The complexities of the Welsh copula

Bob Borsley
University of Essex and Bangor University

Proceedings of the 26th International Conference on
Head-Driven Phrase Structure Grammar

University of Bucharest

Stefan Müller, Petya Osenova (Editors)

2019

Stanford, CA: CSLI Publications

pages 5–25

Keywords: copula, syntax-morphology mismatches, Welsh

Borsley, Bob. 2019. The complexities of the Welsh copula. In Stefan Müller & Petya Osenova (eds.), *Proceedings of the 26th International Conference on Head-Driven Phrase Structure Grammar, University of Bucharest*, 5–25. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2019.1.

Abstract

The Welsh copula has a complex set of forms reflecting agreement, tense, polarity, the distinction between main and complement clauses, the presence of a gap as subject or complement, and the contrast between predicative and equative interpretations. An HPSG analysis of the full set of complexities is possible given a principle of blocking, whereby constraints with more specific antecedents take precedence over constraints with less specific antecedents, and a distinction between morphosyntactic features relevant to syntax and morphosyntactic features relevant to morphology.

1. Introduction

It is probably a feature of most languages that the copula is more complex in various ways than standard verbs. This is true in English, and it is very definitely true in Welsh. The Welsh copula has a complex set of forms reflecting agreement, tense, polarity, the distinction between main and complement clauses, the presence of a gap as subject or complement, and the contrast between predicative and equative interpretations. In this paper, I will set out the facts and develop an analysis within the Head-Driven Phrase Structure Grammar (HPSG) framework. I will draw here on the proposals of Borsley (2015) and especially Bonami, Borsley, and Tallerman (2016). In particular, I will utilize two mechanisms which are employed in the latter. Firstly, I will assume a principle of blocking, whereby if the antecedents of two constraints stand in a subsumption relation, only the more specific constraint may apply. Secondly, I will assume that there is a distinction between two sets of morphosyntactic features, one relevant to syntax and another relevant to morphology. For most words the two sets will be identical, but in some cases there will be a mismatch. These two mechanisms will be crucial for ensuring the correct form of the copula.

The paper is organized as follows. In section 2, I develop an analysis of the basic argument selection properties of the Welsh copula. Then, in section 3, I consider agreement and tense. I go on in section 4 to look at the relevance of polarity and the main-complement distinction. Then, in section 5, I consider the influence of first subject and then complement gaps. In section 6, I look at the distinction between predicational and identity uses. Finally, in section 7, I summarize the paper.

* I am grateful to Bob Morris Jones for help with the data, and to Olivier Bonami, David Willis, Ian Roberts, and Marieke Meelen for helpful discussion of some of the ideas presented here. I am also grateful to various anonymous reviewers and the audience at HPSG19 for their comments and discussion. I alone am responsible for what appears here.

2. Argument selection

Like its counterpart in many languages, the Welsh copula *bod* allows a number of different complements.[1] Perhaps the simplest case is a PP complement, as in (1).

(1) Mae Gwyn yn yr ardd.
be.PRES Gwyn in the garden
'Gwyn is in the garden.'

(This and subsequent examples show that Welsh is a VSO language with verb-subject order in all finite clauses.) It can also have what I will call a Perfect Phrase (PerfP), consisting of the perfect particle *wedi* and a non-finite VP, and what I will call a Progressive Phrase (ProgP), consisting of the progressive particle *yn* and a non-finite VP, as in the following:[2]

(2) Mae Gwyn wedi cysgu.
be.PRES Gwyn PERF sleep.INF
'Gwyn has slept.'
(3) Mae Gwyn yn cysgu.
be.PRES Gwyn PROG sleep.INF
'Gwyn is sleeping.'

Progressive *yn* derives historically from the preposition *yn*, but it triggers no mutation, whereas the preposition *yn* triggers so-called nasal mutation, giving e.g. *yn Neiniolen* for 'in Deiniolen' (a village in North Wales). Finally, it can have what I will call a Predicative Phrase (PredP), consisting of the predicative particle *yn* and an AP or NP, as in the following:

(4) Mae Gwyn yn glyfar.
be.PRES Gwyn PRED clever
'Gwyn is clever.'
(5) Mae Gwyn yn feddyg.
be.PRES Gwyn PRED doctor
'Gwyn is a doctor.'

Unlike progressive *yn*, predicative *yn* triggers soft mutation. The basic forms of *glyfar* and *feddyg* are *clyfar* and *meddyg*, respectively.

---

[1] For general discussion of Welsh syntax, see Borsley, Tallerman, and Willis (2007).
[2] Welsh has a number of other aspectual particles, most of which are homophonous with prepositions, e.g. *ar* 'on', *heb* 'without', and *am* 'about'. See Jones (2010: Chapter 9) for discussion.

As with *be*, coordinations of different phrase types suggest that there is a single verb here.

