putative determinants', *Working Papers in TESOL and Applied Linguistics,* 5.1 (2005), 1–21.

14   Maria Polinsky, 'Reanalysis in adult heritage language: New evidence in support of attrition', *Studies in Second Language Acquisition,* 33 (2011), 305–328.

15   The hierarchy is as follows: subject > direct object > indirect object > oblique object > possessor > standard of comparison. See Edward L. Keenan and Bernard Comrie, 'Noun phrase accessibility and universal grammar', *Linguistic Inquiry,* 8.1 (1977), 63–99. To 'relativize' is to refer to the antecedent with a relative pronoun in this role in the relative clause. For example, the relative clause in the English sentence 'That is the boy *who plays football often*' is a subject relative, but in 'That is the boy *whom I see playing football often*,' the relative clause is a direct object relative. The restrictions discussed here refer to the roles the antecedent may play in the relative clause, not in the main clause.

they are missing components of the HL. Interference from one language to the other, although possible, is not a certainty. This is a point with implications for heritage Welsh as well, as English cannot always be assumed to be responsible for any anomalous grammar manifested in the HL. The other factor in effect in organizing linguistic information in the mind, whether it has been acquired to full proficiency or heritage levels, is Universal Grammar (UG).

The common strategies employed in compensating for any grammatical gaps in the speakers' linguistic proficiency are implemented by that distinctly human capacity to identify and create patterns in verbal communication, i.e., UG.

> A child's linguistic system is shaped to a significant degree by the utterances to which the child has been exposed. That is why a child speaks the language and dialect of his family and community. Nonetheless, there are aspects of the linguistic system acquired by the child that do *not* depend on input data in this way. Some cases of this type, it has been argued, reflect the influence of a genetically prespecified body of knowledge about human language. In the literature on Generative Grammar, the term *Universal Grammar* – commonly abbreviated UG – refers to this body of 'hard-wired' knowledge.[16]

That common toolbox of language creation is fundamentally what we are studying when we look at a heritage language – the non-native grammatical forms and patterns that appear in heritage Welsh, when influence from English is an insufficient explanation, must be produced by UG. Because this linguistic asset is shared by all heritage speakers, their HL output will exhibit common features.

The compilation of a corpus of heritage Welsh samples must be the first step in describing the particular properties of heritage Welsh.[17] From this corpus, it is possible to identify the areas of the heritage grammar which are most aberrant from native speaker norms and to establish the avenues of further research which are likely to be the most informative. In addition to being a transcription of the utterances produced by the heritage speakers, both the occurrence of pausing phenomena and total time used to produce each sample are noted. Not only is it possible to identify the grammatical characteristics of heritage Welsh which distinguish it from fully proficient speech, therefore, but also to measure features such as utterance length and speech rate for comparison with fully proficient speakers. Along with the frequency of embedded structure, vocabulary recall delay, repetitions, and corrections, these are the features of speech most closely tied to the impressions of fluency. With these data recorded, and with certain variables in the background of the heritage speakers known, it is possible to make connections between fluency outcomes and the exposure necessary to attain them. For example,

---

16   David Pesetsky, 'Linguistic Universals and Universal Grammar', in *MITCogNet: References Collection,* (Cambridge, MA: MIT Press, 2009) <http://web.mit.edu/linguistics/people/faculty/pesetsky/Pesetsky_MITECS_Universals_UG.pdf> [accessed 10 June 2014].
17   Erin Boon, 'Heritage Welsh,' in *The Polinsky Language Sciences Lab Dataverse* (2014) <http://thedata.harvard.edu/dvn/dv/polinsky> > [accessed 10 March 2014].

heritage speakers vary in their regular continued use of the language, both in conversational practice and in exposure to Welsh language media, as well as in the extent and nature of their early exposure. Because of the information contained in the corpus, preliminary hypotheses about the connection between reduced childhood language exposure and its resulting incomplete grammar can be made.[18]

