At the interface: selection of the Welsh definite article[1]

S. J. HANNAHS AND MAGGIE TALLERMAN

Abstract

The Welsh definite article has three surface forms, 'r [r], yr [ər] and y [ə]. Though it might appear that all three forms are derived from a single underlying representation, with the surface forms representing a case of simple (morphophonological) allomorphy, in fact such an account is untenable: in this article we show that the interaction of phonetic, phonological, morphological, and lexical considerations all bear on selection of the correct form. The solution to the problem of choosing the correct form of the definite article involves the staggered insertion of various classes of lexemes over the course of the syntactic derivation, with content words inserted first and functional elements inserted later on. Evidence from initial consonantal mutation in Welsh further supports the view that lexical insertion takes place at different stages, rather than as a single operation. We also consider the properties of a number of functional elements in Welsh involving interactions between distinct parts of the grammar. Finally, we show how our solution also sheds light on the system of initial mutation, which displays some interesting features with respect to the functional items that trigger mutations.

1. Introduction

Welsh has no indefinite article, but has a definite article with three surface forms, *'r* [r], *yr* [ər] and *y* [ə]. Given the phonetic similarity of *'r* and *y* to *yr*, one might assume initially that all three are derived from a single underlying representation and that the surface forms represent a case of simple (morphophonological) allomorphy. On closer examination, however, it becomes clear that a straightforward allomorphic account is untenable, because of the interaction of phonetic, phonological, morphological, and lexical considerations which all bear on selection of the correct form of

the definite article. This selection is squarely situated at the interface between different parts of the grammar.

In this article we examine the selection of the Welsh definite article from a number of angles, identifying the various linguistic factors which come into play in choosing the correct surface form. We argue that the selection must be characterized not only in terms of the phonological environment, but also depends crucially on morphological representation and on a particular view of lexical insertion.

Apart from the descriptive value of understanding the selection of the Welsh determiner relative to the grammatical structures involved, the theoretical interest in this problem is essentially twofold. In the first place, it is clearly relevant to ongoing investigations into ''interface'' phenomena, specifically the interactions between various parts of the grammar (e.g., Inkelas and Zec 1990; Lapointe et al. 1998; Kager et al. 1999). Secondly, it also has bearing on recent work on when lexical insertion takes place in the course of the syntactic derivation (e.g., Halle and Marantz 1993; Jackendoff 1997; Emonds 2000, 2002). We will argue that a solution to the problem of choosing the correct form of the definite article involves a staggered insertion of various classes of lexemes into the derivation, with content words inserted first and functional elements inserted later on. Our solution has the further advantage of shedding new light on aspects of the system of initial consonantal mutation in Welsh, which displays some interesting interactions with respect to the functional items that trigger various mutations.

We start in Section 2 by laying out the data at issue and showing where the problems lie. In Section 3, we look at various morphological and phonological issues, comparing and contrasting the definite article with other Welsh function words. Section 4 considers the interaction between lexical insertion, initial mutation and the choice of allomorphs of various function words. Section 5 concludes our discussion.

2.   The data and issues arising from it

The Welsh definite article has three forms, *y*, *yr* and *'r*. Phonetically these are [ə], [ər] and [r], though for the purposes of this article we generally retain the Welsh orthography. The basic distribution of each form before a noun is shown in (1):

(1)   a.   *yr* before a V-initial, h-initial or glide-initial word   yr afon
                                                                        'the river'
        b.   *y* before a C-initial word   y llyfr
                                                                        'the book'

c.	*'r* after a V-final word, preceding V	o'r afon 'from the river'
d.	*'r* after a V-final word, preceding C	o'r llyfr 'from the book'

Note from (1c) and (1d) that *'r* occurs whether the following word is V-initial or C-initial. (2) further illustrates the occurrence of *y* and *yr*, and also shows that when an adjective precedes a noun we find the same distribution as in (1), with *y* and *yr* alternating in just the same way before consonant-initial forms, (2a), and vowel-initial forms such as *unig*, (2b):

(2) a. y brif ddinas
the main city
'the capital city'
   b. yr unig blentyn
the only child

Cardinal and ordinal numerals also precede the noun, so the definite article occurs before a number of different word classes, always with the same distribution of the three variants, regardless of word class.