(6) Mae Gwyn yn ddiog ac yn cysgu.
be.PRES Gwyn PRED lazy and PROG sleep.INF
'Gwyn is lazy and sleeping.'
(7) Mae Gwyn yn sâl ac yn y gwely.
be.PRES Gwyn PRED ill and in the bed
'Gwyn is ill and in bed.'
(8) Mae Gwyn yn ieithydd ac yn astudio Cymraeg.
be.PRES Gwyn PRED linguist and PROG study.INF Welsh
'Gwyn is a linguist and studying Welsh.'

The facts can be handled like similar facts in English and elsewhere by assuming that the Welsh copula takes a [PRED +] complement and that all these phrase types are [PRED +].

Bod takes as its subject whatever its complement requires, including an expletive subject, as the following illustrate:[3]

(9) Mae (hi) 'n bwrw glaw.
be.PRES she PRED strike.INF rain
'It's raining.'
(10) Mae (hi) 'n amlwg bod Mair wedi dod yn ôl.
be.PRES she PRED obvious be Mair PERF come.INF back
'It is obvious that Megan has come back.'

Thus, it appears to be a raising verb.[4] This means an ARG-ST feature of the following form:

(11) $\left[\text{ARG-ST} < [1], \begin{bmatrix}\text{HEAD}[\text{PRED}+] \\ \text{SUBJ} , < [1] >\end{bmatrix} >\right]$

I am assuming here that the subject of a [PRED +] element appears in its SUBJ list. However, I will assume below, following Borsley (1989), that all the arguments of finite verbs, subjects as well as compements, appear in their COMPS lists. Among other things, this accounts for the fact that the subject of a finite verb is always post-verbal.

[3] As Joan Maling has emphasized to me, Welsh is rather unusual in using a feminine pronoun as an expletive.
[4] Cf. Pollard and Sag (1994, 147) and Bender (2001, 48) on *be*.

3. Agreement and Tense

It is not surprising that the Welsh copula has forms reflecting agreement and tense. However, in both areas, it has interesting properties.

Unlike the English copula, but like standard Welsh verbs, the copula only shows agreement with a pronominal subject. Here are examples with third person singular and plural pronouns.

(12) a. Mae o / hi yn y gegin.
be.PRES he she in the kitchen
'He/She is in the garden.'
b. Maen nhw yn y gegin.
be.PRES.3PL they in the kitchen
'They are in the garden.'

With a non-pronominal subject, singular or plural, the form in (12a) appears and not that in (12b).

(13) Mae 'r bachgen / bechgyn yn y gegin.
be.PRES the boy boys in the kitchen
'They boy is/The boys are in the garden.'
(14) *Maen y bechgyn yn y gegin.
be.PRES.3PL the boys in the kitchen

The form in (12a) is sometimes seen as a third person singular form, but I will argue that it is a form unspecified for agreement (hence the gloss).

Borsley (2009) argues that verb-subject agreement is one instance of agreement between a head and an immediately following pronoun. Prepositions show agreement the form of a suffix with a following pronominal object, non-finite verbs show agreement in the form of a preceding clitic with a following pronominal object, and nouns show agreement in the form of a preceding clitic with a following pronominal possessor. In all cases, we also have agreement with a pronominal first conjunct of a coordinate NP in the relevant position. Borsley (2009) proposes that all these heads have an AGR(EEMENT) feature whose value is the relevant index when followed by a pronoun and otherwise *none*.

To capture the distinctive agreement behavior of finite verbs, we can propose that they have five forms in each tense specified for agreement with first and second person singular and plural and third person plural pronouns, and a form in each tense which is not specified for agreement. Following Bonami, Borsley, and Tallerman (2016), I assume that the morphological features which are responsible for the form of verbs and other parts of speech are the value of a feature INFL. Given this, assumption, we can propose constraints like the following, where, following a variety of earlier work, LID

is a feature whose value is unique to each distinct lexeme, the words that realise it, and the phrases that they head.

(15) $\left[\begin{array}{l}\text{INFL}\left[\begin{array}{l}\text{LID } \textit{bod}\\ \text{VFORM } \textit{fin}\\ \text{TENSE } \textit{pres}\\ \text{AGR } [\textit{3rd}, \textit{plur}]\end{array}\right]\end{array}\right] \rightarrow$ [PHON *maen*]

We will have similar constraints for first and second person singular and plural forms. We will also have a constraint of the following form:

(16) $\left[\begin{array}{l}\text{INFL}\left[\begin{array}{l}\text{LID } \textit{bod}\\ \text{VFORM } \textit{fin}\\ \text{TENSE } \textit{pres}\end{array}\right]\end{array}\right] \rightarrow$ [PHON *mae*]

Notice that this does not specify a value for AGR. Given the principle of blocking, (16) will not apply where a constraint specifies a specific value for AGR. Hence, *mae* will not appear with third person plural pronouns or first and second person singular or plural pronouns. But it will appear with a third person singular pronoun and with a non-pronominal NP, singular or plural. This is what we have in (12a) and (13). We will see later that slightly more complex constraints are in fact necessary.

The Welsh copula is just like other verbs where agreement is concerned, but with tense it is different. While standard verbs have three tenses, past, future, and conditional, the copula has five tenses, these three and two more, present and imperfect. Table 1 illustrates the third person singular forms of a standard verb and the copula.

