Welsh clausal *i* and the hierarchical lexicon

Robert D. Borsley
University of Essex and Bangor University

Proceedings of the 30th International Conference on
Head-Driven Phrase Structure Grammar

University of Massachusetts Amherst

Stefan Müller, Elodie Winckel (Editors)

2023

Frankfurt/Main: University Library

pages 6–26

Keywords: Lexicon, Prepositions, Subordinate clauses

Borsley, Robert D. 2023. Welsh clausal *i* and the hierarchical lexicon. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst*, 6–26. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2023.1.

Abstract

A number of types of Welsh subordinate clause are introduced by what looks like the preposition *i* 'to', 'for'. Earlier research has shown that there are three different lexemes here. It is not unusual for a language to have homophonous lexemes, but these lexemes share a variety of properties, and also share properties with the preposition *i*. The similarities and the differences among these lexemes can be captured if they are grouped together as four different realisations of a single 'super-lexeme' within the hierarchical lexicon.

1. Introduction

A number of types of Welsh subordinate clause are introduced by what looks like the preposition *i* 'to', 'for'. It appears with subjectless infinitives in some control sentences such as (1):

(1) Mae Heledd yn awyddus [i weld Rhiannon].
be.PRES.3SG Heledd PRED eager to see.INF Rhiannon
'Heledd is eager to see Rhiannon.'

This is shown to be a control sentence by the ungrammaticality of (2) with a dummy subject in the main clause:

(2) *Mae hi 'n awyddus [i fwrw glaw].
be.PRES.3SG she PRED eager to strike.INF rain
*'It's eager to rain.'

It also appears with subjectless infinitives in some raising sentences such as (3), which is shown to be a raising sentence by the grammaticality of (4):

(3) Mae Heledd yn mynd [i weld Rhiannon].
be.PRES.3SG Heledd PROG go.INF to see.INF Rhiannon
'Heledd is going to see Rhiannon.'
(4) Mae hi 'n mynd [i fwrw glaw].
be.PRES.3SG she PROG go.INF to strike.INF rain
'It's going to rain.'

---

* An earlier version of this paper was presented at the 27th Welsh linguistics seminar, in Bangor, in June 2023. I am grateful to Maggie Tallerman, Frank Van Eynde, Berthold Crysmann, and Jakob Maché for helpful comments, and to Howard Edwards, Peredur Webb-Davies, and Bob Morris Jones for help with the Welsh data. I alone am responsible for what appears here.

Finally, it appears with full clauses with an overt subject reminiscent of English *for-to* clauses such as (5):

(5) Disgwyliodd Heledd [i Sioned weld Rhiannon].
expect.PAST.3SG Heledd to Sioned see.INF Rhiannon
'Heledd expected Sioned to see Rhiannon.'

I will refer to such clauses as *i*-clauses. An obvious question here is: how many *i* lexemes are there in this area? There is evidence that there are three different lexemes (although (1)-(4) involve the same lexeme), but I will show that they share a number of properties, and also share properties with the preposition. I will go on to show that the HPSG hierarchical lexicon allows both the similarities and the differences in this area to be captured.

The discussion is organized as follows: In section 2, I show, drawing especially on Tallerman (1998), that there are three *i* lexemes introducing subordinate clauses. Then, in section 3, I argue that all are complementizers and heads of phrases. In section 4, I show how they share properties with each other and with the preposition *i*. In section 5, I develop an analysis exploiting the hierarchical lexicon of HPSG, and in section 6, I highlight the possibility of similar analyses for some other lexemes. Finally, in section 8, I offer some concluding remarks.

2. How many *i* lexemes are there?

In an early discussion of examples of the kind that we are concerned with here (Borsley 1986), I assumed that *i* is a complementizer, and argued that such an analysis was problematic for the then current Government Binding Theory (GB). The argument was fairly simple.

For GB, subjectless infinitives in a control sentence have a PRO subject and subjectless infinitives in a raising sentence have an NP trace subject. GB assumptions require that (a) PRO must be ungoverned, (b) NP trace must be governed but not case marked, and (c) an overt NP must be case-marked, normally by some governor. It seems to follow that *i* must (a) not govern, (b) govern but not case mark, and (c) govern and case mark. This looks like a problem. Of course, there would be no problem if there were three different *i* lexemes, but it would be unsatisfactory to adopt this position if the only motivation was the maintenance GB assumptions. Tallerman (1998) argues that there are in fact three different *i* lexemes although not in the way GB assumptions require.

There seem to be no reason to think that control and raising complements involve different *i* lexemes, but Tallerman provides evidence that the *i* of subjectless infinitives and the *i* of *i*-clauses are distinct lexemes. She shows that predicates which can take both a full clause introduced by *i* and a subjectless infinitive do not necessarily have *i* with the subjectless infinitive.

Instead, they may be introduced by zero or an element homophonous with the preposition *o* 'from':

(6) a. Disgwyliodd Heledd [i Sioned weld Rhiannon].
expect.PAST.3SG Heledd to Sioned see.INF Rhiannon
'Heledd expected Sioned to see Rhiannon.'

b. Disgwyliodd Heledd [weld Rhiannon].
expect.PAST.3SG Heledd see.INF Rhiannon
'Heledd expected to see Rhiannon.'