   As shown in (1), the *'r* variant occurs whenever the definite article follows a vowel-final word. The examples in (3) show that this preceding word can also come from a wide variety of syntactic categories: in (a) the word to which *'r* attaches is a preposition, in (b) a copula, in (c) a verb, in (d) a noun, in (e) a pronoun and in (f) a complementizer. In fact, as far as we know, *'r* can attach to any vowel-final word. Furthermore, the syntactic constituency facts are irrelevant to the appearance of *'r*: for instance, in (3e), the article belonging syntactically to the direct object is attached to the pronominal subject, although the subject and object clearly do not form a constituent.

(3) a. *o'r*   llyfr
from-the book
   b. Pwy *ydy'r* meddyg?
who is-the doctor
'Who is the doctor?'
   c. yn   *canu'r* emyn
PROG sing-the hymn
'singing the hymn'
   d. *rhieni'r*   ysgolfeistr
parents-the schoolmaster
'the parents of the schoolmaster'

    e.   Brynaist   *ti'r*   llyfr?
        bought:2SG you-the book
        'Did you buy the book?'

    f.   Mi wn   i [*mai'r*   bachgen sydd yn   sgwennu, nid y
        PRT know I COMP-the boy   REL:is PROG write   not the
        ferch].
        girl
        'I know that it's the boy who's writing, not the girl.'

In other words, *'r* appears to be a simple clitic, a morphologically bound element which is syntactically an independent word. Its distribution illustrates a classic case of the independence of syntax and phonology: phonologically, *'r* is enclitic on any preceding vowel-final item, though syntactically it belongs to the following nominal phrase. In the terms of Klavans (1985), it is a typical "dual citizen".

    The major properties of the article which make it an interesting and superficially rather intransigent problem are threefold. First, the environment for the *'r* form generally takes precedence over the environment for the other two forms, *y* and *yr*: if the preceding word is vowel-final, *'r* must be chosen (with a class of exceptions, which we discuss in Section 3.2). For instance, in (1), the (1c) and (1d) forms *o'r afon* and *o'r llyfr* could not normally be replaced by *o yr afon, *o y llyfr, even though the environment for the allomorph *yr* appears to be met by the following vowel-initial word *afon*, and that for *y* is apparently met by the consonant-initial word *llyfr*. So any account must have a way of ensuring that, in the apparent conflict of environments, *'r* is chosen for preference wherever possible. However, as we will see in Section 3.4, it is not the case that Welsh enclitic forms in general must be selected wherever possible. This means that a blanket account covering all enclitics would not be the solution. The solution which we propose in Section 3 relies on the principle that more specific rules take precedence over general rules — the elsewhere principle.

    Secondly, the *'r* form takes precedence over the other two forms in another way: *'r* typically occurs even in particular environments in which it would seem that it could be avoided. As we will outline more fully in Section 3.4.1, some morphemes have an alternation between a vowel-final variant and a consonant-final variant, the latter occurring before a vowel in the following word. Such an item is *gyda/gydag* 'with'. From the data in (4), however, we see that *gyda* + *'r* is chosen rather than *gydag* + *y/yr*:

(4)   a.   gyda'r nod/*gydag y nod
          'with the aim'

b. gyda'r offer/*gydag yr offer
'with the equipment'

We will see in Section 3.4.2 that this distribution is not an inherent fea-
ture of morphemes such as *gyda*, since it's not the case that in all other
constructions, their vowel-final variants must be chosen whenever possi-
ble. The solution which we will present in Section 4 proposes that if forms
such as *gyda* are inserted at an earlier stage in the derivation than the def-
inite article, then the elsewhere principle will again ensure that *'r* takes
precedence.

   Thirdly, an interesting complication arises from the interaction be-
tween the selection of the correct form of the definite article and the phe-
nomenon of consonantal mutation. "Mutation" in the Celtic languages
refers to various sets of morphophonological changes in the initial conso-
nant of a word, triggered in specific morphosyntactic environments, most
particularly following a large array of function words. One environment
in Welsh involves feminine singular nouns, which undergo the lenition
process known as "soft mutation" when following a definite article. Con-
sider the data in (5), in which feminine nouns in (5a) and (5b) are con-
trasted with a masculine noun in (5c), which bears no mutation in this
context:

(5) a. glasog   y   lasog
         'gizzard' [FEM. SG.]   'the gizzard'
     b. gardd   yr   ardd
         'garden' [FEM. SG.]   'the garden'
     c. glo   y   glo
         'coal' [MASC. SG.]   'the coal'