The corpus of narrative samples was collected from 20 heritage Welsh speakers living in London. Once these speakers were recruited, they participated in a brief interview to establish the details of their history with Welsh. Having established that they were, indeed, heritage speakers as I have defined them here, they performed the *Frog*-story task. This involved flipping through the images of a children's picture-book, Mercer Mayer's *Frog, Where Are You?,* in order to familiarize themselves with the story, then recounting the events of the story as they turn through the images a second time. There are no words accompanying the images. The story is of a boy and his dog searching for their missing pet frog in a wilderness behind their house. The same task was performed by fully proficient, native speaker controls in Wales as well, in order to establish the baseline of native Welsh norms against which to compare the heritage speakers. These narratives were recorded, transcribed, coded, and glossed, before being made available online for further investigation.[19]

This is a test of productive ability, of a speaker's ability to plan and organize a narrative in a weaker language, her speed with vocabulary recall, syntactic choices, and so on. To get a rough idea of the narratives in the corpus, the following box-and-whisker plots represent the distribution of word counts and total time measures in the narratives produced by the participants:
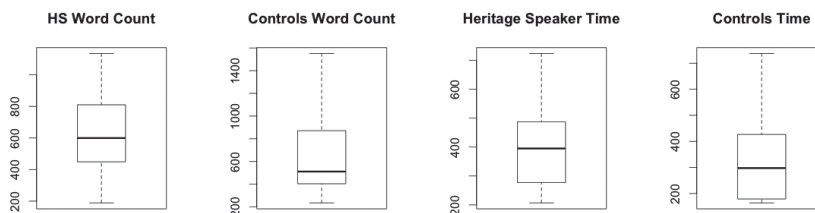

Figure 1 Box-and-whisker plots of the total words in the narratives of the heritage Welsh speakers (HS) and in the narratives of the controls.

Figure 2 Box-and-whisker plots of the total lengths of time taken (in seconds) to produce the narratives of the heritage Welsh speakers and the narratives of the controls.

The heritage speakers generally produced longer narratives, both in terms of word count and temporal length, than did the controls, but the control group

---

18 Erin Boon, 'Variables on fluency in heritage Welsh', paper presented at the *33rd Harvard Celtic Colloquium* (Harvard University, Cambridge MA, October 2013); Erin Boon, 'Measuring the fluency of the Heritage Welsh speaker', in 'Heritage Welsh: a study of heritage language as the outcome of minority language acquisition and bilingualism' (unpublished doctoral dissertation, Harvard University, 2014), pp. 111–60.

19 Boon, 'Heritage Welsh' in *The Polinsky Language Sciences Lab Dataverse.*

showed greater variance on both of those measures.

Superficial differences aside, the two participant populations, the heritage speakers and the controls, also differentiated themselves grammatically. The corpus is glossed for metalanguage (English), word class, initial consonant mutations, as well as any other relevant features. In this form, the data is easily investigated for interesting features of the heritage Welsh grammar. One of the features which divide the two groups particularly well is grammatical gender. Despite grammar book descriptions, the system in standard Welsh is not really a straightforward binary, where all nouns and adjectives show either masculine or feminine gender. Rather, the masculine gender manifests as the default lexical entry, and only the feminine displays gender marking. The soft mutation (SM) is expected on feminine nouns following the definite article and on adjectives following feminine nouns, but no indication is expected to mark the masculine gender in such contexts. This gender pattern makes it particularly difficult to identify every instance in which a speaker truly knows the gender of the noun she has uttered. The manifestation of gender would appear in only a limited set of contexts anyway. As a quick review, SM in standard Welsh is presented in table 1:

Radical form of the word-initial consonant	Form after Soft Mutation/Lenition (SM) *Treiglad Meddal*
/p/ 'p'	/b/ 'b'
/t/ 't'	/d/ 'd'
/k/ 'c'	/g/ 'g'
/b/ 'b'	/v/ 'f'
/d/ 'd'	/ð/ 'dd'
/g/ 'g'	ø
/m/ 'm'	/v/ 'f'
/ɫ/ 'll'	/l/ 'l'
/r̥/ 'rh'	/r/ 'r'