	Cerdded 'walk'	*Bod* 'be'
Future	*cerddith*	*bydd*
Past	*cerddodd*	*buodd*
Conditional	*cerddai*	*byddai*
Present	----------	*mae*
Imperfect	----------	*roedd*

Table 1: Third person forms of *cerdded* 'walk' and *bod* 'be'

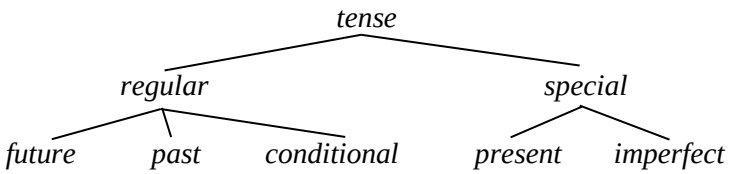
The present and imperfect of *bod* are used to express present and imperfect meanings with standard verbs, as the following illustrate:

(17) Mae Megan yn gadael.
be.PRES Megan PROG leave.INF
'Megan is leaving.'
(18) Roedd Megan yn gadael.
be.IMPF Megan PROG leave.INF
'Megan was leaving.'

One might propose that these are complex or periphrastic present and imperfect forms of the copula. However, all tenses of *bod* can take a ProgP complement. What we have here, then, is not periphrasis but an independent construction which allows the language to express the meanings that certain non-existent forms would have if they existed.[5]

It is not difficult to deal with this contrast between *bod* and standard verbs with respect to tense. Following Bonami, Borsley, and Tallerman (2016), I assume the following system of values for the feature TENSE:[6]

(19)



The following constraint will ensure that standard verbs only have past, future, and conditional forms:

(20) $\begin{bmatrix} \text{LID } \textit{standard-verb} \\ \text{VFORM } \textit{fin} \end{bmatrix} \rightarrow$ [TENSE *regular*]

I assume that *standard-verb* is a supertype of the LID values of all standard verbs. Thus, (20) will ensure that the finite forms of standard verbs are never present or imperfect. There will be no comparable constraint on finite forms of *bod*, and so all five tenses will be possible.

4. Polarity and the main–complement distinction

Some further complexities involve polarity and the distinction between main and complement clauses. The former just involve the third person present tense. The latter are more widespread.

[5] See Brown et al. (2012) for discussion of the nature of periphrasis.

[6] Bonami, Borsley, and Tallerman (2016) call this feature TMA (TENSE-MOOD-ASPECT). What it is called is of no real importance.

As earlier examples indicate, in affirmative declarative clauses, the basic present tense form of *bod* is *mae*. Different forms appear in negative declarative, and interrogative or conditional clauses.[7]

(21) Dydy Gwyn ddim yn yr ardd.
be.PRES Gwyn NEG in the garden
'Gwyn is not in the garden.'

(22) a. Ydy Gwyn yn yr ardd?
be.PRES Gwyn in the garden
'Is Gwyn in the garden?'

b. os ydy Gwyn yn yr ardd
if be.PRES Gwyn in the garden
'if Gwyn is in the garden'

These examples have definite subjects. Different forms appear with an indefinite subject, as the following show:

(23) Does neb yn yr ardd.
be.PRES nobody in the garden
'Nobody in the garden.'

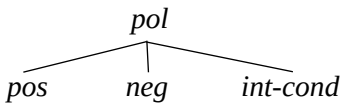
(24) a. Oes unrhyw un yn yr ardd?
be.PRES anybody in the garden
'Is anybody in the garden?'

b. os oes unrhyw un yn yr ardd
if be.PRES anybody in the garden
'if anybody is in the garden'

Clearly, there are some important complexities here.[8]

The facts suggest that we need a POL(ARITY) feature with three values: *pos*(*itive*), *neg*(*ative*), and *int*(*errogative*)-*cond*(*itional*). With *pol*(*arity*) as an unspecified value, this gives us the following values:

(25)



[7] A few ordinary verbs have distinct negative forms in some varieties (see Borsley and Jones 2005: 50-52), but most ordinary verbs take the same form in the three types of sentence that we are distinguishing here.

[8] *Dydy* and *does* are morphologically negative but not semantically negative. As discussed in Borsley and Jones (2005) and Borsley (2006), negative sentences must contain a prominent semantically negative constituent. This entails that *dydy* must co-occur with a negative post-subject adverb such as *ddim* and that *does* must co-occur with a negative subject such as *neb*.

