Glossa a journal of general linguistics

Willis, David. 2019. Dialect syntax as a testbed for models of innovation and change: Modals and negative concord in the *Syntactic Atlas of Welsh Dialects*. *Glossa: a journal of general linguistics* 4(1): 31. 1–30. DOI: https://doi.org/10.5334/gjgl.844

**RESEARCH**

# Dialect syntax as a testbed for models of innovation and change: Modals and negative concord in the *Syntactic Atlas of Welsh Dialects*

David Willis
University of Cambridge, Selwyn College, Cambridge, GB
dwew2@cam.ac.uk

This paper introduces the pilot project for the *Syntactic Atlas of Welsh Dialects*, setting out the procedures for data collection and sketching a case study for one variable within the dataset, namely the patterns of negative concord found with the negative modal *cau* 'won't'. This item is a relatively recent innovation, and it is currently undergoing increasing integration into the negative concord system. The atlas fieldwork establishes current patterns of dialect variation, showing significant age variation indicative of change in progress and the rise of negative concord in this context. On the basis of this, it is argued that the diffuse geographical patterns attested are best interpreted as evidence of multiple innovation across a wide area, with new speakers re-implementing the innovation ("multiple reactuation"). A formal analysis is sketched out, treating the change as moving along a pathway of feature change with semantic features shifting to interpretable syntactic features and then to uninterpretable syntactic ones. This analysis is consistent with the dialect patterns and interspeaker implicational hierarchies found in the data.

**Keywords:** Dialect syntax; Welsh; Modals; Negation; Negative polarity; Grammaticalization

## 1 Introduction

The re-emergence of interest in the application of techniques from geospatial data analysis to address theoretical questions in dialect syntax has provided us with new tools to investigate long-established questions of how linguistic innovations, specifically morphosyntactic ones, arise and diffuse through geographical space.[1] This paper introduces the *Syntactic Atlas of Welsh Dialects* (*SAWD*) project in the context of such research, and shows how it can be used to address theoretical questions in syntax and language change. Specifically, it considers how the geographical distribution of syntactic variants currently undergoing change can inform our understanding of how syntactic innovations arise and diffuse. While some syntactic innovations are transparent to speakers and can be copied by adults, many others are quite abstract, mediated by changes to feature values on functional items. Since speakers have no direct access to the internalized grammars of others, a reasonable hypothesis is that such innovations spread rapidly by being replicated anew in the newly developing grammars of children during language acquisition ("multiple reactuation"); that is, each new child makes the same parsing "error" that led to the first

[1] Among many such projects in recent years are the Syntactic Atlas of Dutch Dialects (Barbiers et al. 2005; 2008), the Dialect Syntax of Swiss German (Bucheli & Glaser 2002; Glaser & Bart 2011), the Syntax hessischer Dialekte project (Fleischer, Lenz & Weiss 2017), the Syntactic Atlas of Northern Italy (Benincà & Poletto 2007), the Scots Syntax Atlas (Thoms et al. 2017), and the Yale Grammatical Diversity Project: English in North America (Zanuttini et al. 2018).

innovation, encouraged in this misparse by the absence of evidence to the contrary in the speech of children and adults who have already incorporated the change into their grammars. Thus, some syntactic innovations may spread by arising independently during acquisition in the grammars of many individuals. Furthermore, we might expect that this type of change is more likely with abstract morphosyntactic features than with those that are readily identifiable by reference to a single surface attestation.

These possibilities will be examined and evaluated here using data on an innovation from the pilot of the *Syntactic Atlas of Welsh Dialects*. The innovation in question involves the grammaticalization of a new negative modal *cau* 'will not, won't' from a lexical verb *nacáu* 'refuse'. Apparent-time age variation shows that this new modal has become increasingly integrated into the Welsh system of negative concord along an implicational hierarchy of contexts. The atlas data show neither wave-like diffusion of this innovation from a single point, nor hierarchical diffusion from one urban centre to another; rather we find a diffuse pattern of negative concord across a wide area in older speakers, crystalizing into a clearer dialect distribution among younger speakers. A formal analysis of the developments offers some clues as to why this should be.

We begin by setting out the fieldwork methodology of the *Syntactic Atlas of Welsh Dialects* (section 2), before turning to the specific variable under investigation. Section 3 outlines the linguistic and historical background to the new modal *cau*, before determining the patterns of variation and diffusion found for it in the atlas fieldwork data. Section 4 proposes a formal analysis according to which this change represents the conversion of semantic negation into a syntactic negation feature (cf. van Gelderen 2015). Section 5 considers the implications of the formal analysis for patterns of variation and change: an abstract grammatical change (creation of a feature) such as this is both expected from general considerations and difficult for one acquirer to copy directly from another. It is argued that these properties make the change more likely to arise multiple times in different speakers and therefore to appear simultaneously in a diffuse pattern across the geographic area experiencing the innovation.

## 2 The Syntactic Atlas of Welsh Dialects

Results reported here are based on the pilot fieldwork for the *Syntactic Atlas of Welsh Dialects* (*SAWD*) (Willis 2014). We begin by considering the background to the project and the methods that were adopted to collect the data for it before turning to look more closely at the specific variable under investigation in section 3.

Two major dialect atlas projects, namely the *Linguistic Geography of Wales* (Thomas 1973) and the *Welsh Dialect Survey* (*WDS*, Thomas 2000), already exist for Welsh. However, as with many other languages, such traditional dialect work focuses primarily on lexicon (in the case of the *Linguistic Geography of Wales*) or phonology (in the case of the *Welsh Dialect Survey*). While the broad outline of syntactic variation in Welsh is known, the details have never been subject to systematic collection; nor has dialectological work sought to focus much on other kinds of variation or on change in progress. It is these gaps that the *Syntactic Atlas of Welsh Dialects* aims to cover.

A number of recent dialect atlas projects have significantly expanded the range of material available for the study of dialect syntax. As these projects have shown, establishing the distribution of syntactic variants requires somewhat different elicitation techniques than those employed in lexical or phonological dialect atlases. The issues are discussed in detail by Cornips & Poletto (2005), Barbiers & Bennis (2007), Benincà & Poletto (2007) and Zanuttini et al. (2018). *SAWD* draws on the experience of these and other projects, but adapts them in various ways to reflect the specific sociolinguistic context of Welsh

and the quite wide range of variability found at all linguistic levels in spoken varieties of the language.

Fieldwork design of dialect atlases must be informed by the social context in which the language under investigation is spoken, particularly the relationship between standard and dialect and speakers' propensity to be influenced by prescribed, standard variants. In the case of Welsh, a written standard language was established in the sixteenth and seventeenth centuries and was used thereafter for the vast majority of writing. Spoken language, however, did not standardize, and dialect variation in speech has remained significant at all levels (phonology, morphology, syntax, lexicon) since then. As the dialects and standard language drifted further and further apart, a diglossic situation emerged, reaching its fullest extent in the nineteenth century and first half of the twentieth century. The standard variety was used for the vast majority of writing and for reading written texts aloud, while all other spoken communication was conducted in dialect. Since the 1970s, however, the written standard and spoken dialects have begun to converge. Written Welsh has adopted many features that are widespread in the dialects and has begun to admit more variation than before. Better communication between speakers of different dialects has led to convergence and levelling between them; the re-introduction of Welsh-medium education has led to the introduction of more standard-like features into speech. The result is that a stylistic continuum from formal to informal writing has developed. Finally, beginning in the nineteenth century, Welsh itself has become a minority language across much of Wales and bilingualism with English has become the norm. Language revitalization efforts since the 1970s, mediated above all through the education system, have also reduced the differences between standard and dialect by promoting standard features, particularly in those regions (mainly in the southeast and northeast) where the inherited dialect is weak or extinct. Nevertheless, while the relationship between spoken and written Welsh and between standard and dialect is now quite complex, most speakers who have control over the written standard have a clear understanding that they speak and write differently. Speaking in the written standard in an informal situation is still generally regarded as amusing.

This sociolinguistic situation impacts on fieldwork methodology in a number of ways. Above all, it motivated the decision to conduct all fieldwork orally in order to make it quite clear that spoken forms were being asked for and to create an environment where such forms would be entirely natural and expected. Furthermore, the extent of variation in spoken dialects ruled out a direct grammaticality judgment task. At the very least, such a task would need an intermediate stage where informants could translate the sentence under discussion into their own variety because all aspects of the sentence are potentially variable.

Fieldwork was thus conducted using a questionnaire administered orally by a fieldworker to Welsh-speaking informants across Wales. Responses were recorded, and transcribed after the interview. After a series of questions used to establish demographic metadata, the questionnaire itself consisted of 60 pairs of sentences. Each pair was read aloud to the informant, and the informant was asked to repeat the second sentence, adapting it so as to express the same idea naturally as they would in everyday speech. The first sentence was given to provide a plausible context, to make the second sentence seem like a natural thing to want to say, and informants were not asked to repeat it (although a few did and, if they persisted in doing this, were allowed to continue in the same way). The first two sentences of the sequence were designed as practice sentences to ensure that the informant had understood the task. They contained well-known dialect variants where we can be confident that any informant doing the task as intended would make changes of

a particular kind. By the second of these, almost all informants had correctly understood the task.

Each test sentence contained one or more features known to be dialectally variable. Consider the following pair of sentences (item 4 of the questionnaire):[2]

(1) Mae hynny 'n syniad da iawn. Ai ti feddyliodd
be.PRS.3SG that PRED idea good very Q.FOC you.SG think.PST.3SG
amdano fo gynta?
about.3MSG it.M first
'That's a very good idea. Was it you that thought of it first?'

The first, carrier sentence contains nothing of particular syntactic interest. The second sentence is designed to test what focus-question marker the informant uses and the form of their second-person singular pronoun: the focus-question particle *ai* is offered, but informants might easily change it to another variant of this particle (either *ife* or *dife*) or omit it entirely; the form of the second person singular pronoun in a focus sentence here is *ti*, but informants might change it to another variant, *chdi*. Additionally, the form of the third-person singular masculine pronoun offered is *fo*, but informants might change it to *fe*.

In principle, the sentence could also be said to test whether a focus structure is appropriate at all in this context, and a few speakers replaced the focus structure (focus marker – focused subject – verb) with an ordinary AuxSVO structure (auxiliary – subject – verb), with focus indicated by intonation.

Test sentences were deliberately constructed so as to contain a mixture of morphosyntactic variants from different dialects, often in combinations that would be highly unlikely to occur in the speech of a single individual. This was done first of all to create an atmosphere where informants felt permitted and even encouraged to use non-standard or dialect forms free from prescriptive pressure. Secondly, it ensured that any informant doing the task correctly would make frequent changes to the test sentences, and anyone merely repeating mechanically what was said to them would be easily recognized. In practice, informants made frequent minor changes not relevant to the grammatical properties under investigation. Examples from the sentences discussed in this article are:

(i) phonological e.g. replacement of *dydi*, the negative concord form of 'be', with a variant such as *tydi* or *di*, or vice versa;
(ii) morphosyntactic e.g. replacement of the bipartite doubling possessive construction with a lone postverbal pronominal possessor (e.g. *ein teledu ni* 'our TV' becomes *teledu ni*); or addition of expletive 'there' (*'na* or *'a*) to a sentence; and
(iii) lexical (e.g. replacement of *teledu* 'television' with *TV*, *teli* or *telefision*; or *cychwyn* 'start' with *startio*).

The rate of paraphrasing varies from question to question across the entire questionnaire. Consistent paraphrasing of the item under investigation is interpreted as rejection of it as part of the informant's grammar. Thus, one informant paraphrased 'Your car won't start' as 'The car is refusing to start' (using a different verb, *gwrthod* 'refuse'); 'Your phone won't work' as 'Your phone doesn't work'; 'He still won't' as 'He still hasn't'; 'Our TV won't work' as 'Our TV doesn't work'; and 'No one will come out with me' as 'No one is coming out with me tonight'. All these sentences were offered with modal *cau* expressing 'won't'

[2] Glosses follow the Leipzig Glossing Rules, except that PRED is used to gloss the predicate marker *yn* and IMPF is used to gloss the imperfect.

and its removal in five cases, set against the background of other speakers who did not paraphrase these sentences, is a good indication that the speaker lacks this item in their grammar.

It was felt to be simply impossible to follow the practice of some other syntactic dialect atlases (e.g. Cornips & Jongenburger 2001: 57, 60; Zanuttini et al. 2018) in asking for grammaticality judgments on a single set of sentences used for all informants. Above all, this was because it proved impossible to construct test sentences in which exactly one feature was potentially dialectally variable: this method would always be vulnerable to the possibility that sentences were being rejected for irrelevant reasons, such as lexical choice, lexical dialect variation, morphological dialect differences etc. For instance, from their responses, it became apparent that some speakers would reject the lexical item *bisged* 'biscuit, cookie' in favour of the alternative form *bisgïen*. In a few cases, unexpected dialect variation became apparent from the responses in items where it had not been intended; for instance, the choice of preposition in *edrych ymlaen at/am* 'look forward to' proved to be variable. Such factors could not be exhaustively foreseen.