In an environment for soft mutation, an initial /g/ in the canonical form
of a word deletes, giving zero. In the case of nouns such as *glasog*, which
begin with a consonant cluster, mutation does not affect the choice of the
article, since the word is still consonant-initial after removal of the initial
/g/: the *y* form is then correct, as (5a) shows. However, the prevowel al-
lomorph *yr* occurs not only before nouns which are canonically vowel-
initial, but also before nouns in which the initial vowel results from leni-
tion, as shown by *yr ardd* in (5b). In other words, the post-mutation form
of the noun determines the form of the article. We seem then to have a
paradox, a classic chicken and egg situation. The lenition on the noun
(*gardd > ardd*) is triggered when the noun follows the article, but the cor-
rect form of the article (*y* or *yr*) can't be inserted unless the initial seg-
ment of the noun is known, which depends on whether or not lenition
has applied. Clearly, we have an issue here which must be addressed by

any account of these phenomena.[2] The solution we propose in Section 4 again relies on the idea of a staggered lexical insertion.

Before turning to the main part of the discussion, note that in many cases of phonologically conditioned morphological alternations, the alternants involved can be accounted for using purely phonological means. It might appear at first glance from (1) that this too is a simple morphophonological question, involving the derivation of three allomorphs from a single underlying form — most likely *yr* /ər/ — and thus an alternation which any model of phonology could easily handle. However, using a standard derivational approach, Coates (1987) demonstrates clearly that whichever variant is chosen as the underlying form, the three alternants cannot be successfully derived without resorting either to undesirable extrinsic ordering or to ad hoc stipulations concerning the interaction of rules of ə-deletion and r-deletion.

An alternative purely phonological approach which might seem promising involves the concept of phonologically conditioned suppletion, in which even the occurrence of suppletive allomorphs may be determined by a phonological trigger: see Carstairs (1988). But the problem of *'r* overriding both *yr* and *y* when the environment for *'r* is met does not fall out from treating the selection as phonologically conditioned suppletion in a three allomorphs arrangement. Something further would have to be built in — something presumably nonphonological — in order to force the precedence of the *'r* variant. There also remains the issue of the interaction of the form of the article with initial mutation, which has no obvious solution under this account.

Given the apparent competition between the three forms of the definite article, another obvious phonological path to explore is optimality theory (cf., Prince and Smolensky 1993; Kager 1999), since the resolution of conflicting requirements is exactly what that model should handle best. And in fact, modeling the behavior of the definite article, independent of other considerations, is quite easy: assuming the constraints *Hiatus ≫ *Coda ≫ MaxSeg, ranked in that order, the correct results obtain for the determiner (MaxSeg requires segments present in the input to also be present in the output). But the existence of the three alternants of the definite article is not the only consideration.

First, the constraint ranking required for the definite article yields the wrong results for the alternating morphemes such as *gyda ∼ gydag*, shown in (4). Since the specific final consonant in *gydag* cannot be predicted from any general facts about Welsh, it must therefore be present in the underlying representation, the input. This means that the appearance of *gyda* (in *gyda'r nod*, *gyda'r offer*) rather than *gydag* gives rise to a violation of MaxSeg which would not be incurred if the forms were

gydag y nod, *gydag yr offer*: these are then incorrectly predicted to be the surface forms. Secondly, it is fatal to any purely phonological account of the correct choice of the article that there is a homophonous morpheme in Welsh, a preverbal particle *y/yr* which also has an associated *'r* variant, but which behaves very differently from the definite article. We will see in more detail in Section 3.3 that the *'r* form of this morpheme does not take precedence over the other two forms, but rather, is an optional clitic variant of *yr* (only). Yet presumably the constraints active in selecting from among *y*, *yr* and *'r* for the definite article would predict an identical result in selecting the correct form of preverbal *y/yr* and *'r*. Thirdly, the interaction between initial mutation and the form of the article again receives no obvious solution under an optimality account. And finally, the required ranking of *HIATUS, *CODA above MAXSEG to select the right form of the definite article is precisely the wrong ranking of these constraints for Welsh phonology in general, which does not avoid hiatus, either internally or at word boundaries. Neither does Welsh syllable structure avoid codas. Ranking *HIATUS and *CODA above MAXSEG — even if it did allow the selection of the determiner to emerge — wrongly predicts that it is more important in Welsh to have CV syllables than it is to parse underlying material in the output.

We therefore conclude that even if a purely phonological account were proposed, it could not constitute a perspicuous solution to the issues here, in light of the interplay of phonological, morphological and lexical factors which govern the selection of forms.