(7) a. Roedd hi 'n siŵr [iddi hi glywed
be.IMPF.3SG she PRED sure to.3SGF she hear.INF
y gwcw].
the cuckoo
'She was sure she heard the cuckoo.'

b. Roedd hi 'n siŵr [o gyrraedd yn
be.IMPF.3SG she PRED sure from arrive.INF PRED
hwyr].
late
'She was sure to arrive late.'

This suggests that there are two distinct lexemes here.

Tallerman (1998) also shows, building on Harlow (1993), that while some *i*-clauses are non-finite clauses and rather like English *for-to* clauses, others are finite. *I*-clauses with *disgwylio* 'expect' and many other verbs are clearly non-finite. They are negated by the negative verb *peidio* like subjectless infinitives. It is mutated as *beidio* in both cases.[1]

(8) Disgwyliodd Heledd [i Sioned beidio â gweld
expect.PAST.3SG Heledd to Sioned NEG with see.INF
Rhiannon].
Rhiannon
'Heledd expected Sioned not to see Rhiannon.'

(9) Disgwyliodd Heledd [beidio â gweld Rhiannon].
expect.PAST.3SG Heledd NEG with see.INF Rhiannon
'Heledd expected not to see Rhiannon.'

But other *i*-clauses appear with verbs which normally take a finite clause such as *meddwl* 'think'. A rather surprising fact about Welsh is that past tense forms of verbs are generally not acceptable in positive complement clauses (Jones 2010: 171). Thus, (10) is quite problematic:

[1] Mutation is ubiquitous in Welsh. I will pass over unimportant instances of mutation without comment, but I will discuss some important instances below.

(10) %Meddyliodd Heledd [aeth Sioned adre'].
think.PAST.3SG Heledd go.PAST.3SG Sioned home
'Heledd thought that Sioned had gone home.'

In colloquial Welsh, a perfect clause involving *bod* 'be' and *wedi* appears instead (Jones 2010: 172):

(11) Meddyliodd Heledd [bod Sioned wedi mynd
think.PAST.3SG Heledd be.INF Sioned PERF go.INF
adre'].
home
'Heledd thought that Sioned had gone home.'

Despite appearances, this is a type of finite clause, as Awbery (1976: 41-43), Tallerman (1998) and Bonami, Borsley & Tallerman (2016) show. In literary Welsh, an *i*-clause appears:[2]

(12) Meddyliodd Heledd [i Sioned fynd adre'].
think.PAST.3SG Heledd to Sioned go.INF home
'Heledd thought that Sioned had gone home.'

The interpretation suggests that this clause is actually finite, and so does the fact that it is in a context where a finite clause is expected. The fact that a negative counterpart of this clause is the ordinary finite clause in (13) points to the same conclusion:[3]

(13) Meddyliodd Heledd [aeth Sioned ddim adre'].
think.PAST.3SG Heledd go.PAST.3SG Sioned NEG home
'Heledd thought that Sioned had not gone home.'

Anaphora also suggests that there are non-finite and finite *i*-clauses. In a non-finite *i*-clause, a pronoun cannot be bound by an NP in the main clause NP, but a reflexive can:

[2] For some discussion of the relation between literary Welsh and other varieties, see Borsley, Tallerman & Willis (2007: section 1.3).

[3] Past tense verbs are also acceptable in interrogative complement clauses, as (i) illustrates (Jones 2010: 174-5):

(i) Dw i 'n gofyn (a) dda'th Mair
be.PRES.1SG I PROG ask.INF Q come.PAST.3SG Mair
ddoe.
yesterday
'I'm asking whether Mair came yesterday.'

(14) a. Dymunai Aled$_i$ iddo fo$_{j/*i}$ fynd.
want.PAST.3SG Aled to.3SGM he go.INF
'Aled wanted him to go.'

b. Dymunai Aled iddo ei hun ddarllen
want.PAST.3SG Aled to.3SGM 3SG REFL read.INF
y llyfr.
the book
'Aled wanted himself to read the book.' (Tallerman 1998: 92)

In contrast, in a finite *i*-clause, a pronoun can be bound by an NP in the main clause or can be free, but a reflexive cannot be:

(15) a. Dywedodd Aled$_i$ iddo fo$_{i/j}$ fynd.
say.PAST.3SG Aled to.3SGM he go.INF
'Aled said he's gone.'

b. *Dywedodd Aled iddo ei hun fynd.
say.PAST.3SG Aled to.3SGM 3SG REFL go.INF
*'Aled said that himself went.' (Tallerman 1998: 90)

Coordination also distinguishes between non-finite and finite *i*-clauses. A non-finite *i*-clause cannot coordinate with a normal finite clause, as noted by Sadler (1988: 40):

(16) ??Disgwyliodd Emrys [i Mair fynd i Gaerdydd]
expect.PAST.3SG Emrys to Mair go.INF to Cardiff
ac [y byddai Siôn yn mynd i
and PRT be.COND.3SG Siôn PROG go.INF to
Abertawe].
Swansea
'Emrys expected Mair to go to Cardiff and that Siôn would be going to Swansea.' (Tallerman 1998: 92)