Other projects have used fieldworkers or native speaker informants to translate the test sentences into the dialect of each informant and then ask for judgments (e.g. parts of the SAND project, Barbiers & Bennis 2007). This was judged impractical in the current context and also raised the problem of what to do if the form simply does not exist in the relevant dialect. For instance, the negative modal *cau* discussed below does not alternate with any particular alternative lexical item in those dialects that lack it: those dialects simply use a paraphrase.

A core of 34 sentences was administered to all speakers. A further 24 sentences were used only in the north, and 24 only in the south. Once the two practice sentences are included, this means that each informant responded to 60 sentences in total. Interviews typically lasted around 15–20 minutes. Informants were recruited across Wales, with most fieldwork conducted in Denbighshire, Bangor, Swansea, the Neath Valley, Cardiff or Cambridge. Since this was a pilot for a larger full-scale project, there were no quotas for particular types of informant, and this leads to some underrepresentation in parts of south Wales (notably Ceredigion and northern Carmarthenshire). In total, 192 informants were interviewed, 98 from the north, and 94 from the south. The date of birth of speakers ranged from 1920 to 1999, with the mean year of birth being 1965. In the full-scale project, quotas for geographic region, age and linguistic background[3] will be implemented to allow a fuller picture of variation to be established.

## 3 Negative modal *cau*

### 3.1 The background to the variable

The variable for consideration here is the syntax of *cau* 'won't' (pronounced /kaɬ/), a new negative modal that has developed in northern dialects. Historically, this item derives from a lexical verb *nacáu* /naˈkaɬ/ 'refuse' (ultimately from *nag*, the negative particle used in responsives, see Jones 1999, plus causative *-ha-* and nonfinite (verbnoun) suffix *-au*). While the lexical verb is attested from the thirteenth century onwards (Thomas et al. 1950–2002: s.v. *nacâf*), the phonological reduction from *nacáu* to *cau* dates only

[3] "Linguistic background" refers to whether a speaker grew up in a Welsh-speaking home or not. This feature should allow us to examine the impact of revitalization on the structure of the language. In practice, speakers from non-Welsh-speaking backgrounds tended to lack confidence and proved rather difficult to recruit. Few are included in the pilot. Furthermore, the bipartite division into "Welsh-speaking" or "non-Welsh-speaking" home is not easy to implement, since homes may be partially Welsh-speaking and changes in circumstance during childhood (migration, divorce etc.) may mean that this status is not clearly defined for a given informant.

from the nineteenth century. Even more recently it has undergone a semantic extension from 'refuse' to a modal meaning 'be predisposed not to' and is usually best translated into English as 'won't'. As such it has become more integrated into the system of modal auxiliaries (cf. *gallu* and *medru* 'be able', *ffaelu* and *pallu* 'be unable') and thus undergone grammaticalization.[4]

In this new meaning it allows non-volitional, non-agentive subjects:[5]

(2) Mae 'r drws (yn)[6] cau agor.
be.PRS.3SG the door (PROG) CAU open.INF
'The door won't open.'

This semantic extension seems to have occurred in all dialects where the phonological reduction occurred. The distribution of *cau* was tested in the *Welsh Dialect Survey* (Thomas 2000: item 448). The *WDS* results are given in Figure 1.[7] As can be seen there, *cau* was reported as present (red) across almost the whole of the northern dialects but is entirely absent (blue) in the south. The patchiness of the distribution is also noteworthy. This may be due to difficulties in eliciting it with the *WDS* methodology (fieldworkers described the item under investigation to encourage informants to say it, but, in general, did not ask about it directly), but, as will been seen below, the same pattern is found in the *SAWD* data.

### *3.2 Innovation of negative concord*

The main focus of interest here is a further change that is currently in progress, integrating this new modal into the negative concord system. To understand this innovation, we need to consider the way in which sentential negation and negative concord are manifested in Welsh. Sentential negation is expressed in two main ways. Verbs in negative sentences undergo morphological or morphophonological changes. This is most evident in the present tense of the verb *bod* 'be', which has a completely distinct negative-concord paradigm. Thus, in (3)–(4), the verb changes from affirmative *mae* to negative *dydi*.

(3) Mae Aled yn chwarae.
be.PRS.3SG Aled PROG play.INF
'Aled is playing.'

(4) Dydi Aled ddim yn chwarae.
NEG.be.PRS.3SG Aled NEG PROG play.INF
'Aled is not playing.'

[4] This order of events, with phonological reduction preceding semantic change and seemingly facilitating it, is an interesting counterexample to the usual hypothesis (Bybee & Pagliuca 1985: 59–60) that the phonological reduction often (but not always) witnessed in grammaticalization changes is the result of semantic change (whether bleaching, generalization or refunctionalization) that increases the frequency of the item undergoing grammaticalization, either following it in time or proceeding in parallel with it: "as the meaning generalizes and the range of uses widens, the frequency increases and this leads automatically to phonological reduction and perhaps fusion" (Bybee & Pagliuca 1985: 76).

[5] All examples are given in northern colloquial forms, reflecting the linguistic level that is the object of study in a dialect atlas. They may differ significantly from standard, written literary Welsh forms.

[6] In spoken northern Welsh, the unmarked ("progressive") aspect marker *yn* is frequently elided in speech and it is not entirely clear that it is even present in the syntax at a more abstract level. Its phonological and perhaps syntactic optionality is indicated by parentheses in this and other examples.

[7] "Rejected" in Figure 1 indicates that the fieldworker actively offered the word under investigation and it was explicitly rejected by the informant. In all other cases, the fieldworker tried to elicit the item through paraphrase and, in some cases, the absence of *cau* in the north may simply be because this elicitation technique did not succeed in bringing the item to mind.

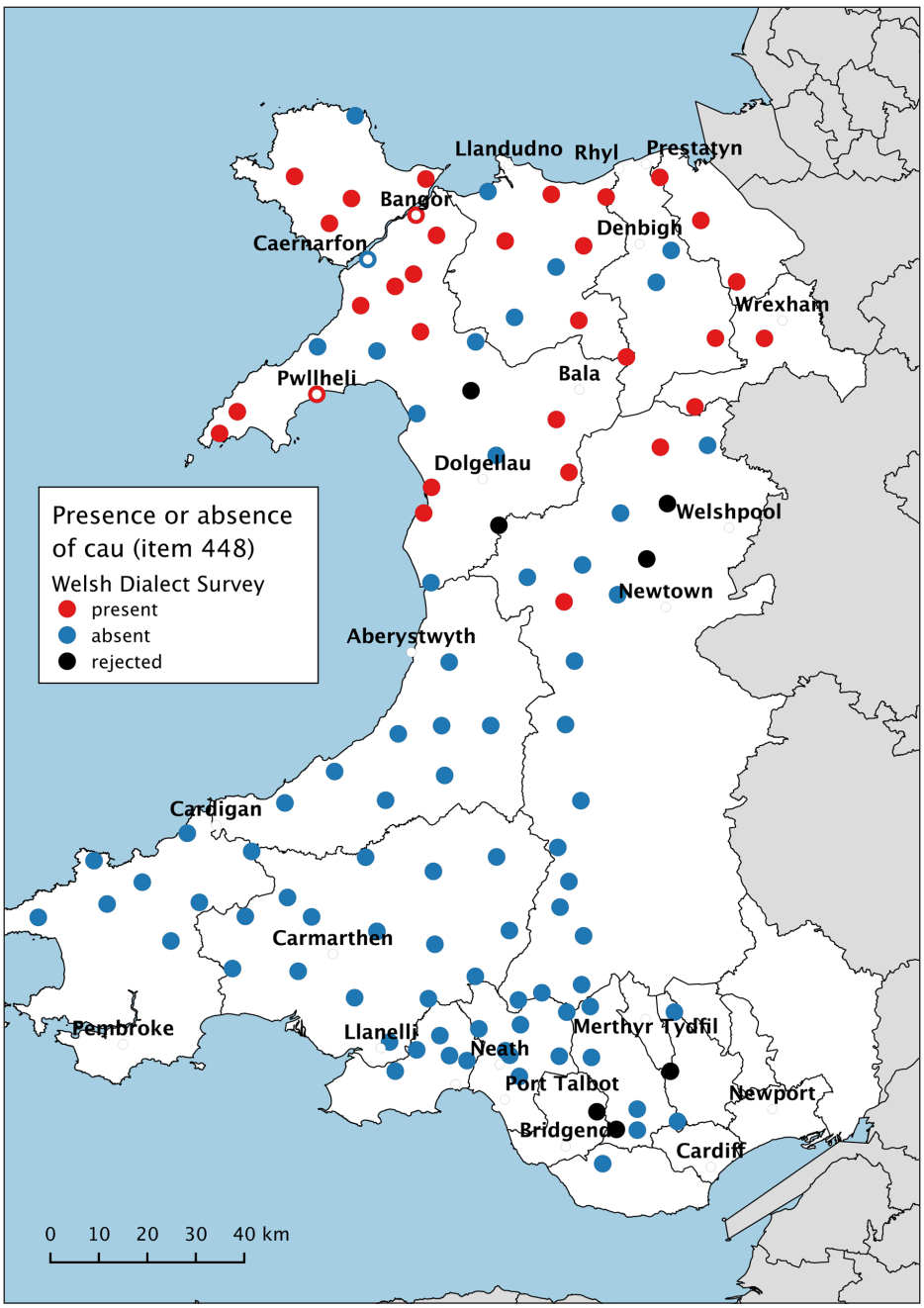



**Figure 1:** Geographical distribution of *cau* (data from Thomas 2000: item 448).

Other verbs show more subtle differences involving the system of initial consonant mutation. In (5), in an affirmative clause, the verb may appear either in its unmutated radical form *ceith* 'will get, will be allowed' or in its soft-mutated form *geith*, while, in a negative clause, in (6), the radical form is impossible, and the possible forms are the soft-mutated form *geith* or the aspirate-mutated form *cheith*.[8]

These changes alone are not sufficient to render a clause negative, however. To do this, the sentential negator *ddim* must also be added in immediately post-subject position, as in (4) above and (6) below. This demonstrates that these changes mark a form of negative concord, and are not an expression of negation in and of themselves.

[8] For further details of the system of initial consonant mutations in Welsh, see Borsley, Tallermann & Willis (2007: 223–254) and the references therein.

(5) Ceith/Geith Aled chwarae.
get.FUT.3SG Aled play.INF
'Aled will be allowed to play.'

(6) Geith/Cheith Aled ddim chwarae.
NEG.get.FUT.3SG Aled NEG play.INF
'Aled will not be allowed to play.'

These changes are also manifested in clauses containing n-words such as *neb* 'no one, anyone':

(7) Dydi Aled ddim (yn) helpu neb.
NEG.be.PRS.3SG Aled NEG (PROG) help.INF no.one
'Aled is helping no one./Aled isn't helping anyone.'

Borsley & Jones (2005), working within HPSG, posit a restriction on n-words that they must appear in one of a small number of specific contexts (the Negative Context Constraint), the most prominent one being a negative context, a head marked [POL *neg*], defined widely to include all of the elements discussed here: *dydi* in (4), *geith* and *cheith* in (6), and *dydi* in (7). Where such an element is not present, n-words are ungrammatical. A further constraint, the Negative Dependent Constraint, requires some verbal elements to have a negative dependent such as *ddim* in order to be grammatical.

One complication is that, if *ddim* would appear adjacent to an n-word, it can (and, in many varieties, must) be omitted. In (8), *ddim* would appear immediately after the subject *Aled*, which would be directly adjacent to *neb*. Note that the mutation triggered by negative concord is compulsory on the verb in (8) (radical *gwelodd* 'saw', soft-mutated *welodd*), indicating that the sentence is not simply opting out of the negative-concord system.

(8) Welodd Aled neb.
NEG.see.PST.3SG Aled no.one
'Aled saw no one.'

The grammaticality of parallel sentences containing *ddim* varies depending on dialect and varies among n-words. Use of *ddim* before *byth* '(n)ever (in the future or generically)' and *erioed* '(n)ever (in the past)' is quite widespread. King (1993: 253) notes of the sequence *ddim erioed* 'not (n)ever' that it "is regarded by many speakers as sub-standard", confirming its widespread existence. Borsley & Jones (2005: 98) consider the sequences *ddim byth* and *ddim erioed* to be characteristic of southern dialects, citing sentences such as the following as grammatical, but only for southern speakers:

(9) Borsley & Jones (2005: 98)
Dw i ddim erioed wedi gweld hyn o'r blaen.
NEG.be.PRS.1SG I NEG never PRF see.INF this before
'I have never seen this before.'

The sequence (*d*)*dim neb* is in fact also acceptable for some speakers, again predominantly, but not exclusively, in southern varieties.[9] A naturally occurring example is given in (10).

[9] The choice between soft-mutated *ddim* with initial /ð/ and radical form *dim* with /d/ is determined by grammatical factors not immediately relevant to the present discussion. *D*(*d*)*im* has multiple functions in the negative system, appearing variously as an n-word *dim* (*byd*) 'anything, nothing', a sentential negator *ddim*, and an emphatic sentence-final negator *dim*. Ultimately all derive historically from a noun *dim* 'thing'. For discussion of the chain of historical innovations that gave rise to the current situation, see Willis (2006; 2010; 2012; 2013).