Mae will be [POL *pos*], *dydy* and *does* [POL *neg*], and *ydy* and *oes* [POL *int-cond*]. This means the following constraint for *mae* instead of (16):

(26) $\begin{bmatrix} \text{INFL} \begin{bmatrix} \text{LID } \textit{bod} \\ \text{VFORM } \textit{fin} \\ \text{TENSE } \textit{pres} \\ \text{POL } \textit{pos} \end{bmatrix} \end{bmatrix} \rightarrow$ [PHON *mae*]

Assuming that the subject of a finite verb is the first member of its COMPS list, *dydy* and *ydy* will have NP[DEF +] as the first member of their COMPS list, and *does* and *oes* will have NP[DEF –]. For *dydy* and *does*, this means the following constraints:

(27) $\begin{bmatrix} \text{INFL} \begin{bmatrix} \text{LID } \textit{bod} \\ \text{VFORM } \textit{fin} \\ \text{TENSE } \textit{pres} \\ \text{POL } \textit{neg} \end{bmatrix} \\ \text{COMPS} < \text{NP[DEF} +], \ldots > \end{bmatrix} \rightarrow$ [PHON *dydy*]

(28) $\begin{bmatrix} \text{INFL} \begin{bmatrix} \text{LID } \textit{bod} \\ \text{VFORM } \textit{fin} \\ \text{TENSE } \textit{pres} \\ \text{POL } \textit{neg} \end{bmatrix} \\ \text{COMPS} < \text{NP[DEF} -], \ldots > \end{bmatrix} \rightarrow$ [PHON *does*]

Ydy and *oes* will be a result of similar constraints with [POL *int-cond*] instead of [POL *neg*].

There is more to be said here. There is evidence that the values *pos* and *neg* form a natural class. Both [POL *pos*] and [POL *neg*] forms appear in many contexts, especially declarative main clauses and many complement clauses. This suggests that they should be grouped together. But there is also evidence that *neg* and *int-cond* form a natural class. Both [POL *int-cond*] and [POL *neg*] forms forms appear in interrogatives and conditionals. The following illustrate the latter:

(29) a. Dydy ’r ddafad ddim yn yr ardd?
be.PRES the sheep NEG in the garden
‘Is the sheep not in the garden?’

b. os dydy 'r ddafad ddim yn yr ardd?
if be.PRES the sheep NEG in the garden
'if the sheep is not in the garden'

Moreover, *bod* has certain reduced forms which can appear where both [POL *neg*] and [POL *int-cond*] forms appear. Thus, (30a) has *dy* where *dydy* might appear, and (30b) and (30c) have it where *ydy* might appear:

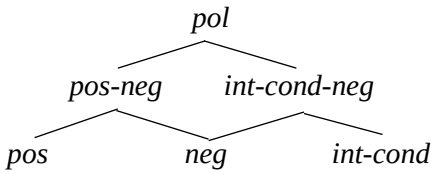
(30) a. Dy 'r ddafad ddim yn yr ardd.
be.PRES the sheep NEG in the garden
'The sheep is not in the garden.
b. Dy 'r ddafad yn yr ardd?
be.PRES the sheep in the garden
'Is the sheep in the garden.'
c. os dy 'r ddafad yn yr ardd?
if be.PRES the sheep in the garden
'if the sheep is in the garden'

Similarly, (31a) has *'s* where *does* might appear and (31b) has it where *oes* might appear.

(31) a. 'S neb yn yr ardd.
be.PRES nobody in the garden
'Nobody in the garden.'
b. 'S unrhyw un yn yr ardd.
be.PRES anybody in the garden
'Is anybody in the garden?'

We can treat both *pos* and *neg* and *neg* and *int-cond* as natural classes by proposing the following system of values:

(32)



With this system we can say that declarative main clauses and many complement clauses are [POL *pos-neg*] and that interrogatives and conditional clauses are [POL *int-cond-neg*]. We can also say that reduced forms like *dy* and *'s* are [POL *int-cond-neg*].

We turn now to the effects of the main-complement distinction. Certain pre-verbal particles are relevant here. In affirmative declarative main clauses,

the copula, like standard verbs, may be preceded by a particle, *mi* in North Wales or *fe* in South Wales. The following illustrates:

(33) Mi/Fe fydd Gwyn yn yr ardd.
AFF be.FUT Gwyn in the garden
'Gwyn will be in the garden.'

In negative complement clauses, verbs, including the copula, may be preceded by a particle *na* (*nad* before a vowel).

(34) Dywedodd Megan [na fydd Gwyn ddim yn yr ardd].
say.PAST Megan NEG be.FUT Gwyn NEG in the garden
'Megan said Gwyn will not be in the garden.'

Harlow (1983), Willis (1998: 70-71) and Borsley and Jones (2005: 57) argue that these particles form a constituent with the following verb. It is not clear whether they are separate words or prefixes, but much the same analytic issues arise on either assumption. In either case, the facts can be handled by labelling bare verbs as [MARKING *unmarked*] and particle + verb combinations as [MARKING *marked*]. *Mi*/*fe* will then combine with an *unmarked* form which is [POL *pos*, ROOT +] and *na*(*d*) will combine with an *unmarked* form which is [POL *neg*, ROOT –].[9]

For some speakers. *mi*/*fe* only occurs with past, future, and conditional forms of the copula, and not with the present and imperfect forms. For such speakers, we can say that the particles only combine with [TENSE *regular*] forms. Other speakers allow *mi*/*fe* with present and imperfect forms of *bod* but not with the third person present tense forms. For these speakers, we can assume that *mi*/*fe* combines with any [MARKING *unmarked*] form but that third person present tense forms are [MARKING *marked*].[10]

Also relevant here are some facts discussed in Bonami, Borsley and Tallerman (2016). As they note, present forms of *bod* and, for some speakers, imperfect forms too are ungrammatical in complement clauses:

---

[9] Bonami, Borsley, and Tallerman (2016) propose that there is a three-way distinction between main clauses, complement clauses, and unbounded dependency clauses and employ a three-valued STATUS feature rather than a two-valued ROOT. Whether this is necessary is not clear to me.