In contrast, a finite *i*-clause can readily coordinate with a normal finite clause:

(17) Meddyliodd Aled [i Alys fynd adre'] ac
think.PAST.3SG Aled to Alys go.INF home and
[y byddai Mair yn mynd yn fuan].
PRT be.COND.3SG Mair PROG go.INF PRED soon
'Aled thought that Alys had gone home and that Mair would be going soon.' (Tallerman 1998: 79)

The preceding discussion focused on *i*-clauses as verbal complements. It seems that both types of *i*-clause may also appear as prepositional

complements, but the facts are complex. The following looks as if it contains a finite *i*-clause:

(18) Ges i air 'dag e cyn [iddo fe
get.PAST.1SG I word with he before to.3SGM he
fynd].
go.INF
'I had a word with him before he went.'

However, this appears to be a non-finite clause deriving a past time interpretation from the context, rather like the non-finite constituent in the following (and its English translation):

(19) Ges i air 'dag e cyn mynd.
get.PAST.1SG I word with he before go.INF
'I had a word with him before going/I went.'

Here is a similar example where the context gives rise to a future interpretation:[4]

(20) Dw i 'n moyn cael gair 'dag e
be.PRES.1SG I PROG want.INF get.INF word with he
cyn [iddo fe fynd].
before to.3SGM he go.INF
'I want to have a word with him before he goes.'

This is like the non-finite constituent in the following:

(21) Dw i 'n moyn cael gair 'dag e
be.PRES.1SG I PROG want.INF get.INF word with he
cyn mynd.
before go.INF
'I want to have a word with him before going/I go.'

But the following naturally occurring example shows an *i*-clause after *oherwydd* 'because', which normally takes a finite clause as its complement:

[4] Bonami, Borsley and Tallerman (2016) call finite *i*-clauses pseudo-non-finite clauses because they look like non-finite clauses but are really finite clauses. From this perspective, examples like (18) and (20) could be called pseudo-pseudo-non-finite clauses.

(22) Dirywiodd y wladwriaeth Carthaginaidd oherwydd
deteriorate.PAST.3SG the state Carthaginian because
[i 'r Rhufeiniaid eu trechu yn y Rhyfeloedd
to the Romans 3PL conquer.INF in the wars
Pwnig].
Punic
'The Carthaginian state deteriorated because the Romans conquered them in the Punic Wars.' https://cy.wikipedia.org/wiki/Algeria

This appears to be a genuine finite *i*-clause as a prepositional complement. Like examples with a finite *i*-clause as a verbal complement, it is quite literary, and a more colloquial example would have *bod* and *wedi*.

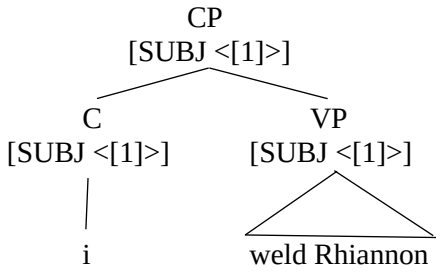
It seems, then, we have quite strong evidence that that there are two *i* lexemes in *i*-clauses: one non-finite, and one finite and past tense. I conclude that there are three clausal *i* lexemes altogether.

3. Three complementizers

What exactly are the three clausal i-lexemes? I will argue that they are all complementizers (essentially as in Borsley 1986) and heads taking complements.

In assuming that the *i* of subjectless infinitives is a complementizer and a head, I am essentially following Tallerman (1998). As is standard in HPSG, I assume that a subjectless infinitive is a VP. I assume, then, that this element is a complementizer taking a non-finite VP complement, and that it has the same value for SUBJ as its complement. In other words, it is a raising predicate, and apart from the fact that it is a complementizer is rather like English *to*. This means structures like the following:

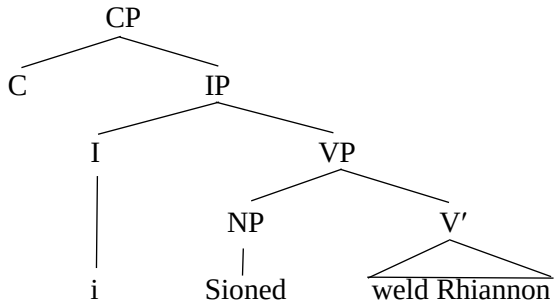
(23)



Not all HPSG work assumes that complementizers are heads taking a complement. Pollard and Sag (1994: 44-46) and others have proposed that they are markers combining with a clausal head of some kind. This looks like a possible alternative here. I will suggest, however, that it not plausible for the *i* lexemes in *i*-clauses.