3. The definite article and other Welsh function words

In this section we first account for the preference for the *'r* form of the definite article over the other two forms: in Section 3.1, we propose that rather than a single underlying form, the article has two forms in its lexical entry, one of which takes precedence. In Section 3.2 we examine environments which favor the suppression of the *'r* form of the definite article, showing that the phonological structure (specifically, the presence of intonational phrase boundaries) can affect the choice of the form of the article. We then compare and contrast the behavior of the definite article with other minor function words of Welsh, establishing in Section 3.3 and 3.4 that the definite article has particular morphological properties which distinguish it from other function words in the language. The points discussed here will lead in Section 4 to our proposal concerning lexical insertion and its interaction with initial mutation.

3.1.   *Selection of the forms of the definite article*

3.1.1.   *The lexical entry and elsewhere.*   We saw in Section 2 that selection of the definite article is not a straightforward (morpho)phonological problem. However, neither can the distribution of the variants be characterized purely morphosyntactically. Given that subcategorization does not typically refer to the PHONOLOGICAL properties of the syntactic context, we cannot state the distribution of the three forms as part of the morphosyntactic subcategorization of the article. Even if we allowed the phonology to be part of the subcategorization, it would seem desirable, even vital, to restrict this to instances where there is a selectional relationship between the items in question. But the choice of allomorph for the article would have to be sensitive to the phonology of whatever word can follow the article. Rather than merely being sensitive to the phonology of a following noun — which might reasonably be considered to be selected by the article — the article would have to know the phonology of any adjectives, numerals and other functional elements that could immediately follow it. Such items are not generally regarded as selected by the article in any way: adjectives and numerals, for instance, are chosen on the basis of semantic compatibility with the noun, not with the article. More worryingly, the distribution of the forms of the article must be sensitive to the phonology of the preceding syntactic context too, which is needed to ensure that the *'r* form always appears after a vowel. Clearly, the preceding context is not selected by the article in any sense, and so cannot possibly form part of its subcategorization.

Moreover, under a subcategorization account, the major properties of the article outlined in Section 2 would also remain a mystery: there would be no obvious way to predict that the *'r* form takes precedence, and the mutation paradox would still have to be addressed. It seems, then, that the choice of the correct allomorph is neither simply phonological nor morphosyntactic.

We propose, therefore, that the solution to the choice of allomorph rests crucially on the lexical entry for the definite article, understanding a lexical entry to contain all the idiosyncratic properties of a lexical item, including such properties as word class, semantics, paradigm membership, pronunciation, allomorphy and so on. As noted in Section 2, any account taking a single underlying form from which all three alternants are derived is doomed to failure, something which Coates (1987) clearly demonstrated. Our alternative proposal is that the lexical representation for the definite article contains not one but two entries, *'r* and *y(r)*, as shown in (6). Potentially, both of these apply in an identical context, in that both occur before either a consonant or a vowel. However, it has long been

recognized in linguistics that if two or more rules can apply to a particular form, then the more specific rule takes precedence over the more general; this is often referred to as the elsewhere condition or disjunctive ordering (cf., Kiparsky 1973, 1982). The elsewhere principle ensures that *'r* takes precedence here: it is chosen in only a single specific context — namely following a vowel. If this environment is met then the occurrence of *y(r)* is blocked. In all other contexts, the *y(r)* variant occurs, including in utterance-initial position, i.e., when there is no preceding context.

(6)   Def Art =   *'r* / V__ ;
                     y(r) / elsewhere

It is the elsewhere condition which ensures that we get *i'r afon* 'to the river' and *i'r llyfr* 'to the book' rather than **i yr afon*, **i y llyfr*. We thus have a simple account of the first question concerning the selection of the definite article, namely why does the enclitic *'r* form take precedence over both *y* and *yr*, even when the context for the occurrence of one of the two latter forms is apparently fulfilled.

  We argue below (in Section 4.1) that the definite article is subject to late insertion, i.e., it is inserted towards the end of the syntactic derivation. It is at that point that *'r* will be selected, provided that the definite article follows a vowel-final word. If, on the other hand, there is no preceding vowel-final word, then selection will be between *yr* and *y*. The choice between *yr* and *y* is a matter of simple allomorphic variation (much like the selection between *a* and *an* in English). At this point *yr* occurs when followed by a vowel-initial word and *y* occurs when followed by a consonant-initial word.