Table 1 The Soft Mutation in Welsh

Overall, SM is a fairly robust feature of the heritage Welsh samples. The heritage speaker participants produced an average of 9.25 instances of SM in contexts where the mutation had been triggered. They produced an average of 1.1 additional instances of SM when the mutation had not been triggered, while also omitting the mutation in a triggered context an average of 7.15 times. Both of these unexpected forms, the overproduction of SM and the omission of SM, are interesting in their own way, especially in the context of grammatical gender. I will return to this point below. The fully proficient native speaker controls, by contrast, produced 23 instances of SM on average. They produced an additional 4.55

mutations which were not in a context where SM was expected, and also omitted the SM in an average of 1.2 triggered contexts. The heritage speakers certainly show a reduced mastery of the mutation when compared to the controls, but the controls do not perform in perfect accord with the standard conventions of the language either. Regardless of the differences in these two groups, there seems to be little reason to doubt that most heritage speakers do acquire the system of SM as a phonological characteristic of their language. What remains uncertain, however, is how many of its grammatical contexts are fully acquired by heritage speakers. The participant samples varied, but one of the most consistent weaknesses related to SM was the manifestation of grammatical gender. The control group identified a feminine mutation context (i.e., performed SM) in 89% of its occurrences. The heritage speakers performed far worse, at a rate of only 62%. The cause of this discrepancy could be dismissed simply as the insufficient exposure of the heritage speaker to the vocabulary of her language and thus her consequent inability to recall the gender of that vocabulary, but the very nature of the gender system in Welsh could be at fault as well.

Welsh does not manifest grammatical gender along a simple masculine-feminine binary similar to the pattern used in many Romance languages. In such a system, gender agreement is essential and apparent in all instances – articles, nouns, and adjectives always show either masculine or feminine gender.[20] In Welsh, although the dictionary may present each noun entry as either masculine or feminine, the form of the article is determined by phonetics, not the gender of the noun. *Yr, y,* and *'r* do nothing to indicate whether the following noun is masculine or feminine, they are instead determined by the letters surrounding them, i.e., their phonetic context. Indeed, even when a following noun is feminine, it and/or its following adjective will only show that gender if the initial consonant is one of the nine mutable consonants affected by SM. However, those nine consonants are not, of course, all of the letters that a feminine word or its following adjective might start with. Notably, no mutation is manifested at all when the word is vowel initial. The limited extent of the manifestation of grammatical gender in Welsh may mean that it is also an exceptionally difficult gender system to acquire. Heritage speakers, however, may be able to shed some light on the topic, as they are a population which is especially useful in determining the ease with which particular features of grammar are acquired.

The heritage speaker is defined by her early exposure to the language. Although that exposure was limited, it was based in the home and proceeded along the natural, native path of language acquisition. This is fundamentally different from the controlled exposure of the late L2 learner in the classroom, and is instead analogous to the language experience of the native speaker who eventually attains full proficiency. The form of the grammar as it survives in the heritage language is determined largely by a process of simplification from the native norms, both because the essentials of communication are the first pieces of language acquired and because they are the last to be lost. The heritage grammar,

---

20   Cf. Grenville Corbett, *Gender*, (Cambridge: Cambridge University Press, 1991). Corbett on gender systems crosslinguistically.

in a sense, appears both as a distilled form of the standard language, in which the grammar is a recognizable simplification, as well as an obviously non-native form of the language, with entirely foreign features grafted onto it from another, possibly dominant, language or from UG itself. The likelihood that a grammatical feature in the standard language will be a part of this simplified heritage language is largely determined by its communicative salience. If the form conveys useful information, or the form itself carries an obvious meaning, then that form is more likely to have been acquired and later maintained by the heritage speaker. For example, the heritage speakers in the corpus sample set overwhelmingly maintained verbal agreement features. The first-person, second-person, third-person, singular and plural forms of the inflected verb *bod* 'to be' consistently agreed with the subject of the sentence.[21] The utility of maintaining these inflections ensures their inclusion in the simplified grammar. The utility of a grammatical system like gender, on the other hand, is debatable.

Given the limited real manifestation of gender in Welsh, as was discussed above, perhaps a more useful description of the gender system in Welsh is not as a masculine-feminine binary, in which every lexical noun is tagged for one or the other, but rather as a [feminine - *marked*] vs. [unspecified] distinction. In this system, only the feminine nouns carry the extra information about mutation contexts. It is a distinction of relative simplicity—the theoretical attraction of constructing the masculine gender as a mirror opposite of the feminine, in which the speaker actively labels these words [masculine] in order to satisfy a binary balance, should not outweigh the more logical ease of not, in fact, needing to remember anything special about the so-called 'masculine' noun at all. [Feminine] nouns do effect SM and at a healthy rate in the control population, therefore they must have a place in the gender system, but when a feature indicates nothing but to maintain the default form, as does the [masculine], it must be questioned whether it is a necessary label at all.