(10) (*Golwg*, "Cyfyngu'r wisg wen i brifeirdd a phrif lenorion", 4 August 2011)
Does dim neb yn bwysicach na 'i gilydd.
NEG.be.PRS.3SG NEG no.one PRED more.important than RECP
'No one is more important than each other.'

The equivalent sentences with "weak" negative concord are exemplified in (11) and (12). These are grammatical in all dialects.

(11) Dw i erioed wedi gweld hyn o'r blaen.
NEG.be.PRS.1SG I never PRF see.INF this before
'I have never seen this before.'

(12) Does neb yn bwysicach na 'i gilydd.
NEG.be.PRS.3SG no.one PRED more.important than RECP
'No one is more important than each other.'

The option in (9) and (10) with *ddim* (or *dim*) adjacent to an n-word may be thought of as "strong" negative concord, as compared to the regular "weak" negative concord system exemplified in (8), (11) and (12). A similar difference will arise in our discussion of *cau*, see example (26) below and discussion in section 3.5.4.

Returning to the syntax of *cau*, consider again example (2) above. Here, *cau* does not participate in this negative concord system and triggers only the affirmative form of the tense-aspect auxiliary *mae*. This option is still widely attested and is clearly the historically older form. However, it appears to be losing ground to a relatively recent innovative variant, in which the negative form of the tense-aspect auxiliary, *dydi*, is used:[10]

(13) Dydi 'r drws (yn) cau agor.
NEG.be.PRS.3SG the door (PROG) CAU open.INF
'The door won't open.'

The central research questions here are: what is the geographic distribution of this variant; why did it arise; and why has it been successful?

Given what we saw in (9) and (10) above, another conceivable possibility is that *cau* might additionally be associated with the presence of the sentential negative marker *ddim*, that is, sentences of the following type:

(14) Dydi 'r drws ddim (yn) cau agor.
NEG.be.PRS.3SG the door NEG (PROG) CAU open.INF
'The door won't open.'

We will also test whether sentences of this type are found and, if so, answer the same questions with reference to them.

### *3.3 Innovation of VP-ellipsis (auxiliary stranding)*

Finally, we want to test how *cau* relates to ellipsis phenomena. The availability of VP-ellipsis provides a way to test the categorial status of *cau* and thus its degree of grammaticalization along the pathway from lexical (control) verb to modal auxiliary. Ellipsis after modal auxiliaries is possible in Welsh, as in (15), leaving the modal auxiliary stranded after the aspect particle:

[10] In the current survey, there are speakers born in the 1920s who have this variant. I know of no attestations in written texts as early as this, so our best guess must be that the innovation dates to the 1920s or thereabouts.

(15) Dydi Mari ddim yn medru agor y drws a
NEG.be.PRS.3SG Mari NEG PROG be.able.INF open.INF the door and
dydi Ifan ddim yn medru chwaith.
NEG.be.PRS.3SG Ifan NEG PROG be.able.INF either
'Mari can't open the door and Ifan can't either.'

After control verbs, however, as in (16), and raising verbs, in (17), we do not find direct ellipsis. Instead, a complement containing *gwneud* 'do' is possible, and its complement may undergo ellipsis. In these examples, omission of *gwneud* leads to partial or complete degradation of grammaticality:

(16) Mae Mari 'n disgwyl gadael yn gynnar ac mae Ifan
be.PRS.3SG Mari PROG expect.INF leave.INF PRED early and be.PRS.3SG Ifan
yn disgwyl *(gwneud) hefyd.
PROG expect.INF do.INF too
'Mari expects to leave early and Ifan expects to (do) too.'

(17) Mae 'r wal wedi dechrau disgyn ac mae 'r tŵr wedi
be.PRS.3SG the wall PRF start.INF fall.INF and be.PRS.3SG the tower PRF
dechrau ??(gwneud) hefyd.
start.INF do.INF too
'The wall has started to fall and the tower has started to too.'

Availability of ellipsis therefore provides us with a test for auxiliary status in Welsh.[11]

In these cases, we can take *gwneud* to be the light verb that is found elsewhere in periphrastic constructions. An example is given in (18), where *gwnaeth* 'did' is followed by a nonfinite VP. This is synonymous with an inflected verb of the same tense, as given in (19).

(18) Gwnaeth Mari adael yn gynnar.
do.PST.3SG Mari leave.INF PRED early
'Mari left early.'

(19) Gadawodd Mari yn gynnar.
leave.PST.3SG Mari PRED early
'Mari left early.'

Given this, we can interpret the VP-ellipsis in a sentence like (16) as involving elision of a VP-complement of light verb *gwneud* in v:

(20) $[_{vP}$ $[_{v}$ gwneud] $[_{VP}$ gadael yn gynnar]]
do.INF leave.INF PRED early

The observation that ellipsis of the VP-complement of v is possible in Welsh leads us to an account of the development of *cau* in which the emergence of ellipsis after *cau* is linked to its ongoing grammaticalization as a modal auxiliary.

Rouveret (2012), building on such work as Merchant (2001), Gallego (2009) and van Craenenbroeck (2010) develops a theory of ellipsis in which ellipsis is licensed for complements of phasal heads, notably v. Before its grammaticalization as an auxiliary, *cau* is a lexical, eventive control verb. In Rouveret's terms it is the spellout of vn + Root in the

[11] It is otherwise rather difficult to construct tests for auxiliary status in Welsh; cf. the discussion in Borsley, Tallerman & Willis (2007: 44–47).

structure [$_{vnP}$ vn [$_{RootP}$ Root [$_{CP}$ …]]], where vn is the functional projection that labels the root as a verb and introduces nonfinite ("verbnoun") morphology. Applying this account to the case at hand, the historical (pre-grammaticalization) *cau* is a full verb, and therefore unable to license ellipsis of its entire complement. Ellipsis is possible only of the complement of the lowest v head in the derivation. As a lexical verb, this *cau* takes a full clausal complement, and it is only the complement of the v head within this structure that can undergo ellipsis. Since this v head is spelled out as *gwneud* 'do', ellipsis cannot occur directly after *cau*, and instead speakers will require the sequence *cau gwneud* (or some other paraphrase).

After grammaticalization has occurred, *cau* is an auxiliary, and is therefore inserted into v. We would therefore expect it to allow full deletion of its complement. We thus predict that, when *cau* grammaticalizes from lexical verb to modal auxiliary, it will gain the ability to license bare VP-ellipsis of its complement. That is, we expect conservative speakers to require insertion of *gwneud* in sentences like the following, while we expect innovative speakers to permit its omission (modulo also variation in negative concord patterns as outlined above):

(21) Mae d' un di 'n agor ond dydi f' un i
be.PRS.3SG 2SG one you PROG open.INF but NEG.be.PRS.3SG 1SG one me
cau (gwneud).
CAU do.INF
'Yours is opening, but mine won't (do).'

In this example, before grammaticalization, *cau* spells out the eventive lexical verbal root (RootP, effectively VP). The structure is:

(22) [$_{vP}$ [$_{AspP}$ [$_{vnP}$ [$_{RootP}$ [$_{Root}$ *cau*] [$_{CP}$ [$_{TP}$ [$_{vP}$ *gwneud* ~~[$_{AspP}$ [$_{vnP}$ [$_{RootP}$ [$_{Root}$ *agor*]]]]~~]]]]]]]]

Ellipsis must target the complement of a vP. Although the highest AspP is complement to a vP, it cannot target this, because doing so would delete semantic content (i.e. the semantics of *cau* itself) that cannot be recovered from the first conjunct. Ellipsis can, however, target the complement of the second vP. If it does, this vP will itself spell out as light verb *gwneud*.

After grammaticalization, *cau* spells out a functional rather than a lexical head (in a much reduced, monoclausal structure), treated as v in the following:

(23) [$_{vP}$ *cau* [$_{AspP}$ [$_{vnP}$ [$_{RootP}$ [$_{Root}$ *agor*]]]]]

Ellipsis again must target the complement of vP, but now this can be the entire complement of *cau* itself, with the result that the structure does not spell out as *gwneud*, and *cau* is left alone in clause-final position.

Given this prediction, we would like to be able to test the extent to which speakers are in fact moving in their judgments along this pathway.

### *3.4* Cau *in the project questionnaire*

Given this context, it was decided to include five test sentences containing *cau* in the pilot *SAWD* questionnaire.[12] Three of them present speakers with different options for negative concord: one without negative concord, in (24); one with negative concord of the

[12] Note here and elsewhere that test sentences from the questionnaire should not be interpreted as grammatical sentences of Welsh. They are left unmarked as to grammaticality here: only a speaker response can be

auxiliary alone (weak negative concord), in (25); and one with both negative concord of the auxiliary and insertion of a concording negative marker *ddim* (strong negative concord), in (26).[13]

(24) (item 16)
Well i ni ffonio'r garej.
**Mae** dy gar chdi 'n cau cychwyn.
be.PRS.3SG 2SG car you PROG CAU start.INF
'(We'd better phone the garage.) Your car won't start.'

(25) (item 17)
Alla i ddim eu ffonio nhw …
**Dydi** dy ffôn di cau gweithio.
NEG.be.PRS.3SG 2SG phone you CAU work.INF
'(I can't phone them.) Your phone won't work.'

(26) (item 41)
Bydd rhaid ni fynd at ffrindiau i wylio'r ffilm …
**Dydi** ein teledu ni **ddim** yn cau gweithio.
NEG.be.PRS.3SG 1PL television we NEG PROG CAU work.INF
'(We'll have to go to friends to watch the film.) Our TV won't work.'

One question, (27), has *cau* accompanying an n-word, namely *neb* 'no one', in subject position, offering a clause which takes the option of including negative concord. We have seen above that *neb* requires negative concord. If a speaker has a grammar that does not allow negative concord with *cau*, sentences of this type should either be ungrammatical for them (because of the conflicting requirement of having both negative concord and its absence) or should yield a double negative interpretation ('No one won't come out with me.'), which is obviously pragmatically infelicitous given the context sentence.

(27) (item 42)
Dwi'n gorfod aros gartre heno …
Does **neb** yn **cau** dod allan efo fi.
NEG.be.PRS.3SG no.one PROG CAU come.INF out with me
'(I've got to stay home tonight.) No one will come out with me.'

Consequently such speakers would be expected to have no grammatical way to combine *cau* with an n-word (without producing a double negation reading). If they use the affirmative form of the auxiliary, as in (28), then *neb* fails to occur within a negative context. If they use the negative form, then *cau* introduces an additional, lexical negation, resulting in a double negation reading ('No one won't come out…'). We expect them to paraphrase, either keeping the form of (28), but replacing *neb* with *pawb* 'everyone', eliminating the licensing problem, or by omitting the modality completely and replacing *cau* with some item in the same general semantic area, such as *isio* 'want', as in (29).

interpreted as (likely) grammatical. In (24), for instance, it is dubious whether *dy gar chdi*, with possessive proclitic *dy* and postnominal (strong) pronoun *chdi*, is ever a grammatical way of saying 'your car'. In general, informants responded by expressing this idea using either a possessive proclitic with a weak postnominal pronoun (*dy gar di*) or no possessive proclitic with a strong postnominal pronoun (*car chdi*, *car ti*).

[13] The progressive particle *yn* (*'n* after a vowel) is often omitted in spoken northern Welsh. It was omitted in (25) and included in (24) and (26). It has generally been put in parentheses in constructed examples, because, in the northern spoken responses of the atlas fieldwork, speakers generally omitted it.

(28) *Mae **neb** (yn) **cau** dod allan efo fi.
be.PRS.3SG no.one PROG CAU come.INF out with me
('No one will come out with me.')

(29) Does **neb** isio dod allan efo fi.
NEG.be.PRS.3SG no.one want come.INF out with me
'No one wants to come out with me.'

The final question, given in (30), tests for the grammaticality of bare VP-ellipsis after *cau*, as discussed in section 3.3 above. From personal observation, it was expected that a few (younger) speakers would permit VP-ellipsis after *cau* in a sentence such as this.

(30) (item 34)
Dwi 'di gofyn i Aled eu ffonio nhw yn barod …
Ond tydi o dal cau.
but NEG.be.PRS.3SG he still CAU
'(I've asked Aled to phone them already.) But he still won't.'