3.1.2.   *A proclitic form of the article.*   One remaining complication concerning the variants of the article is relevant here. Note that another *'r* form of the article also exists which is not an enclitic at all, but is in fact a proclitic. As David Willis points out to us, in informal colloquial usage we find examples like those in (7):[3]

(7)   a.   Faint   ydi oed *'rhen*   Risiart?
            how.much is   age the-old R.
            'How old's old Risiart?'
            (Hughes 1943: 90)
      b.   *'Rarglwydd* mawr!
            the-lord   great
            'Good lord!'

Proclitic *'r* may also be found in formal registers, for instance at the start of a line of poetry, in order to maintain the meter:

(8) '*R un*   nerth   sydd yn fy Nuw
　　 the same strength REL:is in my God
　　 'the same strength which is in my God'
　　 (Thomas 1996: 768)

Clearly, this form is not an enclitic, since there is nothing in any of the examples in (7) or (8) for an enclitic form of the article to cliticize to — it occurs following a consonant in (7a) and in absolute initial position in (7b) and (8). Neither context would permit enclitic '*r*, and the environment for its appearance shown in (6) is not met.

As a proclitic, article '*r* is not obligatory. In fact, it is merely an (optional) contraction: if it is not chosen, the full form of the article will be *yr* in each case — before <h> in (7a) and before vowels in (7b) and (8). The full form *yr* is completely grammatical in these environments. This is crucially unlike the context for enclitic '*r*: recall that in the case of the enclitic article '*r*, we cannot, for instance, substitute *i yr afon for the grammatical *i'r afon*. So we can conclude that proclitic and enclitic '*r* have entirely different properties and status. Enclitic '*r* is obligatory rather than optional, and cannot be treated as a reduced form of *yr*: see also Section 3.2 below. We have also argued that it must receive its own lexical entry, a conclusion which we will support with other evidence as we proceed. As a proclitic, article '*r* is entirely optional, freely alternating with, and deriving from, the full form *yr*. It has no special lexical status, and does not need to appear in the lexical entry for the article, since it can be derived from *yr* by a low-level phonetic rule. It is in fact exactly parallel to another Welsh clitic, the proclitic form '*r* of the preverbal particle, which we discuss below in Section 3.3.

What the contrasting properties of proclitic '*r* confirm is that Welsh clitics cannot all be treated identically, as simple contractions of a full form. Although the majority of the clitics which we discuss indeed have no special properties (lexical, morphological, or phonological) apart from their clitic status, enclitic article '*r* stands out as an exceptional case.

3.2. *Intonational phrase boundaries: suppression of the enclitic* 'r *form*

At the beginning of Section 3 we indicated that in some cases the enclitic '*r* form may be suppressed. We now illustrate the environments in which the enclitic '*r* form fails to appear, even though its phonological environment appears to be met. This suppression will entail a revision to the lexical entry proposed in (6) above.

Comparing the full forms of the article with the enclitic form, the *'r* form displays a greater phonological dependency than *y/yr*: since *'r* consists solely of a consonant it requires a phonological host.[4] It is also fully stressless and cannot be stressed. If, for example, for reasons of semantic focus one wishes to emphasize the definite article, only *y* or *yr* can surface. That is, *y* and *yr* are stressable, although they do not normally attract stress. As we see in (9)[5], although the phonetic environment for *'r* is met it does not appear, because the context requires focus on the article:

(9) Caffi Morgan *ydy Y* lle   i fynd!
     cafe M.   is   the place to go
     'Cafe Morgan is *the* place to go!'

In (9), where focus is realized through stress, the unstressable form *'r* is replaced with stressable *y* (before the consonant in *lle*). And if the following word is vowel-initial, then as we would expect, the stressed form of the article is *yr*, as in *Caffi Morgan ydy yr unig lle i fynd!* 'Cafe Morgan is *the only* place to go!'. It is clear, then, that since both *y* and *yr* can replace enclitic *'r* in the appropriate context, *'r* cannot be regarded as merely a contraction of *yr*.