The difference between these two ways of thinking about gender is most important during the acquisition process. The language acquirer will learn aspects of grammar with greater or less difficulty depending on the transparency of their function, the communicative utility of their form, and the simplicity of their patterns. This is where heritage speakers can be particularly informative. A heritage grammar will highlight the features of a particular language which are acquired early or easily because it developed with a reduced amount of exposure during childhood rather than the amount sufficient to produce a fully proficient native grammar. It is, therefore, a simplified version of the complete grammar, made up of whichever components of the full grammar could be acquired given only this limited language experience. If gender were a necessary feature of every noun spoken in Welsh, then we should expect heritage speakers to have learned it fairly well. As such a frequent and important piece of the language, gender labels would have been necessary within that grammatical system in order to complete each

---

21   Cf. Erin Boon, 'Verbs and agreement in Heritage Welsh', in 'Heritage Welsh: a study of heritage language as the outcome of minority language acquisition and bilingualism' (unpublished doctoral dissertation, Harvard University, 2014), pp. 193–219.

Heritage Welsh and the Bilingual Continuum:
Grammatical Gender Under Incomplete Acquisition

lexical entry in the mind of the speaker. If, on the other hand, the only instances where gender is truly an essential piece of linguistic knowledge are those feminine contexts in which the initial consonant happens to be mutable, and/or the following adjective is mutable, then we should expect a higher rate of errors in the heritage Welsh gender system. The following table summarizes gender marking as it was manifested in the heritage speaker and control narratives of the corpus:

	Heritage Speakers		Controls	
	total	average/ speaker	total	average/ speaker
no SM in a feminine context (unexpected form)	60	3	23	1.15
SM in a masculine context (unexpected form)	48	2.4	33	1.65
SM in a feminine context (expected form)	98	4.9	181	9.05

Table 2 The manifestation of grammatical gender in the narrative samples

The 'feminine context' includes both nouns and adjectives which would have manifested with SM according to the expectations of the standard language. The heritage speakers misidentify a grammatical gender context, either identifying a masculine as a feminine or a feminine as a masculine, an average of 5.4 times per speaker. That is a rate which is higher than the rate at which they correctly identify the feminine context, 4.9 times per speaker. The fact that the masculine context presents no overt indication of that gender not only suggests that the gender system in Welsh is something other than a simple binary, but also presents the researcher with the challenge of how to identify when a speaker does, in fact, recognize the masculine gender. Without an overt marking, there is no way to know whether the speaker is intentionally indicating the masculine gender with the unmutated form, or simply neglecting to consider gender at all. Because the default, unmutated form is the expected form in this context, the only gender whose production we can clearly observe is the feminine. It is only the feminine gender, therefore, which can be analyzed fully here, but there are sufficient data to draw conclusions nonetheless. The heritage speakers identified the feminine gender in 4.9/7.9 contexts, or with a rate of 62% accuracy.

The error rate is certainly higher for the heritage speakers than the controls, but there were mistakes in the narratives that they produced as well. The controls did not mutate in the feminine context 1.15 times per narrative on average, while correctly identifying the feminine context an average of 9.05 times per narrative. That is a rate of roughly 89% accuracy. Referring to an aspect of the language produced by native and fully proficient speakers as 'mistakes,' however, is theoretically problematic. I would not presume to say that these speakers do not know their language implicitly, and instead propose that the rate of 'error' in their

gender marking is the natural imperfection of native speech and the high mark for successful acquisition against which to compare the heritage speakers. Rather than thinking of the heritage speakers as identifying the feminine gender with 62% accuracy, they identified the feminine context with 69.7% of the accuracy of the controls.