### *3.5 Results: Patterns of diffusion*

#### 3.5.1 Weak negative concord

Consider first the overall patterns for the presence or absence of weak negative concord, that is, concord reflected in the form of the finite verb only. Figure 2 shows the raw data for all speakers.[14] Speakers are divided into four types according to their responses across the 5 questionnaire items containing *cau*: (i) those who retained *cau*, but removed

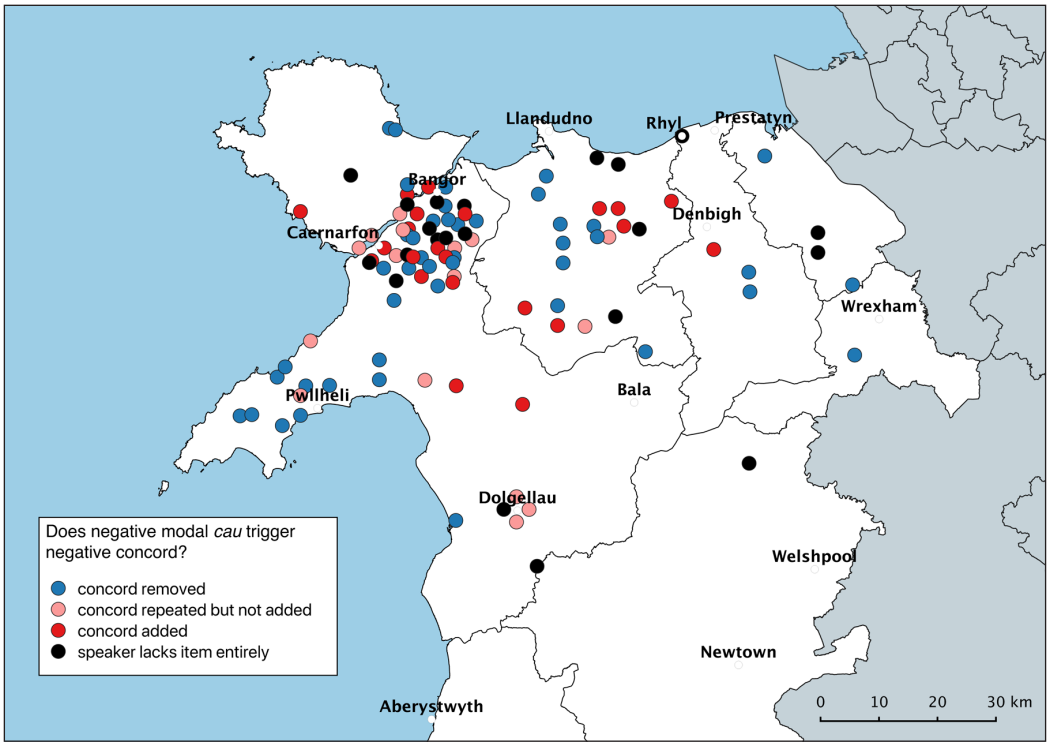



**Figure 2:** Overall distribution of negative concord with *cau* in the raw *SAWD* and Siarad data.

[14] To supplement the pilot data set, the Siarad corpus of spoken Welsh (Deuchar & Davies 2014; Deuchar, Davies & Donelly 2018) was searched for instances of *cau*. Although 15 instances were found, 5 were ambiguous as to whether negative concord had been triggered. The remaining 10 instances are included henceforth in the data.

negative concord in at least one item where it was offered (that is, items 17 = example (25) above, 34 = (30), and 41 = (26)) (blue); (ii) those who retained negative concord with *cau* at least once where it was offered, but did not add it to another item (pink); (iii) those who added negative concord to at least one *cau*-sentence that lacked it (that is, either item 16 = (24) or item 42 = (27)) (red); and (iv) those who rejected use of *cau* entirely by rephrasing all of the test sentences (black). This gives a general overall sense of the data, showing that negative concord with *cau* is found in a diffuse way over much of the area under consideration. However, Figure 2 does not represent the ages of speakers, so gives us no access to age differences or apparent-time change; nor does it allow us to see other more fine-grained differences in the use of negative concord with *cau*.

To see the geospatial patterns more clearly, we can run a geographically weighted regression (GWR) (for further discussion of the application of this technique in linguistics, see Willis 2017). In a GWR model, regressions are performed at different points across a geographic area, with data weighted according to distance from that point in question. Using year of birth as a global factor, and allowing the intercept to vary across geographical space, this allows us to create a model in which change diffuses from one place to another.

A global logistic regression, that is, one that takes no account of geography, was performed using R to establish a baseline against which to compare the GWR model.

In order to produce more easily interpretable results, year of birth was coded relative to 1950; that is, 1950 is coded as 0, 1972 as 22 etc. (cf. Kroch 1989b: 225). This is an arbitrary choice of date, but the move allows us to interpret the intercept of the model as representing the extent of change in 1950 and the age coefficient as representing the subsequent progression of the change for each additional year since then.

This model produced an intercept value of $-1.47199$ logits (standard error = 0.40352, z-score = $-3.648$, one-tailed p-value = 0.000132). This can be interpreted as meaning that a speaker born in 1950 (selected randomly from anywhere for which interviews were conducted) has a probability of accepting negative concord of 0.187 (the link function is $p = \exp(k)/1+\exp(k)$, where k is the intercept value).[15] The age coefficient is 0.05182 logits per year (standard error = 0.01674, z-value = 3.096, one-tailed p-value = 0.000981).[16] This shows year of birth to be a significant factor at the $p < 0.001$ level, with younger speakers more likely to use negative concord.[17] The global probability of a speaker accepting negative concord thus rises to 0.393 for a speaker born in 1970 ($-1.47199 + (20 * 0.05182) = -0.4356$, followed by application of the link function) and to 0.646 for a speaker born in 1990. The AICc of this model (a measure of goodness of fit that takes into account model complexity) is 80.467, and 13.2% of deviance is accounted for.

In the GWR model, we make the following assumptions. These are in conformity with much existing work on language change, and are also the simplest assumptions consistent with the need to allow both change and dialect variation:

[15] The significance (p-value) of the intercept of the global model is given for completeness, but it should be noted that it has little practical significance. The null hypothesis whose rejection is tested by this significance test is the hypothesis that the value of the intercept is in fact zero, that is, that the tipping point of the change (the point where exactly half of speakers accept it) lies exactly in the reference year, namely 1950. This hypothesis is not of any particular interest.

[16] A one-tailed p-value is appropriate, since the variation is known to be due to a historical innovation (rather than of uncertain origin).

[17] Note that year of birth is a continuous variable and the coefficient and p-value are therefore based on differences from one year to the next rather than across decades or arbitarily chosen age groups.

(i) change proceeds via an S-curve (Greenberg, Osgood & Saporta 1954: 155), which we can model using the logistic function (Altmann et al. 1983; Kroch 1989a; b; Denison 2003; Blythe & Croft 2012);
(ii) for a given change, this S-curve has the same slope everywhere; that is, the speed at which a given change diffuses is constant, presumably because the fitness bias that favours the innovation (whether prestige, functional utility or ease of acquisition) is constant across some given language area (Kroch 1989a; b);
(iii) a given change may reach different geographic locations at different points in time.

The first two of these are shared with a global logistic model; however, the third is specific to a geographically weighted model. The model chosen is semiparametric; that is, not all factors vary geospatially: the effect of age is set globally, while the intercept (effectively an estimate of how early or late an innovation reaches a particular location) varies from place to place.

A semiparametric geographically weighted logistic regression was thus performed using the GWR4 software package (version 4.08) (Nakaya 2014). A fixed Gaussian kernel was used, with bandwidth selected so as to produce the optimal model in terms of AICc minimization via the Golden Search function of GWR4. This procedure yielded a bandwidth of 12.36 km. The age coefficient of this model is 0.05825 logits per year (standard error = 0.02020, z-score = 2.88382, one-tailed p-value = 0.001965).

The results of this model are most easily visualized in map form, as given in Figure 3. Here sample locations are coloured according to the earliest date at which a majority of speakers born in that year (who use *cau*) find it acceptable with negative concord. Henceforth, speakers who reject the *cau*-construction entirely are excluded from the analysis, since they are neither conservative nor innovative with respect to the innovation at hand. A local intercept of 0.000 indicates that this situation was reached in 1950 (given the recoding of dates discussed above). In fact, all locations had negative intercept scores,

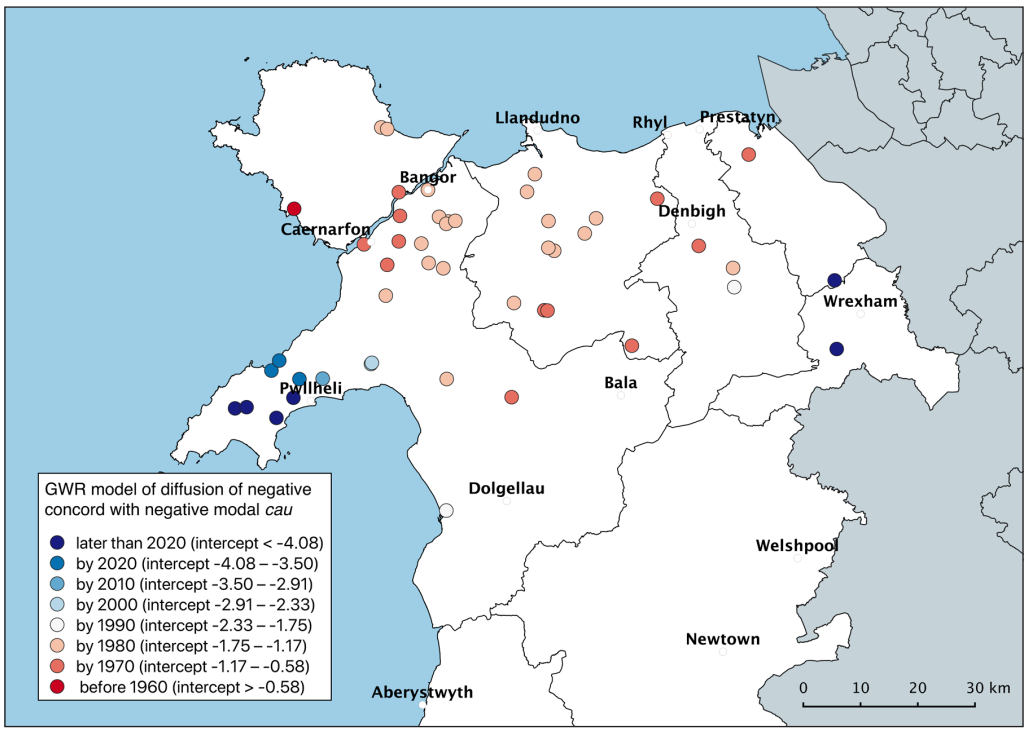



**Figure 3:** Plot of intercept values for a GWR model of the diffusion of negative concord with *cau*.

indicating that this situation was reached only later. Given the model's age coefficient of 0.05825, a local intercept of −0.5825 would indicate that negative concord had become acceptable to the majority ten years later, in 1960, and so on.

The AICc of this model is 75.316. Lower AICc values indicate improved model fit even allowing for the greater complexity of a GWR model. Since this is lower than the AICc of the global model (80.467), we conclude that the geographic model performs substantially better that the comparitor global one set out above, confirming the geographic basis of the variation. Consistent with this, the model also accounts for a substantially higher proportion of the deviance in the data than the global model, 33.6% compared to 13.2% for the global model.

Looking at Figure 3, we see a diffuse, non-wave-like pattern of geospatial diffusion. Early-adopting locations, with majority negative concord by 1970, are found sporadically in both the west (around Caernarfon) and the east (around Bala and Denbigh). Subsequently, a more focused dialect distribution rapidly emerges, so that, by 1980, we find a coherent central zone with majority negative concord. The far west (the Llŷn Peninsula around Pwllheli) and the far east (around Wrexham) remain untouched by the innovation.

These patterns are rather unexpected, fitting neither a wave model (Schmidt 1872: 27; Bailey 1973: 64–109; Britain 2013: 478) nor a gravity model (Hägerstrand 1967; Trudgill 1974) of diffusion. Indeed, one significant problem is that it is difficult to identify a single locus of innovation. While the highest intercept values are in the west, on the island of Anglesey and in Caernarfon, there are also high intercept values further east. It is hardly plausible that the innovation began in the west and jumped to more easterly rural locations before spreading via contagious diffusion to the intervening areas.

Related to this is another feature of the change, namely its speed. An age coefficient of 0.058 indicates a relatively rapid increase over time. It lies at the upper end of the range of values typically found (for instance, Willis 2017: 58 finds values for age coefficients between 0.017 and 0.049 for those syntactic contexts currently undergoing change in the form of the second-person singular pronoun in Welsh). The result of this is that the change achieves a coherent dialect distribution rather quickly after the initial innovation(s).

### 3.5.2 Interaction between negative concord and n-words

Consider next the results for item 42 of the questionnaire, given as (27) above, testing for the availability of n-words in the subject position of *cau*-clauses. The raw data are given in Figure 4 in the same format as Figure 2 above. We see limited acceptance of the structure, concentrated above all around Caernarfon with a few scattered outliers elsewhere. A global logistic regression shows age to be a significant factor at the $p < 0.01$ level, with younger speakers more likely to accept n-words in the subject position of *cau*-clauses. The global model produces an intercept value of −2.06441 logits (standard error = 0.448981, z-score = −4.598061, two-tailed p-value = $4.26 \times 10^{-6}$). This means that an arbitrary speaker born in 1950 has a probability of accepting the construction of 0.113. The age coefficient is 0.043785 logits per year (standard error = 0.016870, z-score = 2.595396, one-tailed p-value = 0.004724). The AICc of this model is 75.726, and 9.6% of deviance is accounted for.