[10] Southern dialects have certain special negative present tense forms of the copula. Here is an example:

(i) So 'r ddafad yn yr ardd.
be.NEG.PRES the sheep in the garden
'The sheep is not in the garden.'

These forms are confined to main clauses and hence must be [POL *neg*, ROOT +].

(35) *Dywedodd Megan [mae Gwyn yn yr ardd].
say.PRES Megan be.PRES Gwyn in the garden
'Megan said Gwyn is in the garden.'
(36) %Dywedodd Megan [roedd Gwyn yn yr ardd].
say.PRES Megan be.IMPF Gwyn in the garden
'Megan said Gwyn was in the garden.'

Instead of present forms of *bod* and for some speakers imperfect forms as well, what looks like the non-finite form *bod* appears.

(37) Dywedodd Megan [bod Gwyn yn yr ardd].
say.PRES Megan be.INF Gwyn in the garden
'Megan said Gwyn is/was in the garden.'

Bod shows agreement in the form of a clitic with a following pronoun like an ordinary non-finite verb. Thus, we have the same agreement in (38) and (39).

(38) Dywedodd Megan [ei fod o yn yr ardd].
say.PRES Megan 3SGM be.INF he in the garden
'Megan said he is/was in the garden.'
(39) Dylai Megan ei weld o.
ought Megan 3SGM see.INF he
'Megan ought to see him.'

The only difference is that the clitic marks agreement with a subject in (38) and with an object in (39). Thus, *bod* seems to be morphologically non-finite. But there is evidence that it is syntactically finite. Only finite verbs precede their subject, as *bod* does here. Moreover, only finite verbs are negated by the negative adverb *ddim*, and *bod* has this property:

(40) Dywedodd Megan [bod Gwyn ddim yn yr ardd].
say.PRES Megan be.INF Gwyn NEG in the garden
'Megan said Gwyn is/was not in the garden.'

It seems, then, that *bod* in these clauses is a form of the copula which is syntactically finite but morphologically non-finite. Thus, we need an approach which distinguishes between morphological and syntactic finiteness.

Before we outline an analysis, we should note that there is one situation in which present and imperfect forms of *bod* may appear in complement clauses. This is in complement clauses affected by an unbounded dependency such as the following (Willis 2000, 2011, Borsley 2013):[11]

[11] Some speakers have *bod* in such sentences, but others prefer present and imperfect forms.

(41) Beth mae Aled yn credu [mae Elen yn
what be.PRES Aled PROG believe.INF be.PRES Elen PROG
ei ddarllen]?
3SGM read.INF
'What does Aled believe that Elen is reading?'

(42) Beth mae Aled yn credu [roedd Elen yn
what be.PRES Aled PROG believe.INF be.IMPF Elen PROG
ei ddarllen]?
3SGM read.INF
'What does Aled believe that Elen was reading?'

It seems, then, that present and imperfect forms of *bod* are only morphologically non-finite when they are not affected by an unbounded dependency. On standard HPSG assumptions, this means when they are [SLASH {}].

Bonami, Borsley and Tallerman (2016) show that it is easy to accommodate the facts given a distinction between morphosyntactic features relevant to syntax (the value of HEAD) and morphosyntactic features relevant to morphology (the value of INFL). Normally, HEAD and INFL will have the same value as a result of the following constraint:

(43) $[\ ] \rightarrow \begin{bmatrix} \text{HEAD}\ [1] \\ \text{INFL}\ [1] \end{bmatrix}$

In [ROOT –] clauses which are [SLASH {}], the positive present tense of *bod* will be [HEAD [VFORM *fin*]] but [INFL [VFORM *inf*]] as a result of the following constraint:

(44) $\begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{LID}\ \textit{bod} \\ \text{VFORM}\ \textit{fin} \\ \text{ROOT}\ - \\ \text{TENSE}\ \textit{pres} \\ \text{POL}\ \textit{pos} \end{bmatrix} \\ \text{SLASH}\ \{\} \end{bmatrix} \rightarrow$ [INFL [VFORM *inf*]]

For speakers who have *bod* instead of imperfect forms as well the constraint will refer to [TENSE *special*].

Notice that the constraint in (44) refers to [POL *pos*] forms. What about [POL *neg*] and [POL *int-cond*] forms? [POL *neg*] forms may be *bod* (as in (40)) but may also be the ordinary present tense forms. This suggests that they

require a constraint with a disjunctive consequent. [POL *int-cond*] are ordinary present tense forms. So nothing special is required here.

5. The effect of gaps

We can turn now to examples where one of the arguments of *bod* is an unbounded dependency gap. In some cases, we see the forms of *bod* that appear in ordinary affirmative or negative clauses, but in others, we have something different.