The standard against which to compare the linguistic attainment of the heritage speaker must be set as realistically as possible. The 'baseline' is the expected target language of the speaker, not the language as it is presented in grammar texts and written media. The subject of study is not the metalinguistic awareness of the speaker, rather it is her naturally acquired and produced speech. The natural speech of even a native and fully proficient speaker still contains errors if we judge it by the standards of the prescriptive grammarian. The baseline for the heritage speaker, therefore, is the set of natural language choices made by the population to which the heritage speaker was exposed in childhood. It is the dialect of the family and community, casual and error-ridden, and it is this language which represents the best possible attainment for the natural acquisition process. It is this level of attainment which was the target the heritage speakers have failed to reach. In the case of gender marking, then, full baseline proficiency is roughly 89% in accord with the dictionary.

Errors of mistaking the masculine as the feminine gender context are particularly interesting. While failing to mutate in a feminine context is an error of omission, and presumably less cognitively taxing than remembering to produce the mutation, mutating a consonant where no mutation is triggered is an added level of unnecessary processing. The fact that these gender misidentifications go both ways, mistaking a feminine for masculine and mistaking a masculine for feminine, and in both the heritage and fully proficient grammars, could indicate that, indeed, all nouns are marked for gender in the Welsh system. The bi-directionality in the errors produced by both participant groups could be taken as evidence that all nouns are tagged as either masculine or feminine in a speaker's lexicon, and the misidentification of a masculine noun is just as likely, therefore, as the misidentification of a feminine noun. If, on the other hand, the gender system only required the tagging of feminine nouns, then we would only expect to see errors of omitting the SM in feminine contexts. In all other instances, the masculine form would appear, i.e., the un-mutated or default form.

These figures may also, however, lend support to the view that gender is not a particularly salient feature of Welsh grammar and would be tagged in the lexicon only when absolutely necessary, i.e. [feminine]. The fact that there are so many errors in the speech of the heritage speakers, performing at a rate of just 69.7% of the accuracy of the controls in marking the feminine context, may indicate that gender is not a critical feature of nouns during the acquisition process. As the heritage speakers are uniquely informative of the acquisition process, we might conclude from their performance that nouns are not learned with their standard gender category from the outset. The acquisition of gender may require more input and exposure to the baseline language than does the acquisition of a noun's phonology and meaning alone, which is evidently sufficient for the purpose of communication. The nominal lexical entry, therefore, is learned without the necessity of assigning

a grammatical gender to that item. Gender assignment comes later. In the case of the heritage speakers, that gender assignment has only been completed to a degree which would identify the feminine at a much lower rate than was exhibited by the controls.

A third possibility may exist as an alternative to the two presented above. The acquisition of grammatical gender may be a process developed in stages, in which the initial state of a recently acquired noun is non-specific to gender, or the default form, which happens also to be the correct form if the noun is masculine. Only after sufficient exposure to the target language is the gender of the noun firmly identified and stored as a part of that lexical entry. This option would also leave room for the likelihood that the gender marking would exist, at first, in a state of flux, and errors would be expected even as the acquirer is beginning to assign a gender to the noun. The lower rate of identification of a feminine context in the speech of the heritage speakers, then, would be explained by the fact that their acquisition was incomplete. The heritage speaker's nominal lexicon was never allowed to fully develop through these stages. Being proficient enough to perform the narrative task, these speakers are also fundamentally aware that the definite article triggers SM for some words, and that those words in turn also trigger SM on a following adjective, but their intuition does not extend to the full set of feminine nouns. They are missing the complete system of gender assignment that can be expected from the fully proficient native speaker. It is a component of the language which, even to the controls interviewed for this project, is intuitive rather than actively memorized. Without sufficient exposure to (and practice with) these manifestations of gender in Welsh, the heritage speakers cannot be expected to have acquired the sense for when a gender-triggered mutation 'sounds right.'

Given the ability of these heritage speakers to complete the narrative task, which I take as evidence of their basic communicative competence in Welsh, the salience of grammatical gender is again called into question. If the correct identification of gender were important to effective communication, we would expect that the gender errors in the heritage speaker narratives would inhibit their ability to tell the story. It is not so, however. It seems, rather, that gender truly is of dubious functional necessity in Welsh, and is merely a formal aspect of the language. This is not altogether a surprise, of course, because the system of gender agreement is also not particularly extensive even in fully proficient Welsh. The only observable instances of gender are in the feminine context, and only in those feminine contexts in which the initial consonant is one of the nine consonants susceptible to SM. Viewed from this perspective, it is impressive that the heritage speakers, with their dramatically curtailed acquisition process, identify the feminine gender as frequently as they do.