Again, a GWR model is more successful at accounting for variation and allows us to see a pattern of diffusion. This was done in the same way as above. The bandwidth was kept at 12.36 km as before to ensure comparability with the model for negative concord overall, even though a somewhat higher bandwidth would have produced a slightly better fit. In practice, varying the bandwidth did not substantially alter the geospatial patterns of

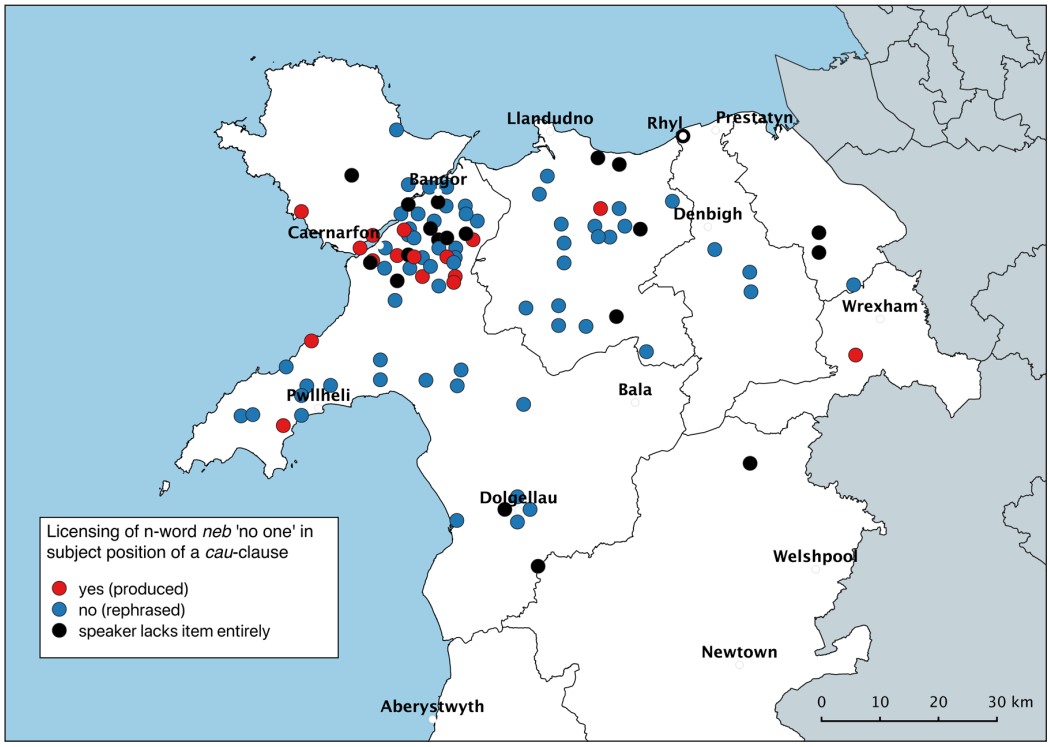



**Figure 4:** Distribution of the availability of n-word *neb* 'no one' in subject position with *cau* in the *SAWD* data.

diffusion produced by the model.[18] The age coefficient of this model is 0.041808 logits per year (standard error = 0.019316, z-score = 2.164408, one-tailed p-value = 0.015217). The AICc of this model is 74.504638, a modest improvement over the global model, and 26.2% of deviance is accounted for.

The age coefficient is very similar to that found above for the overall diffusion of negative concord (0.042 logits per year compared to 0.058), suggesting that both innovations are progressing at more or less the same rate, the age coefficient being the slope of the S-curve. Such close S-curve slopes are normally considered (as a result of the Constant Rate Hypothesis) to be evidence that the changes in question reflect a single underlying innovation (Kroch 1989a: 153–163; 1989b; 2001: 720–721 etc.).

However, the rise of this construction is rather later in time than negative concord overall. This can be seen clearly from the plot of intercept values across geographic space given in Figure 5. Compared to Figure 3, we typically see a 20-year time lag in the model in the emergence of acceptability of n-words with *cau*. Furthermore, the geospatial pattern of diffusion is more wave-like. A clear origin is observable around Caernarfon, and the extent to which the innovation has progressed decreases broadly in proportion to the distance from this area.

### 3.5.3 Licensing of VP-ellipsis (*cau*-stranding)

Recall that item 34 of the questionnaire, given in (30) above, tests for the innovation of licensing of VP-ellipsis after *cau* (*cau*-stranding). The distribution of acceptance of this structure, given in Figure 6, is even more limited, with only 10 informants repeating it faithfully in their interview. The majority of informants either added *gwneud* 'do'

[18] The best-performing model had a bandwidth of 17.1 km, and yielded an AICc of 73.990732 with 19.8% of deviance accounted for. It produced an age coefficient of 0.038499 logits per year (standard error = 0.017706, z-score = 2.174371, one-tailed p-value = 0.014841).

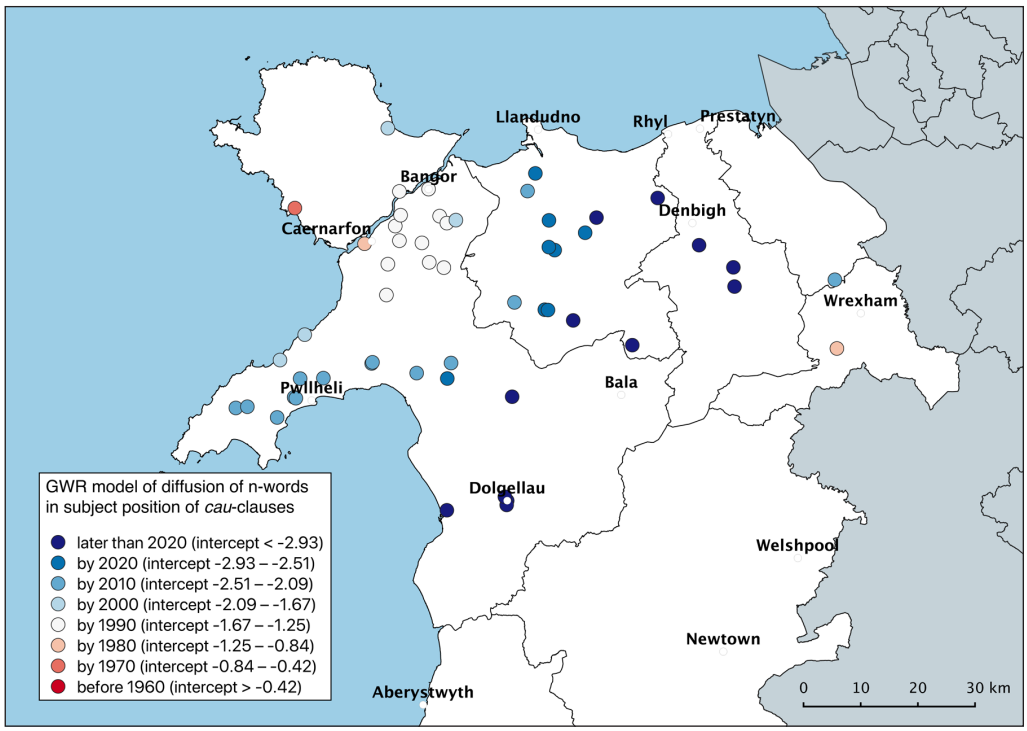



**Figure 5:** Plot of intercept values for a GWR model of the diffusion of the acceptability of n-words in the subject position of *cau*-clauses.

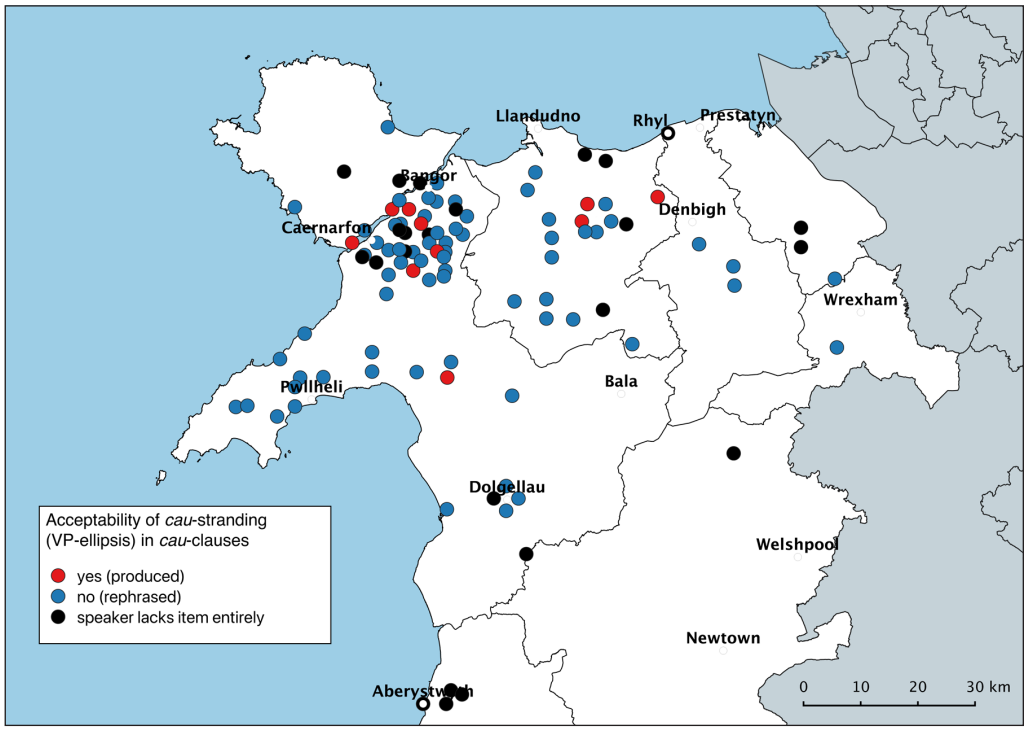



**Figure 6:** Distribution of the availability of VP-ellipsis (*cau*-stranding) in *cau*-clauses in the *SAWD* data.

as a pro-VP ('…but he's refusing to do.') or rephrased without the modality ('…but he hasn't.'). This limited acceptance makes it difficult to be confident of the age and dialect distribution of the feature and the following comments should be treated with some caution. Dialectally, acceptance is spread fairly widely, albeit at fairly low levels, across the core central area for *cau*.

A global logistic regression shows age to be a significant factor at the $p < 0.05$ level, with younger speakers more likely to accept the construction. The global model produces an intercept value of $-2.57803$ logits (standard error $= 0.54912$, z-score $= -4.695$, two-tailed p-value $= 2.67 \times 10^{-6}$). This means that an arbitrary speaker born in 1950 has a probability of accepting the construction of 0.071. The age coefficient is 0.039156 logits per year (standard error $= 0.01974$, z-score $= 1.983$, one-tailed p-value $= 0.0237$). The AICc of this model is 59.592, and 7.2% of deviance is accounted for.

A GWR model was performed as before with the same bandwidth of 12.36 km. This is plotted in Figure 7. For this structure, the GWR model does not perform better than the global model: the distribution of informants that accept the construction does not show sufficient spatial autocorrelation. While the GWR model performs somewhat better in terms of overall data fit, its AICc is worse (64.382223 compared to 59.592214 above) due to its increased complexity compared to the global model.[19] The age coefficient of the model is 0.036776 logits per year (standard error $= 0.021800$, z-score $= 1.686956$, one-tailed p-value $= 0.045811$), and 19.4% of deviance is accounted for.

Note that the age coefficient for this innovation (0.037 logits per year) is somewhat lower than that for the diffusion of acceptability of n-words in subject position of *cau*-clauses (0.042 logits per year) and much lower than that for the diffusion of negative concord itself (0.058 logits per year). Indeed, it lies just outside the standard error of the age coefficient of the diffusion of negative concord, which suggests that this is not a manifestation of the same change. We will see later that evidence from implicational hierarchies of speaker responses supports this conclusion.

Finally, the pattern of diffusion here is weak. There is little geospatial variability in the data and, in so far as we can tell, the innovation seems to begin at much the same time and progresses to much the same extent across much of the central *cau* region.