Cliticization of *'r* is also normally suppressed when giving a citation form, so that in (10) we do not find **mai'r*; compare (3f), in which cliticization to *mai* has occurred:

(10) er   *mai* 'y pobloedd' a   ddisgwylir
      although COMP the peoples   REL expect:PRES:PSV
      'although *y pobloedd* is expected'
      (Morgan 1952: 10)

The same lack of cliticization typically occurs when the article is part of a name or a title:

(11) *Mae yr* Arolwg Ordnans yn   adolygu 30,000 o fapiau
       is   the survey ordnance PROG review   of maps
       y   flwyddyn.
       the year
       'The Ordnance Survey revises 30,000 maps a year.'
       (www.ordnancesurvey.co.uk/cymraeg/busnes/cynnyrchbusnes1
       .html)

The environment illustrated in (12) is somewhat different: the site at which cliticization fails to occur is the start of a series of conjoined noun phrases, all of which are in apposition to a name:[6]

(12)   *Mae yr* actor, y   cynhyrchydd a'r   cyfarwyddwr ffilm
       is   the actor the producer   and-the director   film
       enwog, Yr Arglwydd Richard Attenborough, wedi canmol
       famous the lord   PERF praise
       'menter ragorol' . . .
       initiative excellent
       'Lord Richard Attenborough, the famous actor, producer and
       director, has praised the "brilliant initiative" . . .'
       (http://sgrinasain.llgc.org.uk/archif024.htm)

In this environment it seems more appropriate to have a full form of the
article, presumably in order to indicate that the series of noun phrases in
apposition all refer to the same entity. The lack of cliticization empha-
sizes the constituency of the article with the following material.

   The *'r* form is also suppressed (and replaced by *y* or *yr* as appropriate)
following a deliberate pause, as Thorne (1993: 97) notes, and in particular
this is likely when the NP containing the article is an adjunct:

(13)   Ni   alwodd neb   *yma* ‖ *y*   dydd o'r blaen   ‖ = pause
       NEG called no-one here   the day   before
       'No one called here the other day.'
       Thorne (1993: 97)
(14)   Rhaid ichi   hefyd astudio'r modiwlau a   restrir
       must to:you also study-the modules   PRT list:PRES:PSV
       yma y   sesiwn hwn.
       here the session this
       'You must also study the modules listed here this session.'
       (www.aber.ac.uk/smba/welsh_ver/students/current_students/mod
       .shtml)

A rather nice example of this environment is also shown at the end of the
sentence in (11), where we find *y flwyddyn*, despite the fact that the article
follows a vowel-final word. The suppression of the *'r* form allows the
phrase meaning 'a year/every year' to be parsed as an adjunct to the
verb, rather than forming a constituent with the preceding noun *mapiau*
(here mutated to *fapiau*). The alternative, *mapiau'r flwyddyn*, has the
form of a possessive noun phrase, meaning 'maps of the year'; the lack
of cliticization in (11) thus prevents a misparse.

   In (9) through (14), various prosodic effects give rise to the suppression
of *'r* in favour of *y* or *yr*, an effect which is clearly related to prosodic
structure, in the sense of Nespor and Vogel (1986). Drawing on evidence
from a number of languages, Nespor and Vogel argue that pauses, paren-
theticals, adjuncts and the edges of listed items (among other structures)
frequently coincide with an intonational phrase boundary, I. The I-

domain (the stretch of speech defined between two I boundaries) is often associated with the semantics of focus; indeed I may be restructured as the result of focus. Comparable effects can be observed in the data in this section, where we see that pauses, the start of a list, and focus all co-occur with the suppression of *'r*; citation forms and titles can reasonably be seen as analogous phenomena. The I-boundary is argued by Nespor and Vogel to block various kinds of segmental rules, such as nasal assimilation in Spanish, and the lenition process of Tuscan Italian known as "Gorgia Toscana". We therefore propose that the suppression of *'r* similarly involves I, and that all the examples in this section involve a (sometimes optional) I-boundary. When *'r* is separated from its phonological host by an intonational phrase (I) boundary it is suppressed: the presence of an I-boundary between the enclitic and its host prevents encliticization. Suppression of *'r* then forces the appearance of *yr* or *y*.[7]

   In light of these facts, the distributional information given in the lexical entry shown in (6) needs to be revised to allow for the suppression of *'r*. Evidently, there is a condition on the occurrence of *'r* which prevents this form from appearing when an intervening intonational phrase boundary separates *'r* from its host.

(15)   Def Art =   *'r* / V__ ;   Condition: there is no intervening intonational phrase boundary separating *'r* from its host.
                        y(r) / elsewhere

With this revision to the lexical entry we can capture both the normal precedence of the enclitic *'r* over *y/yr* and the suppression of the enclitic form in those cases when stress, focus, citation forms, list readings, etc. lead to the otherwise unexpected occurrence of one of the full forms.