The observations discussed here are preliminary, but they do argue that the manifestation of these forms in heritage and standard Welsh deserves further investigation. The corpus has provided enough data to suggest an incomplete system of gender assignment in heritage Welsh, which in turn may indicate an acquisition process which features a period of fluctuation in nominal gender assignment, but without further directed study which is specifically aimed at the question of grammatical gender acquisition it will not be possible to draw any firm

conclusions. A comparison to the speech produced by children acquiring Welsh, for example, would be informative as to whether native-like gender assignment had been acquired and then lost by the heritage speakers through a process of L1 attrition, or the adult heritage Welsh system represents an incompletely acquired grammar, halted during a stage in which parts of the lexicon were still in flux for nominal gender. An overt gender labelling task may also shed some light on the heritage speaker's accuracy in gender labelling. Given a list of nouns to mark as [feminine] or [masculine], rather than a narrative task, perhaps the heritage speakers would exhibit more native-like accuracy and show that the errors produced in the corpus were caused more by the pressures of real-time speech performance than any deviations from native speaker norms in their lexicon. Regardless of the results of any future investigation, however, it is clear that the system of grammatical gender is reduced in the heritage grammar.

The most obvious implications of HL research are in language pedagogy. In a bilingual community where one language is clearly the minority language, some of the students in language courses for adults and older children may actually be heritage language learners – heritage speakers who are actively trying to develop their proficiency in the HL. A thorough understanding of the strengths and weaknesses of these learners would, therefore, be a huge advantage to the language teacher, especially because their strengths and weaknesses do not necessarily match those of the traditional second language learner. Grammatical gender, for example, may be closer to native norms in the speech of second language learners because they tend to learn gender deliberately, whereas the heritage speaker does not tend to think about grammatical gender in such a metalinguistic way. This is a preliminary hypothesis, but it is an example of the grammatical divergences that could exist between these two learner populations because of their dissimilar acquisition experiences. This line of thinking is not a revelation to Welsh tutors. The terminology and research framework of heritage Welsh may not have existed until recently, but the concept is not new to the professionals who work with these students. If for no other reason, the profile of the heritage Welsh speaker should be thoroughly researched in order to best serve the heritage language learners who may make up a significant portion of adult Welsh language learners. These learners have the advantage of their childhood experience with the language, but they would still benefit from a curriculum that caters to their particular grammatical profile. The analysis presented above, for example, provides empirical evidence of the need for overt instruction in grammatical gender for heritage speakers who are otherwise comfortable with these lexical items.

Regarding the heritage speaker as a true bilingual is not always the most intuitive label to assign, especially when we occasionally must consider the needs of such a speaker in the language classroom as well. Even during the course of this project, many of the heritage speakers interviewed were unable to produce any sort of coherent narrative and were left out of the corpus as a result. If the ability to produce the *Frog*-story in Welsh were a valid criterion for Welsh proficiency, however, the challenge of labelling bilingualism would be easily answered. Unfortunately, this would an entirely arbitrary qualification. As individuals with any Welsh proficiency at all, they are included on the bilingual continuum. These are speakers whose

first language was, in fact, Welsh, despite the fact that English later became the dominant language. With that early childhood home exposure, a period privileged for the acquisition of phonology at the very least, the heritage speaker undeniably has some amount of linguistic intuition stored in the mind. Acknowledging that there will inevitably be differences of opinion as to what constitutes Welsh fluency, a concept of bilingualism which is as well-defined as possible is essential. As the state of the research stands now, and given the minority status of the language in question, erring on side of over-inclusion for lower proficiency bilinguals in any counts of Welsh speakers is far safer than excluding those speakers. The framework of the bilingual continuum allows for the inclusion of such speakers. The concept of Welsh-English bilingualism should, therefore, be extended to include heritage speakers regardless of the level of their speaking abilities, and the term 'semi-speaker' must be recognized as derogatory, in which 'semi-' prejudices the researcher to look for disability rather than ability in the speaker's grammar. As heritage speakers, the incomplete nature of these individuals' acquisition process is still recognized, but not in such a necessarily negative light. These are Welsh speakers after all.