Tallerman (1998) in fact assumes that *i* of *i*-clauses is not a complementizer. Assuming a fairly orthodox Chomskyan view of clause structure, in which there is a distinction between C(omplementizer) and I(nflection), she proposes that this *i* is in the I position. Thus, she has structures like the following:

(24)



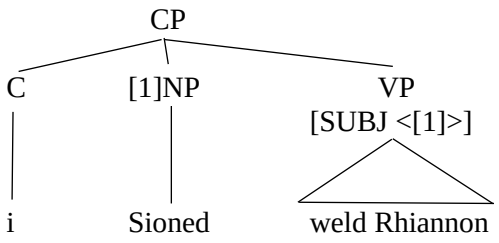
Tallerman sees the *i* of *i*-clauses as similar to finite verbs, which she assumes are in I because they can be preceded by certain particles, e.g. the affirmative particles *mi* in North Wales or *fe* in South Wales, which she assumes are in C:

(25) Mi/Fe welodd Sioned Rhiannon.
PART see.PAST.3SG Sioned Rhiannon
'Sionedd saw Rhiannon.'

However, there is evidence in Willis (1998: 70-71) and Borsley and Jones (2005: 57) that preverbal particles form a constituent with the following verb. This suggests that both are in C, and this is explicitly assumed by Willis working within a Chomskyan framework. Thus, the argument for this analysis seems quite weak even within Chomskyan assumptions. Outside those assumptions there is no reason to think that *i* occupies a different position in *i*-clauses and subjectless infinitives. I will assume, then, that the *i* lexeme in *i*-clauses is a complementizer.

I will also assume following Borsley (1999) that these elements are omplementizers taking two complements: an NP and a VP, where the NP is the subject of the VP. This is essentially the analysis that Sag (1997) proposes for English *for-to* clauses. It means structure like the following:

(26)



I think there is an objection here to an analysis in which complementizers are markers. As we will see in the next section, the *i* of *i*-clauses agrees with a following pronominal subject. As shown in Borsley (2009, 2022), agreement in Welsh generally involves a head and an immediately following complement. On the analysis in (26), *i*-clauses are just another example of this pattern. On a marker analysis, they would be something rather different. Markers combine with a single sister. Hence, on such an analysis, NP and VP would have to form a constituent and the agreement would involve a non-head and an element which is not its sister but a daughter of its sister. It seems preferable to maintain the assumption that agreement in Welsh involves a head and an immediately following complement, and the analysis in (26) allows one to do this.[5]

I conclude then there are three complementizers, one taking a single complement, a VP, and two taking two complements, an NP and VP. I will assume that *verb* and *complementizer* are subtypes of a type *verbal* as in Sag (1997: 457). This makes it unsurprising that there are positions in which both verb-headed and complementizer-headed constituents appear, and especially that finite *i*-clauses appear in the same positions as clauses headed by a finite verb.

4. **Similarities between the four *i* lexemes**

It is obviously not unusual for a language to have homophonous lexemes. Commonly, they have no other shared properties. In English, the preposition *to* and the infinitive marker *to* seem to have no other shared properties. The following Welsh examples illustrate a similar situation:

(27) a. Mae Heledd yn **Neiniolen**. (Deiniolen)
be.PRES.3SG Heledd in Deiniolen
'Heledd is in Deiniolen.'

b. Mae Heledd yn dawnsio.
be.PRES.3SG Heledd PROG dance.INF
'Heledd is dancing.'

These feature the preposition *yn* 'in' and the homophonous progressive marker. In (27a), the preposition triggers the alternation known as nasal mutation. Thus, the place name *Deiniolen* appears as *Neiniolen*. (Here and subsequently, I put important mutated words in bold and give the basic form

[5] It has been suggested to me that *i* could be a weak head in the sense of Tseng (2002), a head which derives many of its properties from its complement(s). But the various forms of *i* have little in common with their complements. The first of the forms is non-finite like its complement, and the second is non-finite like its second complement, but the third form has essentially no properties in common with its complements. Hence, I don't see any reason to think that we have weak heads here.

of the word in brackets.) In (27b), there is no mutation with the progressive marker. Historically, these are the same element (Sims-Williams 2015), but in the contemporary language they seem to be just two separate lexemes. However, the three complementizers that we are concerned with here are quite different. They have a variety of properties in common other than just their phonological form, and they all share properties with the preposition *i*.

Unlike the two *yn* lexemes just considered, which have different mutation properties, the four *i* lexemes have the same mutation property: they all trigger soft mutation on the following constituent. This is an NP in (28), (30) and (31), and a VP in (29):

(28) i **Fangor** (Bangor)
to Bangor

(29) Mae Heledd yn awyddus [i **weld**
be.PRES.3SG Heledd PRED eager to see.INF
Rhiannon]. (gweld)
Rhiannon
'Heledd is eager to see Rhiannon.'

(30) Disgwyliodd Heledd [i **ddau** dyn weld
expect.PAST.3SG Heledd to two man see.INF
Rhiannon]. (dau)
Rhiannon
'Heledd expected two men to see Rhiannon.'

(31) Meddyliodd Heledd [i **ddau** dyn fynd
think.PAST.3SG Heledd to two man go.INF
adre']. (dau)
home
'Heledd thought that two men went home.'

(The mutation of *weld* in (30) is triggered not by *i*, but by the preceding subject *ddau dyn*.)

The preposition *i* and the *i* of non-finite and finite *i*-clauses are also similar in showing agreement with a following third person pronoun:

(32) iddo fo / iddi hi / iddyn nhw
to.3SGM he to.3SGF she to.3PL they
'to him/her/them'

(33) Disgwyliodd Heledd [iddo fo / iddi hi /
expect.PAST.3SG Heledd to.3SGM he to.3SGF she
iddyn nhw weld Rhiannon].
to.3PL they see.INF Rhiannon
'Heledd expected him/her/them to see Rhiannon.'