   Having proposed that the lexical entry for the definite article is responsible for the observed distribution of the full and enclitic forms, and for the general precedence of enclitic *'r*, we next demonstrate the contrasting behavior of another function word of Welsh, the preverbal proclitic *y/yr* (mentioned briefly in Section 2). The relevance of this particular word is that despite having full forms which are homophonous with the full forms of the definite article, this proclitic has no enclitic *'r* form. This supports our claim that the key to the distribution of the definite article lies in its lexical entry and is not derivable from the general behavior of clitics in Welsh.

3.3.   *The preverbal proclitic y/yr*

As well as the definite article, there is a separate lexical item with two allomorphs *y/yr* in Welsh; this is a preverbal particle (sometimes

considered to be a complementizer) which occurs in finite VSO clauses. In subordinate clauses, use of the particle is restricted to formal varieties of Welsh, but it can occur with any finite verb form[8]; the embedded clauses in (16) and (17) illustrate. Importantly, no enclitic form *'r* is possible:

(16) a. Dywedodd hi [CP *y*   bydd   cyngerdd heno].
          said:3SG   she   PRT be:FUT:3SG concert   tonight
          'She said there will be a concert tonight.'
      b. *Dywedodd *hi'r* bydd cyngerdd heno.

(17) a. Dw i'n   credu [CP *yr* hoffai   Eluned fynd].
          am I-PROG believe   PRT like:COND:3SG E.   go
          'I believe that Eluned would like to go.'
      b. *Dw i'n *credu'r* hoffai Eluned fynd.

The examples in (16) and (17) might appear to show simply that cliticization of the preverbal particle is blocked across a clause boundary, CP, perhaps in an analogous way to the blocking of the enclitic article across an intonational phrase boundary. But this is not the solution, since even when preverbal *y/yr* could cliticize to an initial element within its own clause, it does not, as (18) shows. The particle here fails to cliticize to *lle* 'where', despite the fact that this word ends in a vowel:

(18) a. Yr oedd yn y fan [*lle*   *y*   cyfarfuasai   Martha
          PRT was in the place where PRT meet:PLUP:3SG M.
          ag   ef].
          with him
          'He was in the place where Martha had met him.'
      b. *Yr oedd yn y fan *lle'r* cyfarfuasai Martha ag ef.

Compare this failure of the preverbal particle to encliticize with the behavior of the article, which we have shown to encliticize freely to any preceding item: (3f), repeated here as (19), contrasts neatly with (18) in illustrating cliticization of the article *'r* to the initial element in CP:

(19) Mi wn   i [*mai'r*   bachgen sydd yn   sgwennu, nid y ferch].
      PRT know I COMP-the boy   REL:is PROG write   not the girl
      'I know that it's the boy who's writing, not the girl.'

Moreover, in contrast with (16) and (17), at least some speakers allow cliticization of the article *'r* ACROSS a CP clause boundary, as shown in (20) and (21):[9]

(20) Dw i'n   *credu'r*   ceffyl du   ennillodd y   ras.
      am I-PROG believe-the horse black won   the race
      'I think it was the black horse that won the race.'
      Dw i'n credu [CP *'r* ceffyl du ennillodd y ras]

(21) Dw i'n   *dychmygu'r* plant   drws nesa' sy   wedi torri'r
      am I-PROG imagine-the children door next REL:is PERF break-the
      ffenest.
      window
      'I imagine that it was the children next door who broke
      the window.'
      Dw i'n dychmygu [CP 'r plant drws nesa' sy wedi torri'r ffenest]

What we have in the case of the article and the preverbal particle is a simple case of homophony: two different morphemes have the word forms *y/yr* as allomorphs, but only one of these morphemes, the definite article, has an enclitic form *'r*. Clearly, encliticization is not merely a low-level phonetic phenomenon which is automatically triggered in any [ə] + C function word (see also Section 3.4.2). This bolsters our view that a distinct lexical treatment of the definite article is in order, and further supports our claim that the choice of variants cannot be handled by a simple phonological treatment.

Before leaving the distinction between article and preverbal particle, there is one more type of data we need to consider. At first glance, examples such as (22) appear to indicate that there is an enclitic form of the preverbal particle, apparently cliticized to *lle*:

(22) Yr oedd yn y   fan   [*lle'r*   oedd Martha wedi ei
      PRT was in the place where-PRT was M.   PERF 3MSG
      gyfarfod].
      meet
      'He was in the place where Martha had met him.'