The simplest of these cases is where a gap appears in a present tense subject position. We have examples like the following:

(45) y dyn [*mae / sy(dd) yn yr ardd]
the man be.PRES in the garden
'the man who is in the garden'

(46) y dyn [*dydy / sy(dd) ddim yn yr ardd]
the man be.PRES NEG in the garden
'the man who is not in the garden'

Here, we have not the expected forms *mae* and *dydy* but a special form *sy*(*dd*). To accommodate such examples, the constraints that are responsible for *mae* and *dydy* must be constrained to require a canonical subject. In the case of *mae*, this means the following constraint:

(47)
$$\begin{bmatrix} \text{INFL} \begin{bmatrix} \text{LID } \textit{bod} \\ \text{VFORM } \textit{fin} \\ \text{TENSE } \textit{pres} \\ \text{POL } \textit{pos} \end{bmatrix} \\ \text{COMPS} < [\textit{canon}], \ldots > \end{bmatrix} \rightarrow \quad [\text{PHON } \textit{mae}]$$

Sydd can then be analyzed as the product of the following constraint, which requires the subject to be a gap:

(48)
$$\begin{bmatrix} \text{INFL} \begin{bmatrix} \text{LID } \textit{bod} \\ \text{VFORM } \textit{fin} \\ \text{TENSE } \textit{pres} \\ \text{POL } \textit{pos} - \textit{neg} \end{bmatrix} \\ \text{COMPS} < [\textit{gap}], \ldots > \end{bmatrix} \rightarrow \quad [\text{PHON } \textit{sydd}]$$

This assumes, following Borsley (2009, 2013), that gaps appear in VALENCE lists and not just in ARG-ST lists.

We turn now to complement gaps. The copula takes the expected form if the gap is a PP, PerfP, or ProgP. The following are emphatic counterparts of (1) and (2) with a PP gap and a PerfP gap in complement:

(49) Yn yr ardd mae Gwyn.
In the garden be.PRES Gwyn
'Gwyn is IN THE GARDEN.'
(50) Wedi cysgu mae Gwyn.
PERF sleep.INF be.PRES Gwyn
'Gwyn has SLEPT.'

In both, the copula is *mae*, as we would expect. I assume the following is an emphatic counterpart of (3) with a ProgP gap in complement position:

(51) Cysgu mae Gwyn.
sleep.INF be.PRES Gwyn
'Gwyn is SLEEPING.'

There is no progressive *yn* here. But *yn* appears when the ProgP has some sort of adverbial element in initial position, as the following illustrates:

(52) Wrthi yn golchi 'r car mae Mair.
at.3SGF PROG wash.INF the car be.PRES Mair
'Mair is in the process of washing the car.'

Borsley (2015) proposes that predicative *yn* is normally deleted or suppressed when it is in initial position, hence its absence from (51). In the present context, however, the important point about (51) (and (52)) is that the copula is *mae*, as expected. The situation is different if the gap is a PredP. The following are emphatic counterparts of (4) and (5):

(53) Clyfar *mae/ydy Gwyn.
clever be.PRES Gwyn
'Gwyn is CLEVER.'
(54) Meddyg *mae/ydy Gwyn.
doctor be.PRES Gwyn
'Gwyn is A DOCTOR.'

There is no predicative *yn* in these examples just as there is no progressive *yn* in (51). However, like progressive *yn*, it appears when the PredP has some sort of adverbial element in initial position:

(55) Bron yn barod *mae/ydy Mair.
almost PRED ready be.PRES Mair
'Mair is ALMOST READY.'

(56) Bron yn fradychwr *mae/ydy o.
almost PRED traitor be.PRES he
'He is ALMOST A TRAITOR.'

But in all these examples, the copula is not *mae*, which is expected in an affirmative declarative clause, but *ydy*, which is normally confined to interrogatives and conditionals.

These examples appear to be affirmative declarative clauses. In fact they must be affirmative clauses. They have no ordinary negative counterparts.[12] The only way to negate such sentences is by negating the initial constituent with *nid*/*dim*. Thus, (57a) is ungrammatical, and only (57b) is possible:[13]

(57) a. *Cysgu dydy Gwyn ddim.
sleep.INF be.PRES Gwyn NEG
'Gwyn is SLEEPING.'

b. Nid/dim cysgu mae Gwyn.
NEG sleep.INF be.PRES Gwyn
'Gwyn is not SLEEPING.'

This suggests that these clauses are [POL *pos*], and one would expect the verb that heads them to be the same. But the verb looks like a [POL *int-cond*] form. This seems to be a second case where HEAD and INFL have different values, in this case for the feature POL. We can attribute the facts to the following constraint:

(58) $\begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{LID } \textit{bod} \\ \text{VFORM } \textit{fin} \\ \text{TENSE } \textit{pres} \\ \text{POL } \textit{pos} \end{bmatrix} \\ \text{COMPS} < [], \begin{bmatrix} \textit{gap} \\ \text{PredP} \end{bmatrix} > \end{bmatrix} \rightarrow$ [INFL [POL *int-cond*]]

[12] It seems that complement gaps are generally bad with negated forms of *bod*.