(34) Meddyliodd Heledd [iddo fo / iddi hi /
think.PAST.3SG Heledd to.3SGM he to.3SGF she
iddyn nhw fynd adre'].
to.3PL they go.INF home
'Heledd thought that he/she/they had gone home.'

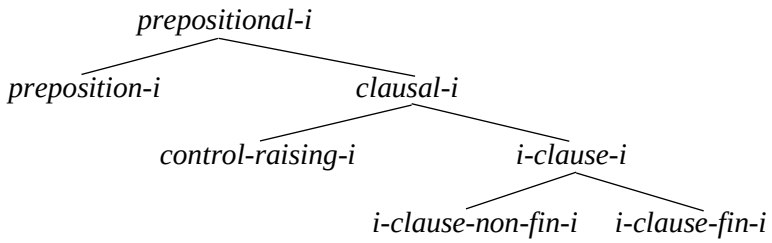
The preposition is unusual among prepositions in only showing agreement with a third person pronoun and not with all pronouns, and the complementizers have the same property. The *i* of subjectless infinitives does not show agreement, but it does not have the opportunity to because it is never immediately followed by a pronoun. Thus, we can say that all four lexemes have the same agreement potential, and more generally that they have the same morphological properties.

The four lexemes have different syntactic properties, but the three complementizers have in common the fact that they are complementizers, and the two *i*-clause complementizers have the same complement selection properties. Thus, it seems that there are four distinct *i* lexemes, but that they show a range of similarities. A satisfactory analysis needs to capture both the similarities and the differences in this area.

5. A hierarchical lexicon analysis

Standard HPSG assumptions about the lexicon stemming from Flickinger (1987) allow a fairly simple approach to situations like this. They allow the four lexemes to be analysed as four realisations of a 'super-lexeme' and all the shared properties to be specified just once. We can propose the type hierarchy in (35) for this part of the lexicon:

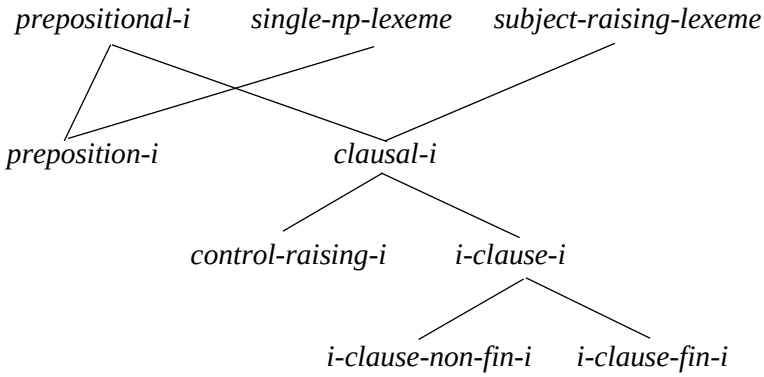
(35)



Note that *prepositional-i* and *preposition-i* are quite different types, and that I am using the type *control-raising-i* for the *i* of subjectless infinitives. This type hierarchy groups together all four lexemes as instances of *prepositional-i*, the three complementizers as instances as *clausal-i*, and the two *i*-clause complementizers as instances of *i-clause-i*. This hierarchy provides a basis for capturing the similarities and the differences in this area. However, it needs to be extended to take account of the fact that there is nothing unusual about the complement selection properties of the four lexemes. The preposition *i* is like

many prepositions in taking a NP complement. The *i* of subjectless infinitives is a raising predicate, as noted earlier. Finally, the two complementizers that appear in *i*-clauses are essentially raising predicates with both their arguments as complements.[6] These complement selection properties should be largely inherited from various argument selection types. I will assume two such types *single-np-lexeme* and *subject-raising-lexeme* and make *preposition-i* a subtype of the former, and *clausal-i* a subtype of the latter. This gives the following extended type hierarchy:

(36)



We begin with *prepositional-i* and the properties that are shared by all four lexemes. We have seen that all have the same morphological properties. I assume that these properties are a reflection of two features. First, following Borsley (2009, 2022), I assume that agreement in Welsh is the realization of a feature AGR, whose value is the index of a following pronoun with its PERSON, NUMBER, and GENDER features, or *none* when there is no following pronoun. I assume that the mutation-triggering property of a lexeme reflects a feature MUT(ATION)-TR(IGGER) with the values *soft*, *nasal*, and *aspirate* for the three kinds of mutation that occur in Welsh, or *none*. (Only the first is important here.) With these assumptions, we can attribute the phonological and morphological properties of the four lexemes to the following constraint on *prepositional-i* (where the MARKING feature allows heads to select a constituent headed by one of these lexemes):

(37) *prepositional-i* $\Rightarrow$ $\begin{bmatrix} \text{MARKING } i \\ \text{AGR } index \vee none \\ \text{MUT} - \text{TR } soft \end{bmatrix}$

[6] Following Borsley (1989), I assume that finite verbs also have all their arguments as complements. Thus, the *i*-clause complementizers are like finite raising verbs in their complement selection properties.