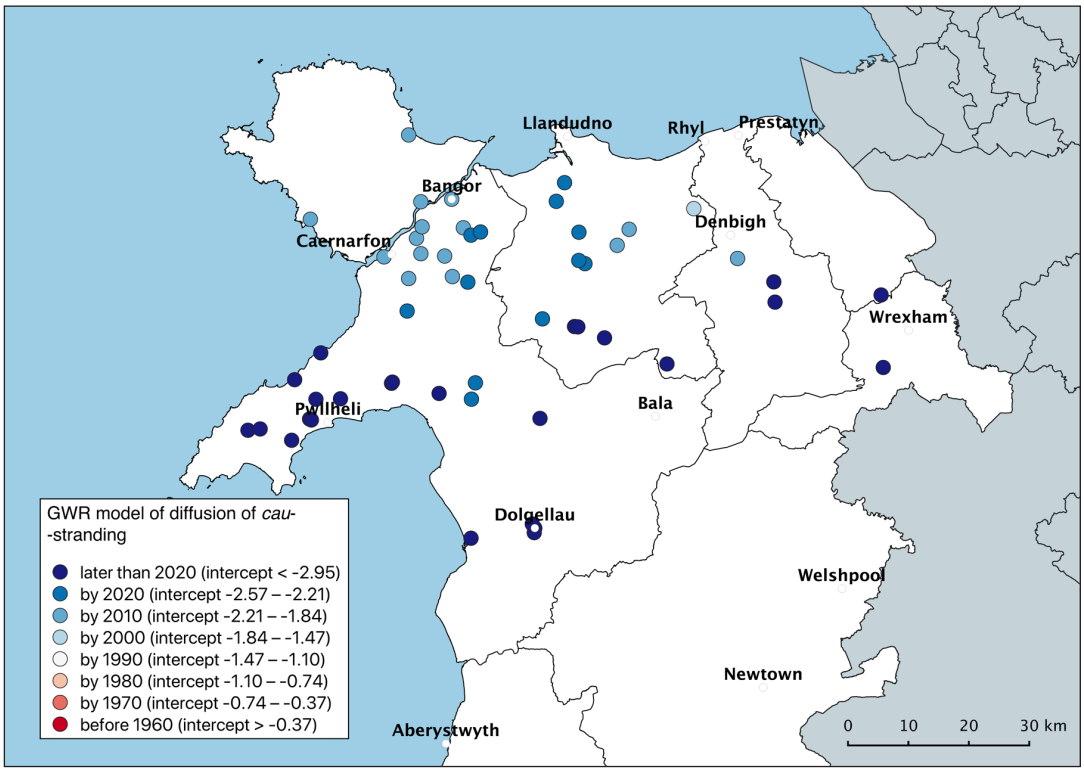



**Figure 7:** Plot of intercept values for a GWR model of the diffusion of acceptability of *cau*-stranding.

[19] The optimal bandwidth, as determined by AICc minimization, is thus the entire area of study.

#### 3.5.4 Strong negative concord

The final structure tested was that of strong negative concord, the insertion of the sentential negation marker *ddim* into the *cau*-construction. This was accepted by only one speaker. However, this speaker did it consistently, both with item 41 (the question concerning this structure, (26) above) and with other items containing *cau* in the questionnaire. This suggests a real phenomenon, and one that it will be interesting to watch over the coming years. However, at present, there is little that can be said about it within the current analysis.

### *3.6 Implicational hierarchies*

Another way of making sense of the variation is to look at the grammars of individual speakers, to see what combinations of structures they accept. Of the 96 speakers in the northern part of the questionnaire, 19 lacked *cau* entirely. The distribution of the grammars of the remaining 77 speakers is given in Table 1. Numbers in parentheses refer to the number of speakers who used negative concord actively (adding it to at least one item where it was not offered) rather than passively (merely repeating it as offered). Combinations not listed were not attested.

Note that one of the advantages of the systematic investigation with a structured questionnaire carried out for a syntactic dialect atlas is that it allows grammatical systems for individual speakers to be established in this way. This would not be so readily possible for data from a spoken corpus, since most speakers present in a corpus will provide relevant data only for a subset of environments, possibly only one.

As Table 1 shows, the possible combinations within a single grammar are rather constrained. These combinations yield the implicational hierarchy (implicational scale) given in (31). This shows, for instance, that a speaker with n-words in subject position of *cau*-clauses always has weak negative concord with *cau*, while the relationship between *cau*-stranding and allowing n-words in subject position of *cau*-clauses is free and neither implicates the other.

(31) weak neg. concord < n-words in subject position < strong neg. concord
< stranding

This distribution of grammars suggests that the emergence of *cau*-stranding is independent of the development of negative concord, while the other innovations are all syntactically interconnected. This is consistent with the suggestive evidence from the rate of change of the GWR model developed above, in which we found the rate of change in *cau*-stranding to be substantially slower than in the other two cases examined.

In the next section, we will develop a formal syntactic account of the innovation consistent with these empirical observations.

**Table 1:** Distribution of grammars among the speakers surveyed for *SAWD*.

| Speaker has weak negative concord? | Speaker allows n-words in subject position of *cau*-clauses? | Speaker has stranding? | Speaker has strong negative concord? | Number of speakers |
|---|---|---|---|---|
| – | – | – | – | 40 |
| + | – | – | – | 19 (10) |
| + | + | – | – | 9 (3) |
| + | – | + | – | 5 |
| + | + | + | – | 4 |
| + | + | + | + | 1 |

## 4 A formal analysis

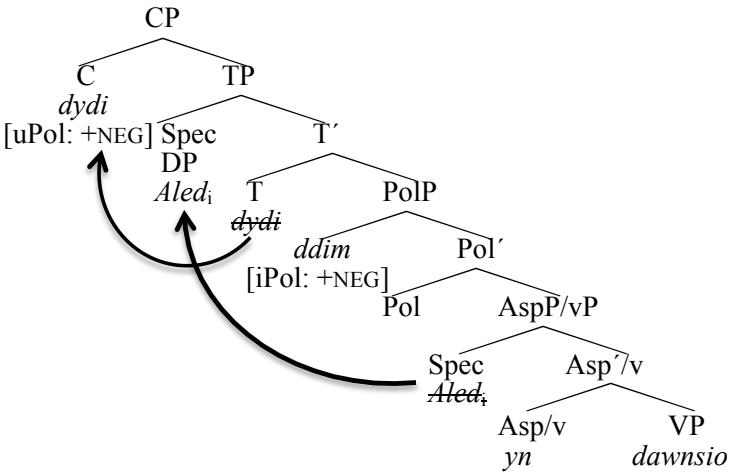
### 4.1 Background assumptions about Welsh grammar

Declarative main clauses in Welsh manifest VSO word order. Where the inflected verb is an auxiliary, this is manifested as AuxSVO. In generative analyses, this word order is assumed to be derived by head movement of the finite verb (whether an auxiliary or a lexical verb) from V to some projection beyond TP. This position is Aux for Jones & Thomas (1977: 19–23), I (Infl) for Sproat (1985) and Rouveret (1994: 51–90), and AgrS for Willis (1998: 19–26) and Roberts (2005: 7–48). The evidence for this movement comes from the position of the finite verb and subject, both of which precede the negative marker:

(32) Dydi Aled ddim yn dawnsio.
NEG.be.PRS.3SG Aled NEG PROG dance.INF
'Aled isn't dancing.'

Given this ordering, we must assume that both subject and inflected verb raise out of the VP beyond the polarity phrase, PolP, to a position preceding the negative marker. For the sake of concreteness, it is assumed here that this movement takes subjects, being projected initially to the specifier of an aspectual light verb projection labelled here as AspP/vP, to Spec, TP, the projection immediately dominating PolP. Consistent with this, the verb must raise via T to C. This would be a low head position, such as Fin, in a system in which C was split into multiple projections (Rizzi 1997). This system is shown in the tree in (33), representing the sentence in (32).

(33)




This commits us to the view that preverbal particles, such as the main-clause affirmative marker *mi* in (34), in so far as they survive in spoken Welsh, are proclitics or prefixes to the finite verb and thus do not block movement of the verb into the C-domain (cf. Willis 2007: 437).

(34) (Mi) naeth Aled ddawnsio.
PRT do.PST.3SG Aled dance.INF
'Aled danced.'

Adverb placement suggests that some non-specific indefinite subjects remain in AspP/vP; however, this is not relevant to the current discussion. Aspectual particles such as the unmarked particle *yn* or the perfect particle *wedi* are the spellout of the head of this projection. For further detail on the system of Welsh clause structure assumed here, see Borsley, Tallerman & Willis (2007: 48–52).

Furthermore, I assume an account of negative concord in Welsh that extends and adapts Zeijlstra's (2004) account of negative concord in a range of mostly Romance, Germanic

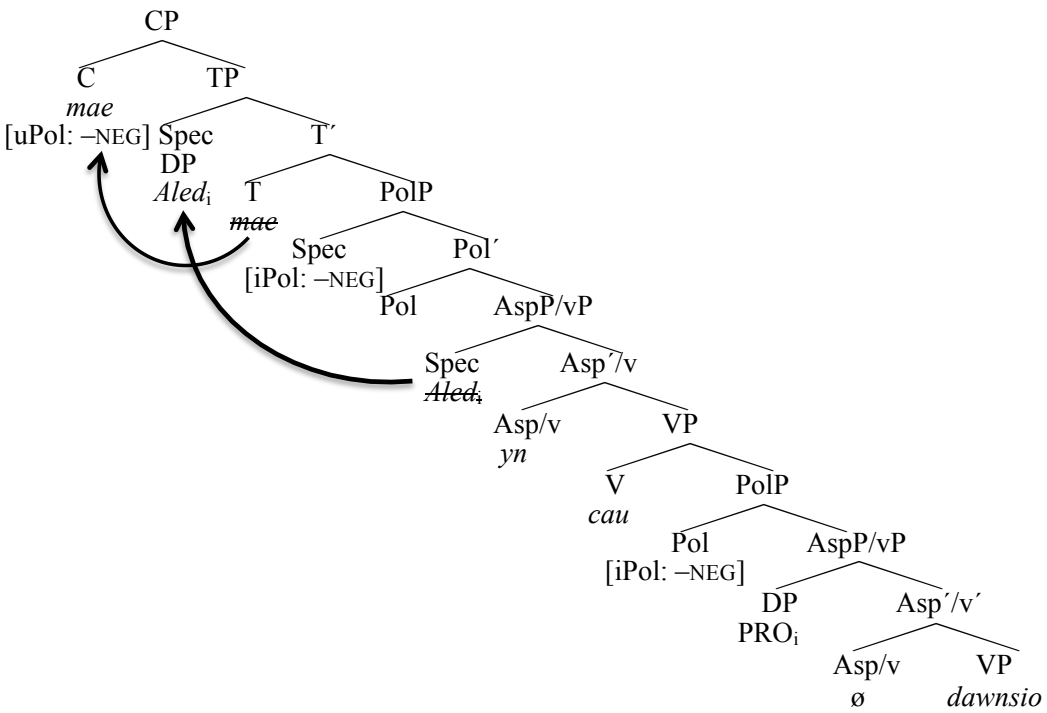
and Slavic languages. On this approach, in Welsh, a clause is negative if Spec, PolP contains an interpretable negative operator [iPol: +NEG], generally spelled out as *ddim*. Other elements that participate in the negative concord system bear uninterpretable features that agree (including upwards Agree) with this element. The auxiliary *bod* 'be' will therefore be specified either as [uPol: +NEG], spelled out as *dydi*, if the negative operator is present, or as [uPol: –NEG], spelled out as *mae*, if it is not. This accounts for the difference between (3) and (4) above.

Welsh indefinites of the main *neb*-series (*neb* 'anyone, no one', *dim byd* 'anything, nothing' etc.) are n-words in the sense of Laka (1990: 107–9): they convey a negative meaning when used as sentence fragments, but must be accompanied by some other negative element in a negative-concord configuration in finite clauses. On the approach adopted here, these n-words are reflexes of indefinites that have undergone agreement with this operator. So, in (7), *neb* is the spellout of the [+human] indefinite that is [uPol: +NEG] by virtue of having undergone agreement with *ddim*, which is marked [iPol: +NEG] in Spec, PolP. In (8), the same process applies, except that the negative operator in Spec, PolP is not spelled out as *ddim*, but rather is covert. Spec, PolP is generally spelled out as *ddim*, but the spellout rules allow for a sequence of two negative elements to be spelled out as one: *ddim neb* 'not no one' is spelled out as *neb* 'no one' and *ddim dim* 'not nothing' is spelled out as *ddim* 'nothing'. The exact conditions in which this happens and the mechanism for producing it are not immediately relevant here.

### 4.2 A formal analysis of the development of cau

Having laid out the relevant background assumptions, we can now consider the status of *cau* at the more conservative stage. In the initial grammatical system, attested in the nineteenth century, *cau* is a phonologically reduced form of *nacáu* 'refuse', a control verb of a familiar type. The structure is therefore biclausal, as given in (35), representing a typical example sentence such as (36).