[13] Notice that *yn* does not appear here although it would not be in initial position if it did. See Borsley (2015) for some discussion.

6. Identity interpretations

We turn finally to sentences in which the copula has an identity interpretation. As discussed in Zaring (1996) and Borsley (2015, section 3), it has some distinctive properties in this use. The following is a typical example:[14]

(59) Y meddyg ydy Gwyn.
the doctor be.PRES Gwyn
'Gwyn is the doctor.'

Here, the initial constituent is understood as a complement, and there is presumably an NP gap in the normal complement position. Again, the form is *ydy*, and *mae* is not possible.

(60) *Y meddyg mae Gwyn.
the doctor be.PRES Gwyn

Examples like (59) have no verb-initial counterparts. Hence, (61) is not possible with either *mae* or *ydy*.

(61) *Mae/ydy Gwyn y meddyg.
be.PRES Gwyn the doctor

This suggests that there is a separate identity copula with a distinctive syntax. However, all its forms are identical to forms of the predicational copula, and a satisfactory analysis needs to take account of this.

Before we outline an analysis, we should note a further fact about the identity copula. As we might expect, sentences with the identity copula have no ordinary negative counterparts, and can only be negated by negating the initial constituent with *nid*/*dim*.

(62) *Y meddyg ydy Gwyn ddim.
the doctor be.PRES Gwyn NEG
'Gwyn is not the doctor.'
(63) Nid/dim y meddyg ydy Gwyn.
NEG the doctor be.PRES Gwyn
'It's not the doctor that Gwyn is.'

[14] The very different syntax of identity sentences such as (59) and sentences with a predicative nominal such as (5) argues against the approach of Van Eynde (2015), in which the latter are analysed as examples of the former.

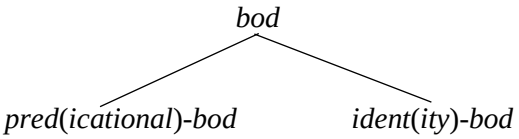
However, the identity copula can appear in both interrogatives and conditionals:[15]

(64) a. Y meddyg ydy Gwyn?
the doctor be.PRES Gwyn
'Is Gwyn the doctor?'
b. os y meddyg ydy Gwyn.
if the doctor be.PRES Gwyn
'if Gwyn is the doctor'

This suggests that the identity copula must be [POL *pos*] or [POL *int-cond*].

The facts that we are concerned with here can be handled by assuming that the two copulas are two forms of a single copula, i.e. by assuming an index *copula* with two subtypes, as follows:[16]

(65)



The syntactic and semantic properties of the two subtypes can be attributed to the following constraints:

(66) [LID *pred-bod*] $\rightarrow$

$$\begin{bmatrix} \text{ARG-ST} < [1], \begin{bmatrix} \text{HEAD}[\text{PRED}+] \\ \text{SUBJ}, < [1] > \\ \text{CONTENT}[2] \end{bmatrix} > \\ \text{CONTENT}[2] \end{bmatrix}$$

[15] Some speakers would have *mai*, which is generally viewed as complementizer, after *os* in a conditional clause, but assuming *os* combines with a [POL *int-cond*] clause, it seems reasonable to assume that *ydy* is [POL *int-cond*] in (66b).

[16] A rather similar approach is taken to the Arabic copula in Alotaibi and Borsley (forthcoming).

(67) [LID *ident-bod*] →

$$\begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{VFORM } \textit{fin} \\ \text{POL } \textit{pos} \vee \textit{int-cond} \end{bmatrix} \\ \text{ARG-ST} < [\text{INDEX}[1]], \begin{bmatrix} \textit{gap} \\ \text{INDEX}[2] \end{bmatrix} > \end{bmatrix} \\ \text{CONT} \begin{bmatrix} \textit{identity-rel} \\ \text{ARG}[1] \\ \text{ARG}[2] \end{bmatrix} \end{bmatrix}$$

The constraint in (66) ensures that the predicational copula takes a [PRED +] complement, has a subject which is the subject of its complement, and has the same interpretation as its complement. The constraint in (67) ensures that the identity-copula is finite and not negative, has a complement which is a gap, and has an identity interpretation.

But what about the forms of the two versions of the copula? In earlier discussion I have attributed the forms of the copula to constraints referring to [LID *bod*]. I will assume that all forms of the copula are the product of such constraints. With no further assumptions this would entail that parallel slots in the paradigms of two versions of the copula are filled by the same form. This is overwhelmingly what we find. The following imperfect tense examples illustrate the typical situation:

(68) Oedd Gwyn yn yr ardd.
be.IMPF Gwyn in the garden
'Gwyn was in the garden.'

(69) Yr athro oedd Gwyn.
the teacher be.IMPF Gwyn
'Gwyn was the teacher.'