What about the form of the four lexemes? The grammar just needs to impose the following pairings of AGR value and form for *prepositional-i*:

AGR value	Form
3nd, sing, masc	*iddo*
3nd, sing, fem	*iddi*
3rd, plur	*iddyn*
Any value	*i*

AGR value-form pairings for *prepositional-i*

Following Bonami, Borsley and Tallerman (2016), I assume that more specific constraints take precedence over more general ones and hence that a general constraint does not apply if a more specific constraint requires something different. This means that the basic form *i* will not appear with a third person pronoun, but will appear in all other circumstances, i.e. with a first or second person pronoun or a non-pronominal NP.

We can turn now to the two immediate subtypes of *prepositional-i*. Here, we can propose the following simple constraints:

(38) *preposition-i* $\Rightarrow$ $\begin{bmatrix}\text{HEAD } prep \\ \text{SUBJ } <>\end{bmatrix}$

(39) *clausal-i* $\Rightarrow$ [HEAD *comp*]

The former will inherit properties from *single-np-lexeme* and the latter from *subject-raising-lexeme*. I assume these are subject to the following simple constraints:

(40) *single-np-lexeme* $\Rightarrow$ [ARG-ST <NP>]
(41) *subject-raising-lexeme* $\Rightarrow$ [ARG-ST <[1]NP, VP[*inf*, SUBJ <[1]>]>]

Numerous lexemes will inherit properties from these two types. I also assume the Argument Realization Principle in (42):

(42) *word* $\Rightarrow$ $\begin{bmatrix}\text{SUBJ } [1] \\ \text{COMPS } [2] \\ \text{ARG-ST } [1] \oplus [2]\end{bmatrix}$

As a subtype of *single-np-lexeme*, preposition *i* will have a single NP in its ARG-ST list. The SUBJ <> restriction in (38), interacting with the Argument Realization Principle, will ensure that this NP appears in its COMPS list. As a

subtype of *subject-raising-lexeme*, *clausal-i* will have the ARG-ST list specified by (41).

The two subtypes of *clausal-i*, *control-raising-i* and *i-clause-i*, will be subject to the following constraints:

(43) *control-raising-i* $\Rightarrow$ $\begin{bmatrix}\text{HEAD [VFORM } inf\text{]} \\ \text{SUBJ} < [] >\end{bmatrix}$

(44) *i-clause-i* $\Rightarrow$ [SUBJ <>]

Both types inherit a two member ARG-ST list from *subject-raising-lexeme*. The constraint on *control-raising-i* ensures that only the second member appears in its COMPS list. The constraint on *i-clause-i* ensures that both members appear in its COMPS list.

Finally, for the two subtypes of *i-clause-i*, we can propose the following quite simple constraints:

(45) *i-clause-non-fin-i* $\Rightarrow$ [HEAD [VFORM inf]]

(46) *i-clause-fin-i* $\Rightarrow$ $\left[\text{HEAD } \begin{bmatrix}\text{VFORM } fin \\ \text{TENSE } past\end{bmatrix}\right]$

(45) requires *i-clause-non-fin-i* to be non-finite, and (46) requires *i-clause-fin-i* to be finite and past tense. The past tense requirement ensures that finite *i*-clauses have the sort of interpretation that one would expect to be expressed by a complement clause with a past tense verb. The constraint in (46) could be extended to include the information that finite *i*-clauses are literary. It could be reformulated as follows:

(47) *i-clause-fin-i* $\Rightarrow$ $\begin{bmatrix}\text{SS|CAT|HEAD [VFORM } fin, \text{TENSE } past\text{]} \\ \text{REGISTER } literary\end{bmatrix}$

Within this analysis the four *i* lexemes have a variety of properties inherited from the various supertypes. Here are fairly full syntactic categories for each:

(48) *preposition-i*:

$$\left[\begin{array}{l}\text{HEAD } \begin{bmatrix}prep \\ \text{MARKING } i \\ \text{AGR } index \vee none \\ \text{MUT} - \text{TR } soft\end{bmatrix} \\ \text{SUBJ} <> \\ \text{COMPS} < \text{NP} >\end{array}\right]$$

(49) *control-raising-i*:

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} comp \\ \text{MARKING } i \\ \text{AGR } index \vee none \\ \text{MUT} - \text{TR } soft \\ \text{VFORM } inf \end{bmatrix} \\ \text{SUBJ} < [1] > \\ \text{COMPS} < \text{VP}[inf, \text{SUBJ} < [1] >] > \end{bmatrix}$$

(50) *i-clause-non-fin-i*:

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} comp \\ \text{MARKING } i \\ \text{AGR } index \vee none \\ \text{MUT} - \text{TR } soft \\ \text{VFORM } inf \end{bmatrix} \\ \text{SUBJ} <> \\ \text{COMPS} < [1]\text{NP}, \text{VP}[inf, \text{SUBJ} < [1] >] > \end{bmatrix}$$

(51) *i-clause-fin-i*:

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} comp \\ \text{MARKING } i \\ \text{AGR } index \vee none \\ \text{MUT} - \text{TR } soft \\ \text{VFORM } fin \\ \text{TENSE } past \end{bmatrix} \\ \text{SUBJ} <> \\ \text{COMPS} < [1]\text{NP}, \text{VP}[inf, \text{SUBJ} < [1\} >] > \end{bmatrix}$$

All these categories are [AGR *index* ∨ *none*]. Assuming the analysis of agreement developed in Borsley (2009, 2022), the value of AGR is an index when there is a following pronoun and otherwise *none*. The preposition and the two *i*-clause complementizers may be followed by a pronoun, but control-raising *i* is never followed by a pronoun. Thus, on the proposed analysis, it has an agreement potential which is never realised.