(35)



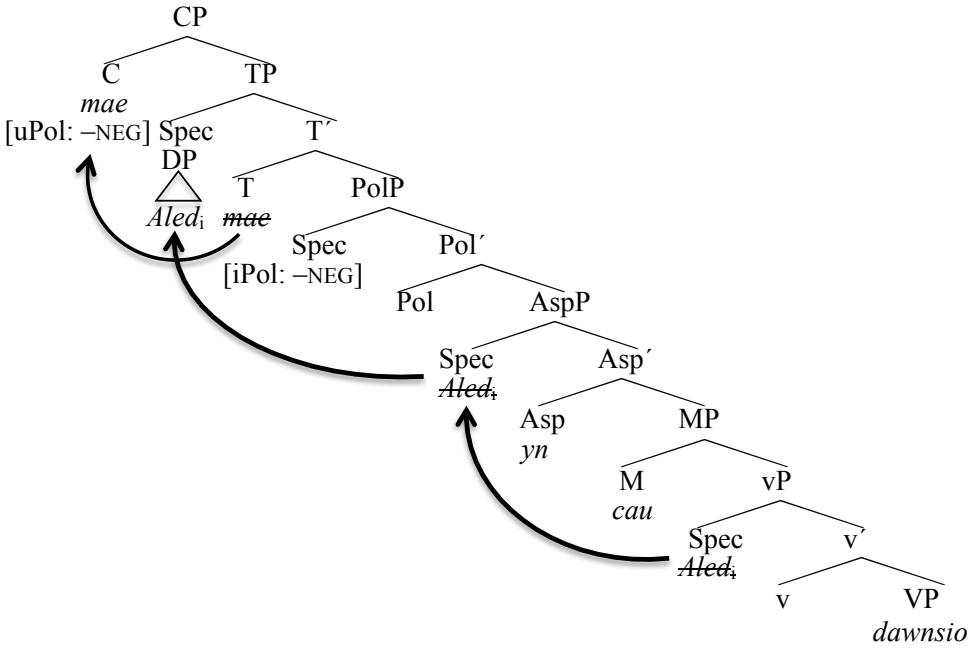
(36) Mae Aled (yn) cau dawnsio.
be.PRS.3SG Aled PROG refuse.INF dance.INF
'Aled is refusing to dance.'

In this grammar, *cau* itself has full argument structure, introducing an agent external argument, *Aled*, in Spec, AspP/vP in (35). By projecting an agent, it imposes the usual semantic restrictions on this argument, namely that it is volitional and thus capable of functioning as an agent. Each of the two clauses has its own independent polarity value. Formally, polarity is represented in (35) by a polarity projection PolP, whose head bears the feature [–NEG]. Negative concord on the finite verb is treated as the morphological reflex of a polarity feature on T. *Cau* is a semantically adversative predicate with negative connotations, comparable to adversative predicates in English such as *refuse*, *deny*, *be afraid to*, *be reluctant to* (Klima 1964: 313–14), which license negative polarity items in their complements. However, it is not grammatically negative and therefore does not trigger negative concord within its clause: polarity items are semantically licensed, while negative concord is based on a syntactic configuration. The Pol head in the main clause in (35) is therefore marked as [–NEG]. The unvalued Pol feature on the auxiliary in T agrees with this, taking the value [Pol: –NEG], before itself moving to C. The spellout of the auxiliary is therefore the [–NEG] form, *mae*.

Finally, the structure is a biclausal control structure. This is represented as a second predication in a complement AspP/vP with PRO as the external argument of *dawnsio* 'dance'. Nothing crucial, however, depends on the exact representation of control and additional functional projections in the complement or a movement analysis of control (Hornstein 1999) could be accommodated without changing the essential intuition.

Having established an analysis for the historical input system, we can now consider the syntactic representation of the reanalysis. Reanalysis of *cau* as a modal is reflected in all the dialects under investigation here. The new structure is shown in (37).

(37)




In the new grammar, the structure is acquired as monoclausal. *Cau* is no longer a lexical verb, and does not introduce an external argument of its own. Instead, it effectively acts as a raising verb (assuming auxiliaries in general essentially to be a type of raising verb)

with a modality feature (identified by its treatment as a modality head M in (37)). The external argument is merged as the specifier of vP and moves to become the subject in Spec, TP, while the finite auxiliary *mae* 'is' moves from T to C. Less structure is projected than previously, as is typical of reanalysis involved in grammaticalization. This reanalysis is purely a case of auxiliation, interpreted as the failure of acquirers to posit the entirety of the syntactic structure of their input data (loss of structure, and, in a sense, upwards reanalysis, cf. Willis 1999; 2000 on the Russian conditional; Roberts & Roussou 2002; 2003: 35–72 on the Romance and Greek future and English modals). Its surface effects are above all semantic, in that the emergent transparency of the new modal to the assignment of argument structure means that semantic restrictions on the subject are relaxed to include only those imposed by the lexical verb. Thus sentences like (2) above, with non-agentive subjects, are possible if the lexical verb is semantically consistent with such subjects. These are possible in all contemporary dialects with *cau*; hence, we can be confident that the reanalysed structure has been adopted in all these dialects.

Polarity here operates in the same way as before: T bears an uninterpretable polarity feature, which is valued under Agree with an interpretable polarity feature on the Pol head. In this case, the Pol head retains the [–NEG] value of the previous generation's grammar, presumably on the basis of observed instances of affirmative rather than negative forms of the verb in clauses containing *cau*.

Note that, in this grammar, n-words are ungrammatical in subject position:

(38) *Mae neb (yn) cau dawnsio.
be.PRS.3SG n-one PROG CAU dance.INF
('No one will dance.')

The structure of (38) will be the same as that in (37), with the appropriate lexical substitutions. This structure fails to spell out as (38), however, because the source of the polarity is lexical rather than grammatical in this structure. Consequently, the subject indefinite is marked as [uPol: –NEG]. Since *neb* is the spellout of [uPol: +NEG], it fails to be generated. Since there is no way to generate (38) in this grammar, this analysis predicts that no dialect can have n-words in *cau*-clauses (without a double negation interpretation). This is indeed what was found in sections 3.5.2 and 3.6 above.

Next, we turn to the grammatical system with innovated negative concord. Here, we can suggest that the negative semantic element of *cau* has grammaticalized, that is, gone from being represented lexically in the semantic representation of *cau*, to being represented in the grammatical specification of the item (cf. van Gelderen's 2015 postulation of the diachronic linguistic cycle as a shift in feature status from semantic to interpretable to uninterpretable). Syntactically, this means that it gains an [iPol: +NEG] feature, which triggers negative concord throughout the clause. One problem with this is that it predicts that n-words, being the reflex of agreement with [iPol: +NEG] should immediately become available in such a grammar. We have seen that only a subset of speakers with negative concord allow n-words in *cau*-clauses. Further research is needed to establish whether this is a function of the particular sentences in the questionnaire or a wider finding, motivating either a more fine-grained analysis or a treatment in terms of the gradual actualization of the innovations, with n-words being a relatively resistant context.[20]

[20] One solution, if actualization turns out to be stepwise, would be to propose, in the spirit of Haegeman & Lohndal's (2010) binary Agree analysis of West Flemish negative concord, that *cau* initially bears an additional feature that can only form an Agree relation with tensed (rather than nominal) elements. This feature would be absent from more innovative speakers' grammars.

What triggers this shift to integration of *cau* into the polarity system? On the one hand, it is of course expected from the general direction of grammaticalization and is accounted for in the same way that this pattern is accounted for more generally. However, specific to this particular instance is the syntactic analogy of other quasi-negative elements in the Welsh grammatical system. Parallel variation in negative concord exists with other negative modals and aspect markers in Welsh dialects, specifically, in (39), the negative modal *methu* 'be unable', and, in (40), the negative perfect aspect marker *heb* (< preposition *heb* 'without').

(39) Mae/Dydi o 'n methu dawnsio.
be.PRS.3SG/NEG.be.PRS.3SG he PROG be.unable.INF dance.INF
'He can't dance.'

(40) Mae/Dydi hi heb fynd eto.
be.PRS.3SG/NEG.be.PRS.3SG he NEG.PRF go.INF yet
'She hasn't gone yet.'

The dialect distribution of this feature has never been investigated, but it seems to be widely distributed geographically.

## 5 Integrating formal syntax and dialect geography

Syntactically, this is a moderately abstract innovation, requiring reference in speakers' grammars to uninterpretable agreement features in negative concord. When *cau* is lexical, negation is a semantic feature of its lexical meaning and thus is not represented syntactically. Once it is reanalysed as a modal auxiliary, it contributes formal interpretable modal and negative features to the derivation (e.g. [iPol: +NEG]). This development follows a pathway of change from semantic to interpretable syntactic features (van Gelderen 2015).

Negative concord targets the newly emerged interpretable negative feature of *cau*, making the absence of negative concord a morphosyntactic exception and thus difficult to acquire. Acquirers thus posit negative concord with this item even though they are presented with moderately strong positive evidence for its absence. This example also suggests that consistent presence or absence of negative concord is the preferred option and that mixed, partial negative-concord systems are dispreferred in language change and thus typologically.

The dialect pattern associated with this sort of innovation seems to be a diffuse one. We saw above (section 3.5) that negative concord with *cau* is spreading, being commoner in younger speakers, but that its distribution is quite diffuse: it is found in some speakers across most of the north, but it is hard to identify its heartland. This can be interpreted as meaning that it innovated in more than one location, and that new speakers who adopt it must do so by reanalysing the system in the same way as the original innovators. We can term this phenomenon "multiple reactuation" of change.

Contrast this with the distribution of other less abstract innovations, which show contiguous patterns from the very beginning. For instance, another variable in the pilot for the *Syntactic Atlas of Welsh Dialects*, the presence of absence of a new second-person singular pronoun, namely *chdi*, shows very clear wave-like patterns of diffusion in all age groups (Willis 2017). A morphosyntactic innovation such as this is straightforwardly acquirable on the basis of positive evidence, and thus unlikely to be the result of multiple reactuation in different locations.

As Table 1 shows, one speaker consistently used strong negative concord with *cau*, that is, negative concord on the auxiliary alongside insertion of the sentential negator *ddim*,

as in (26) above. This innovation fits naturally into the analysis proposed here if we treat it as a further step along the path of feature changes that are associated with grammaticalization. Grammaticalization often seems to involve a shift in status of features from interpretable to uninterpretable, a shift which can be thought of as a move in the direction of features becoming purely formal rather than expressing semantic distinctions (van Gelderen 2015). In the current case, a shift in the status of *cau* from [iPol: +NEG] to [uPol: –NEG] is therefore a natural one. A speaker with such a grammar will need an interpretable negation in Spec, PolP. Such negation is normally spelled out as *ddim*. Emergence of the sequence *ddim cau* is therefore expected as the simplest option if this final stage of grammaticalization is reached. Clearly, however, data from additional speakers will be needed to confirm or reject this suggestion.

Finally, we established empirically that availability of *cau*-stranding is the result of an independent change: it does not spread at the same rate as the other innovations discussed here, and it did not participate in the same implicational hierarchy in (31) as the other innovations. Innovation of *cau*-stranding is logically independent of innovation of negative concord and negative concord is not a prerequisite for it. This is in contrast to the case with innovation of n-words in subject position, which requires innovation of negative concord to create the necessary licensing environment.

If we adopt the position suggested above that ellipsis is licensed (only) of the complement of v, then this must reflect variation in the status of *cau*. One possibility is that, for speakers who allow *cau*-stranding, *cau* is an element that spells out Asp, M and v in (37), and thus qualifies as a v-element that licenses ellipsis. Another way of thinking of it is that it reflects interspeaker variation in the set of heads that license ellipsis. Speakers vary in their acceptance of ellipsis after both aspectual head and the predicative head *yn*, the structures with ellipsis all being relatively recent innovations:

(41) %Doedd Mari ddim yn chwarae heddiw ond roedd Ifan yn.
NEG.be.IMPF.3SG Mari NEG PROG play.INF today but be.IMPF.3SG Ifan PROG
'Mari wasn't playing today but Ifan was.'

(42) %Dydi Mari ddim wedi dod heddiw ond mae Ifan wedi.
NEG.be.PRS.3SG Mari NEG PRF come.INF today but be.PRS.3SG Ifan PRF
'Mari hasn't come today, but Ifan has.'

(43) %Dydi Mari ddim yn hwyr ond mae Ifan yn.
NEG.be.PRS.3SG Mari NEG PRED late but be.PRS.3SG Ifan PRED
'Mari isn't late but Ifan is.'

Variation with *cau* belongs in with this variation and would be expected to show the same kinds of dialectal and sociolinguistic patterns.

## 6 Conclusions

This paper has introduced the pilot fieldwork for the *Syntactic Atlas of Welsh Dialects*, and shown by close examination of a single case study involving negative concord and the grammaticalization of a new modal how that work can yield a fruitful interaction between dialectological research and more theoretical work in syntax and language change. Specifically, this work suggests the following conclusions:

(i) more abstract syntactic innovations, such as those that involve a change in the status of a feature from interpretable to uninterpretable (cf. van Gelderen's 2015

notion of Feature Economy), are more likely to manifest multiple reactuation, that is, diffuse patterns of innovation across a wide area, resulting from multiple points of innovation, rather than diffusing out in either a wave-like pattern or via hierarchical diffusion from a single point of innovation;

(ii) as a minimum, our syntactic analyses need to be consistent with patterns in dialect distributions of variants and implicational hierarchies in speakers' use of those variants: a syntactic analysis in which feature x can only be found in a speaker's grammar if feature y is also present is to be preferred if such a relationship holds among the variation found in inter-dialect and inter-speaker variation;

(iii) the speed of diffusion of an innovation may be used to establish whether it forms a sub-part of a larger innovation, and thus whether a syntactic analysis which treats it as such is to be preferred or dispreferred.