But an issue obviously arises in the present tense, where identity *bod* has *ydy* and not *mae*. I propose that this is a third case where HEAD and INFL have different values, again in the value of POL. This can be attributed to the following constraint:

(70) $$\begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{LID } \textit{identity-bod} \\ \text{TENSE } \textit{pres} \\ \text{AGR}[1] \\ \text{POL } \textit{pos} \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \text{INFL} \begin{bmatrix} \text{LID } \textit{identity-bod} \\ \text{TENSE } \textit{pres} \\ \text{AGR}[1] \\ \text{POL } \textit{int-cond} \end{bmatrix} \end{bmatrix}$$

As a result of this constraint the present tense of the identity-copula will have *ydy* not only when it is [HEAD [POL *int-cond*]], as in (64a, b), but also when it is [HEAD [POL *pos*]], as in (59). Elsewhere, the identity-copula will have the same value for INFL as HEAD, and its forms will be identical to the corresponding forms of the predicational copula.

There is one further point to note about the identity-copula. This is that it does not take the form *bod* in complement clauses. We have example like the following:

(71) Dywedodd Megan [mai/taw y meddyg ydy Gwyn].
say.PAST Megan COMP the doctor be.PRES Gwyn
'Megan said that Gwyn is the doctor.'

This suggests that the constraint in (44) should be revised to refer not to [LID *bod*] but to [LID-*pred*-*bod*].

8. Concluding remarks

In the preceding pages I have developed an HPSG analysis for all the main complexities of the Welsh copula *bod*. I have assumed a variety of features, some very familiar, others less so, and I have proposed a variety of constraints to ensure that just the right forms appear. Following Bonami, Borsley, and Tallerman (2016), I have assumed a principle of blocking, whereby if the antecedents of two constraints stand in a subsumption relation, only the more specific constraint may apply. I have also made crucial use of a distinction between morphosyntactic features relevant to syntax, which are the value of HEAD, and morphosyntactic features relevant to morphology, which are the value of INFL. Normally these features have the same value, but I have proposed that there are three situations were forms of *bod* have different values for these features, one where *bod* appears rather than expected finite forms of the copula, and two where what looks like an interrogative-conditional form of *bod* appears rather than the expected positive declarative form. In all these situations, the principle of blocking ensures that certain unexpected forms appear and not the expected forms. The principle of blocking also allows a simple account of the way that what looks like the third person singular form of the verb appears with a non-pronominal subject, singular or plural.

REFERENCES

Alotaibi, A. and R. D. Borsley (forthcoming), The copula in Modern Standard Arabic, in A. Abeillé and O. Bonami (eds.), *Constraint-based Syntax and Semantics: Papers in Honor of Danièle Godard*, Stanford: CSLI Publications.

Bender, E. (2001), *Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence*, PhD thesis, Stanford University

Bonami, O., R. D. Borsley, and M. O. Tallerman (2016), On pseudo-non-finite clauses in Welsh, *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, 104–124.

Borsley, R. D. (1989), An HPSG approach to Welsh, *Journal of Linguistics* 25, 333–354.

Borsley, R. D. (2006), A linear approach to negative prominence, in S. Müller (ed.), *Proceedings of the HPSG06 Conference*, Stanford: CSLI Publications, 60–80.

Borsley, R. D. (2009), On the superficiality of Welsh agreement, *Natural Language and Linguistic Theory* 27, 225-265.

Borsley, R. D. (2013), On the nature of Welsh unbounded dependencies, *Lingua* 133, 1–29.

Borsley, R. D. (2015), Apparent filler–gap mismatches in Welsh, *Linguistics* 53-995-1029.

Borsley, R. D. and B. M. Jones (2005), *Welsh Negation and Grammatical Theory*, Cardiff: University of Wales Press.

Borsley, R. D., M. Tallerman and D. Willis (2007), *The Syntax of Welsh*, Cambridge: Cambridge University Press.

Brown, D., M. Chumakina, G. G., Corbett, G. Popova and A. Spencer, (2012), Defining 'periphrasis': key notions, *Morphology*, 22, 233–275.

Harlow, S. (1983), Celtic relatives, *York Papers in Linguistics* 10, 77–121.

Jones, B. M. (2010), *Tense and Aspect in Informal Welsh*. Berlin & New York: Mouton de Gruyter.

Pollard, C. and I. A. Sag (1994), *Head-driven Phrase Structure Grammar*, Chicago: University of Chicago Press.

Van Eynde, F. (2015), *Predicative constructions: from the Fregean to a Montagovian treatment*, Stanford: CSLI Publications.

Willis, D. W. E. (1998), *Syntactic Change in Welsh: A Study of the Loss of Verb-Second*, Oxford: Clarendon Press.

Willis, D. W. E. (2000). On the distribution of resumptive pronouns and wh-trace in Welsh, *Journal of Linguistics* 36, 531–573.

Willis, D. W. E. (2011). The limits of resumption in Welsh wh-dependencies. In: Rouveret, A. (Ed.), *Resumptive Pronouns at the Interfaces*, Amsterdam: Benjamins, pp. 189–222.

Zaring, L. (1996), Two 'be' or not two 'be': Identity, predication and the Welsh copula, *Linguistics and Philosophy* 19, 103–142.