With these categories, the examples that we are concerned with here are all fairly ordinary head-complement phrases, two with one complement, and two with two. In each case, the head assigns soft mutation, and in each case, it will agree with an immediately following pronoun (but, as we have emphasized, control-raising *i* will never be immediately followed by a pronoun). (49) and (50) both head a non-finite clause, but (51) crucially heads a finite and past tense clause.

But what about the fact that a positive past tense verb is generally ungrammatical in a complement clause? One possibility is an analysis of the kind outlined in Bonami, Borsley & Tallerman (2016), in which finite *i* is literally a positive past tense form of the associated verb. However, as noted above, finite *i* is generally confined to the literary language, and in more colloquial Welsh a perfect clause involving *bod* 'be' and the particle *wedi* appears. I will assume, then, that there is a constraint ruling out a past tense verb in a positive complement clause, and that different varieties have different ways of expressing the meanings which cannot be expressed by a past tense verb, finite *i* fulfilling this role in the literary language.

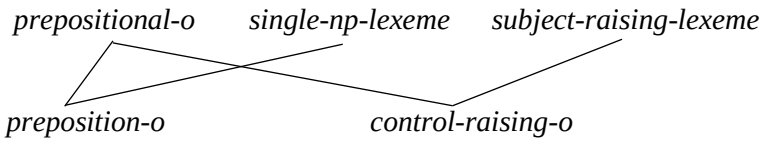
6. Some other super lexemes

There are some other cases in Welsh of homophonous lexemes which should probably be analysed as alternative realizations of a single super lexeme. I assume the element *o* in (7b) is another complementizer homophonous with a preposition. This element triggers soft mutation (the unmutated form of the following verb is *cyrraedd*). In this, it just like the proposition:

(52) Dw i wedi dôd o
be.PRES.1SG I PERF come.INF from
Gaernarfon. (Caernarfon)
Caernarfon
'I have come from Caernarfon.'

This suggests the type hierarchy in (53) and the constraints in (54)-(56):

(53)



(54) *prepositional-o* $\Rightarrow$ $\begin{bmatrix} \text{MARKING } o \\ \text{AGR } index \vee none \\ \text{MUT} - \text{TR } soft \end{bmatrix}$

(55) *preposition-o* $\Rightarrow$ $\begin{bmatrix} \text{HEAD } prep \\ \text{SUBJ } <> \end{bmatrix}$

(56) *control-raising-o* $\Rightarrow$ $\begin{bmatrix} \text{HEAD } \begin{bmatrix} comp \\ \text{VFORM } inf \end{bmatrix} \\ \text{SUBJ } < [] > \end{bmatrix}$

This is essentially a simplified version of the analysis for *i*.

There are at least two other cases for which an analysis of this kind seems appropriate. Welsh has a number of aspectual particles which are homophonous with a preposition. The most common, progressive *yn* and perfect *wedi*, which is homophonous with a preposition meaning 'after', seem to share no other properties with the preposition, but two others are different. The preposition *ar* 'on' and the homophonous aspect marker of imminence assign soft mutation (Jones 2010: 336-9).

(57) a. Mae 'r wylan ar **graig**. (craig)
be.PRES.3SG the seagull on rock
'The seagull is on a rock.'

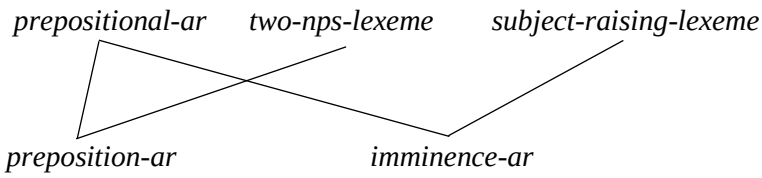
b. Mae o ar **ganu**. (canu)
be.PRES.3SG he on sing.INF
'He's about to sing.'

The preposition *ar* is predicative and has an object and a subject. This suggests that it inherits properties from a type *two-nps-lexeme* subject to the following constraint:

(58) *two-nps-lexeme* ⇒ [ARG-ST <NP, NP>]

This allows us to propose the following type hierarchy and constraimts:

(59)



(60) *prepositional-ar* ⇒ $\begin{bmatrix} \text{MARKING } ar \\ \text{AGR } index \vee none \\ \text{MUTR } soft \end{bmatrix}$

(61) *preposition-ar* ⇒ $\begin{bmatrix} \text{HEAD } prep \\ \text{SUBJ } < [] > \end{bmatrix}$

(62) *imminence-ar* ⇒ $\begin{bmatrix} \text{HEAD } imminence \\ \text{SUBJ } < [] > \end{bmatrix}$

We have a similar situation with the preposition *heb* 'without' and the homophonous negative perfect aspect marker (Jones 2010: 333-6). Both assign soft mutation:

(63) a. Dw i heb **gar** yr wythnos 'ma. (car)
be.PRES.1SG I without car the week here
'I'm without a car this week.'

b. Maen nhw heb **gyrraedd** eto. (cyraedd)
be.PRES.3PL they without arrive.INF yet
'They haven't arrived yet.'