## Acknowledgements

I am grateful to Maggie Tallerman and Bob Borsley for their support for the project, to Laura Arman and Peredur Webb-Davies for help with grammaticality judgments, and to fieldworkers and transcribers Jonathan Stammers, Elidir Dafydd, Mared Gwyn, Mair Roberts, Swyn Haf and Owain Rhys James. Parts of this research were presented at the British Academy Development of the Welsh Language network, Cardiff, 2013; the 20th Welsh Linguistics Seminar, Gregynog, 2013; the 21st International Conference on Historical Linguistics, Oslo, 2013; the 8th Celtic Linguistics Conference, Edinburgh, 2014; Cambridge Comparative Syntax 3, Cambridge, 2014; Formal Ways of Analyzing Variation 4, York, 2017; Theoretical and Empirical Approaches to Microvariation (TEAM), Padua, 2017; and the Linguistics Association of Great Britain Annual Meeting, Canterbury, 2017. I am grateful to these audiences for their comments and feedback.

## Funding Information

This research was funded in part by grants from the British Academy ("Syntactic variation in contemporary Welsh", British Academy Mid-career Fellowship Grant no. MC110579) and the Isaac Newton Trust. Their support is hereby gratefully acknowledged.

## Competing Interests

The author has no competing interests to declare.

## References

Altmann, Gabriel, Haro von Buttlar, Walter Rott & Udo Strauss. 1983. A law of change in language. In Barron Brainerd (ed.), *Historical linguistics*, 104–15. Bochum: Brockmeyer.

Bailey, Charles-James N. 1973. *Variation and linguistic theory*. Washington DC: Center for Applied Linguistics.

Barbiers, Sjef & Hans Bennis. 2007. The syntactic atlas of the Dutch dialects: A discussion of choices in the SAND-project. *Nordlyd* 34. 53–72. DOI: https://doi.org/10.7557/12.89

Barbiers, Sjef, Hans Bennis, Gunther De Vogelaer, Magda Devos & Margreet van der Ham. 2005. *Syntactische atlas van de Nederlandse dialecten, deel I* [Syntactic atlas of the Dutch dialects, volume I]. Amsterdam: Amsterdam University Press.

Barbiers, Sjef, Johan van der Auwera, Hans Bennis, Eefje Boef, Gunther De Vogelaer & Margreet van der Ham. 2008. *Syntactische atlas van de Nederlandse dialecten, deel II* [Syntactic atlas of the Dutch dialects, volume II]. Amsterdam: Amsterdam University Press. DOI: https://doi.org/10.5117/9789053567791

Benincà, Paola & Cecilia Poletto. 2007. The ASIS enterprise: A view on the construction of a syntactic atlas for the northern Italian dialects. *Nordlyd* 34. 35–52.

Blythe, Richard A. & William Croft. 2012. S-curves and the mechanisms of propagation in language change. *Language* 88. 269–304. DOI: https://doi.org/10.1353/lan.2012.0027

Borsley, Robert D. & Bob Morris Jones. 2005. *Welsh negation and grammatical theory*. Cardiff: University of Wales Press.

Borsley, Robert D., Maggie Tallerman & David Willis. 2007. *The syntax of Welsh*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486227

Britain, David. 2013 [2002]. Space, diffusion and mobility. In J. K. Chambers & Natalie Schilling (eds.), *Handbook of language variation and change*, 471–500. Chichester: Wiley–Blackwell.

Bucheli, Claudia & Elvira Glaser. 2002. The Syntactic Atlas of Swiss German Dialects: Empirical and methodological problems. In Sjef Barbiers, Leonie Cornips & Susanne van der Kleij (eds.), *Syntactic microvariation*, 41–74. Amsterdam: Meertens Institute.

Bybee, Joan & William Pagliuca. 1985. Cross-linguistic comparison and the development of grammatical meaning. In Jacek Fisiak (ed.), *Historical semantics – Historical word-formation*, 59–83. Berlin: Mouton.

Cornips, Leonie & Cecilia Poletto. 2005. On standardising syntactic elicitation techniques (part 1). *Lingua* 115. 939–57. DOI: https://doi.org/10.1016/j.lingua.2003.11.004

Cornips, Leonie & Willy Jongenburger. 2001. Elicitation techniques in a Dutch syntactic dialect atlas project. *Linguistics in the Netherlands*, 53–63.

Denison, David. 2003. Log(ist)ic and simplistic S-curves. In Raymond Hickey (ed.), *Motives for language change*, 54–70. Cambridge: Cambridge University Press.

Deuchar, Margaret & Peredur Davies. 2014. Documentation file for the Bangor *Siarad* Corpus. Bangor: Bangor University.

Deuchar, Margaret, Peredur Davies & Kevin Donnelly. 2018. *Building and using the Siarad Corpus*. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/scl.81

Fleischer, Jürg, Alexandra N. Lenz & Helmut Weiss. 2017. *SyHD-atlas*. Philips-Universität Marburg, Universität Wien, Goethe Universität Frankfurt am Main. http://www.syhd.info/.

Gallego, Ángel J. 2009. Ellipsis by phase. *Paper presented at the 19th Colloquium on Generative Grammar, Euskal Herriko Unibertsitatea*. Vitoria-Gasteiz.

Glaser, Elvira & Gabriela Bart. 2011. Discovering and mapping syntactic areas: Old and new methods. In Xosé Afonso Álvarez Pérez, Ernestina Carrilho & Catarina Magro (eds.), *Proceedings of the International Symposium on Limits and Areas in Dialectology (LimiAr)*, 345–63. Lisbon: Centro de Linguística da Universidade de Lisboa.

Greenberg, Joseph H., Charles E. Osgood & Sol Saporta. 1954. Language change. In Charles E. Osgood & Thomas A. Sebeok (eds.), *Psycholinguistics: A survey of theory and research problems*, 146–63. Baltimore: Waverly Press.

Hägerstrand, Torsten. 1967. *Innovation diffusion as a spatial process*. Chicago, IL: University of Chicago Press.

Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30. 69–96. DOI: https://doi.org/10.1162/002438999553968

Jones, Bob Morris. 1999. *The Welsh answering system*. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110800593

Jones, Morris & Alan R. Thomas. 1977. *The Welsh language: Studies in its syntax and semantics*. Cardiff: University of Wales Press for the Schools Council.

King, Gareth. 1993. *Modern Welsh: A comprehensive grammar*. London: Routledge.

Klima, Edward S. 1964. Negation in English. In Jerry A. Fodor & Jerrold J. Katz (eds.), *The structure of language*, 246–323. Englewood Cliffs, NJ: Prentice-Hall.

Kroch, Anthony S. 1989a. Function and grammar in the history of English: Periphrastic *do*. In Ralph W. Fasold & Deborah Schiffrin (eds.), *Language change and variation*, 133–72. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/cilt.52.09kro

Kroch, Anthony S. 1989b. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1. 199–244. DOI: https://doi.org/10.1017/S0954394500000168

Kroch, Anthony S. 2001. Syntactic change. In Mark R. Baltin & Chris Collins (eds.), *The handbook of contemporary syntactic theory*, 699–729. Malden, MA: Blackwell. DOI: https://doi.org/10.1002/9780470756416.ch22

Laka, Itziar. 1990. *Negation in syntax: On the nature of functional categories and projections*. Cambridge, MA: Massachusetts Institute of Technology PhD dissertation.

Merchant, Jason. 2001. The syntax of silence: Sluicing, islands, and the theory of ellipsis. Oxford: Oxford University Press.

Nakaya, Tomoki. 2014 [2009]. *GWR4 user manual*. https://geodacenter.asu.edu/drupal_files/gwr/GWR4manual.pdf.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar: Handbook of generative syntax*, 281–337. Dordrecht: Kluwer. DOI: https://doi.org/10.1007/978-94-011-5420-8_7

Roberts, Ian G. 2005. Principles and parameters in a VSO language: A case study in Welsh. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780195168211.001.0001

Roberts, Ian G. & Anna Roussou. 2002. The history of the future. In David W. Lightfoot (ed.), *Syntactic effects of morphological change*, 23–56. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199250691.003.0002

Roberts, Ian G. & Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486326

Rouveret, Alain. 1994. *Syntaxe du gallois*. Paris: CNRS Éditions.

Rouveret, Alain. 2012. VP ellipsis, phases and the syntax of morphology. *Natural Language and Linguistic Theory* 30. 897–963. DOI: https://doi.org/10.1007/s11049-011-9151-3

Schmidt, Johannes. 1872. *Die verwantschaftsverhältnisse der indogermanischen sprachen*. Weimar: Hermann Böhlau.

Sproat, Richard. 1985. Welsh syntax and VSO structure. *Natural Language and Linguistic Theory* 3. 173–216. DOI: https://doi.org/10.1007/BF00133840

Thomas, Alan R. 1973. *The linguistic geography of Wales: A contribution to Welsh dialectology*. Cardiff: University of Wales Press.

Thomas, Alan R. 2000. *Welsh dialect survey*. Cardiff: University of Wales Press.

Thomas, R. J. (ed.). 1950–2002. *Geiriadur Prifysgol Cymru: A dictionary of the Welsh language*. Cardiff: Gwasg Prifysgol Cymru.

Thoms, Gary, David Adger, Caroline Heycock & Jennifer Smith. 2017. Shades of negative concord in dialects of Scots. *Paper presented at the LAGB Annual Meeting*. University of Kent.

Trudgill, Peter. 1974. Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society* 3. 215–46. DOI: https://doi.org/10.1017/S0047404500004358

Van Craenenbroeck, Jeroen. 2010. *The syntax of ellipsis: Evidence from the Dutch dialects*. New York: Oxford University Press.

Van Gelderen, Elly. 2015. Semantic and formal features: Feature economy in language change. In Chiara Gianollo, Agnes Jäger & Doris Penka (eds.), *Language change at the syntax–semantics interface*, 33–56. Berlin: De Gruyter.

Willis, David. 1998. *Syntactic change in Welsh: A study of the loss of verb-second*. Oxford: Clarendon Press.

Willis, David. 1999. The structure of Old Russian periphrastic verbal constructions. In István Kenesei (ed.), *Crossing boundaries: Advances in the theory of central and eastern European languages*, 45–65. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/cilt.182.05wil

Willis, David. 2000. Verb movement in Slavonic conditionals. In Susan Pintzuk, George Tsoulas & Anthony Warner (eds.), *Diachronic syntax: Models and mechanisms*, 322–48. Oxford: Oxford University Press.

Willis, David. 2006. Negation in Middle Welsh. *Studia Celtica* 40. 63–88.

Willis, David. 2007. Specifier-to-head reanalyses in the complementizer domain: Evidence from Welsh. *Transactions of the Philological Society* 105. 432–80. DOI: https://doi.org/10.1111/j.1467-968X.2007.00194.x

Willis, David. 2010. Motivating the emergence of new markers of sentential negation: The case of Welsh *ddim*. *Diachronica* 27. 110–56. DOI: https://doi.org/10.1075/dia.27.1.04wil

Willis, David. 2012. A minimalist approach to Jespersen's cycle in Welsh. In Dianne Jonas, John Whitman & Andrew Garrett (eds.), *Grammatical change: Origins, nature, outcomes*, 93–119. Oxford: Oxford University Press.

Willis, David. 2013. The history of negation in the Brythonic Celtic languages. In David Willis, Christopher Lucas & Anne Breitbarth (eds.), *The history of negation in the languages of Europe and the Mediterranean, volume 1: Case studies*, 239–98. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199602537.003.0007

Willis, David. 2014. Newid ac amrywiaeth mewn Cymraeg Cyfoes [Change and variation in Contemporary Welsh]. In Delyth Prys (ed.), *Ysgrifau ar ieithyddiaeth a geiriaduraeth Gymraeg*, 84–102. Carmarthen: Coleg Cymraeg Cenedlaethol.

Willis, David. 2017. Investigating geospatial models of the diffusion of morphosyntactic innovations: The Welsh strong second-person singular pronoun *chdi*. *Journal of Linguistic Geography* 5. 41–66. DOI: https://doi.org/10.1017/jlg.2017.1

Zanuttini, Raffaela, Jim Wood, Jason Zentz & Laurence Horn. 2018. The Yale Grammatical Diversity Project: Morphosyntactic variation in North American English. *Linguistic Vanguard* 4. 1–15. DOI: https://doi.org/10.1515/lingvan-2016-0070

Zeijlstra, Hedde H. 2004. *Sentential negation and negative concord*. Amsterdam: University of Amsterdam PhD dissertation.

**How to cite this article:** Willis, David. 2019. Dialect syntax as a testbed for models of innovation and change: Modals and negative concord in the *Syntactic Atlas of Welsh Dialects*. *Glossa: a journal of general linguistics* 4(1): 31.1–30. DOI: https://doi.org/10.5334/gjgl.844

**Submitted:** 22 October 2018 **Accepted:** 24 January 2019 **Published:** 22 February 2019



]u[ *Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by Ubiquity Press. **OPEN ACCESS**