This suggests an analysis like that proposed for *ar*.

7. Conclusions

I have argued in preceding pages that clausal *i* is three different lexemes (essentially as Tallerman 1998 showed), but that they are related lexemes with shared properties. I have also shown that the preposition *i* is a further related lexeme sharing various properties. I have shown that it is not too difficult to capture the similarities and differences among the four lexemes with the hierarchical lexicon of HPSG. With an appropriate type hierarchy the shared properties can be all be specified just once. There are a number of other cases in Welsh where an analysis of this kind may be appropriate.

A similar treatment is probably appropriate for a variety of phenomena in a variety of languages. In Borsley (2019), I analyze the Welsh predicational copula and identity copula as two realisations of a super-lexeme (without using the term), and Alotaibi and Borsley (2020) argue for a similar approach to the copula in Modern Standard Arabic. Also relevant here is recent unpublished work by Jacob Maché, who proposes an analysis rather like this for Germanic 'need' verbs, and also discusses how the type hierarchy it involves could emerge diachronically. It looks, then, as if the type of analysis developed here has considerable potential.[7]

[7] Naturally there are other matters that could be explored here. For example, the constructions we have been discussing may appear in relative clauses and other unbounded dependency clauses. The following, from (Borsley, Tallerman and Willis 207: 134), illustrate:

(i) Dw i 'n chwilio am rywbeth [i ('w)
be.PRES.1SG I PROG search for something to 3SGM
ddarllen ___].
read.INF
'I'm looking for something to read.'

(ii) Mae e wedi canu ar bob albwm [i ni
be.PRES.1SG he PERF sing.INF on every album to us
ei wneud erioed].
3SGM do.INF ever
'He's sung on every album we've ever done.'

But this is perhaps more a topic for research on Welsh unbounded dependencies.

REFERENCES

Alotaibi, Ahmad. & Robert D. Borsley 2020. The copula in Modern Standard Arabic. In Anne Abeillé & Olivier Bonami (eds.), *Constraint-based Syntax and Semantics: Papers in Honor of Danièle Godard*, 15-35. Stanford: CSLI Publications.

Awbery, Gwenllian 1976. *The Syntax of Welsh: A Transformational Study of the Passive*. Cambridge: Cambridge University Press.

Bonami, Olivier, Robert D. Borsley & Maggie O. Tallerman. 2016. On pseudo-non-finite clauses in Welsh. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller, *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, 104–124. Stanford, CA: CSLI Publications.

Borsley, Robert D. 1986. Prepositional complementizers in Welsh. *Journal of Linguistics* 22, 67–84.

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

Borsley, Robert D. 1999. Mutation and constituent structure in Welsh. *Lingua* 109, 263–300.

Borsley, Robert D. 2009. On the superficiality of Welsh agreement. *Natural Language & Linguistic Theory* 27, 225–265.

Borsley, Robert D. 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*, 5–25. Stanford, CA: CSLI Publications.

Borsley, Robert D. 2022. On the structure of Welsh noun phrases. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 29th International Conference on Head-Driven Phrase Structure Grammar*, Online (Nagoya/Tokyo), 27–47. Frankfurt/Main: University Library.

Borsley, Robert D. & Bob M. 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.

Flickinger, Daniel P. 1987. *Lexical Rules in the Hierarchical Lexicon*. Stanford, CA: Stanford University dissertation.

Harlow, Stephen 1993. Finiteness and Welsh clause structure. In Hans-Georg Obenauer & Anne Zribi-Hertz (eds), *Structure de la Phrase et Thėorie du Liage*, 93-119. Paris: Presses Universitaires de Vincennes.

Jones, Bob Morris. 2010. *Tense and Aspect in Informal Welsh*. Berlin & New York: Mouton de Gruyter.

Pollard, Carl & Ivan A. Sag 1994. *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Sadler, Louisa 1988. *Welsh Syntax: A Government-Binding Approach*. London: Croom Helm.
Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33, 431–484.
Sims-Williams, Patrick 2015. Four types of Welsh YN. *Transactions of the Philological Society* 113:3, 286–304.
Tallerman, Maggie O. 1998. The uniform case-licensing of subjects in Welsh. *The Linguistic Review* 15, 69–133.
Tseng, Jesse 2002. Remarks on Marking. In Frank van Eynde, Lars Hellan & Dorothee Beermann (eds), *HPSG 2001: Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, 267-283. CSLI Publications.
Willis, David W. E. 1998. *Syntactic Change in Welsh: A Study of the Loss of Verb-Second*. Oxford: Clarendon Press.