([18] and [22] are taken from two different Bible translations [*John 11:30*] cf., Thorne 1993: 303.)

In fact, the orthographic representation in (22) is misleading. The clitic form in (22) is not an enclitic, but instead, once again we have a PRO-CLITIC form which — just like the proclitic article *'r* shown in Section 3.1.2 — only occurs when a vowel-initial word follows. Most commonly it occurs with vowel-initial forms of the (highly suppletive) verb *bod* 'be'. So as well as the forms *yr oedd* 'was (3SG)', *yr wyf* 'am', *yr wyt* 'are (2SG)' (see Note 8), there are stylistic alternatives *roedd, rw i, rwyt* (represented orthographically either with or without an apostrophe before the initial *'r*). Note that once again only *yr* (and crucially, not *y*) can replace the proclitic form, as is appropriate if the proclitic is a reduced form of the prevocalic variant *yr*.

Three pieces of evidence support our assertion that preverbal *'r* is a proclitic. First, as noted, it only occurs with vowel-initial verb forms.

Compare (22) with (23), where we have the future form of *bod* 'be', which is consonant-initial, so cannot host a proclitic. No encliticization to the preceding word is possible:

(23) a.   Bydd   yn y fan [*lle*   *y*   bydd   Martha wedi ei
            be:FUT in the place where PRT be:FUT:3SG M.   PERF 3MSG
            gyfarfod].
            meet
            'He will be in the place where Martha will have met him.'
       b.   *Bydd yn y fan *lle'r* bydd Martha wedi ei gyfarfod.

If the verbal particle had an enclitic form *'r*, then the environment for its occurrence would presumably be met in (23), since *lle* is vowel-final, yet we find no enclitic. Similarly with other forms of *bod* which begin with consonants, we find *lle y bu* 'where PRT was' and *lle y mae* 'where PRT is' rather than **lle'r bu*, **lle'r mae*.

   Secondly, the *'r*-initial forms of *bod* 'be' occur both following a consonant, as in (24), and also sentence-initially, as in (25):

(24)   Ni   ddywedais i ddim, oblegid *'roedd*   hi wedi rhoi arian
       NEG said:1SG I NEG   since   PRT-was she PERF give money
       i mi.
       to me
       'I said nothing, since she'd given me money.'
(25)   *'Roedd* hi'n   mynd.
       was:3SG she-PROG go
       'She was going.'

In neither case is there is any preceding material for *'r* to be enclitic to, which is compelling evidence that *'r* is not an enclitic at all here.

   Thirdly, if the particle were an enclitic in such data as (22), then the statement of its distribution would have to take into account not only the preceding material, but also the following material: the only forms of *bod* 'be' that co-occur with *'r* are the vowel-initial forms. As far as we know, no other enclitics have restrictions on the following context as well as the preceding context; in fact, this would be a very strange kind of "enclitic". It seems clear, then, that the preverbal particle *y/yr* has just those two allomorphs, plus a proclitic form which is merely an optional reduction of *yr*, and which is phonetically restricted to occurring before vowel-initial verbs.

   Verbs other than *bod* can also take proclitic *'r*, as (26) illustrates with a vowel-initial form of another highly suppletive verb, *mynd* 'go':

(26) I *b'le'r*   awn   ni?
       to what-place-PRT go:FUT:1PL we
       'Where are we going?'
       (Hughes 1943: 144)

That the proclitic form of the particle is not obligatory is illustrated in
(27) with another vowel-initial form of *mynd*:

(27) Dw i'n   meddwl *yr a'*   i ar fy mhen fy hun.
       am I-PROG think   PRT go:FUT:1SG I on my head my self
       'I think I'll go on my own.'
       (Doherty 1993: 51)

The optional nature of the proclitic *'r* is in fact rather an important
way in which the preverbal particle contrasts with the article enclitic *'r*:
as we have noted, the latter cannot typically be considered optional, since
it takes precedence over the other two forms of the article. This distinc-
tion again supports our analysis of the article *'r*. The grammar obliges
definite article *'r* to be chosen wherever possible, resulting in the blocking
of the other forms of the article.

To summarize, the facts concerning proclitic preverbal *'r* are as fol-
lows: (i) crucially, only *yr* but not *y* can replace it; (ii) *'r* is never oblig-
atory in any environment; (iii) *'r* is an optional contraction of *yr*; and
(iv) *'r* is transparently derivable from the full form *yr* by regular pho-
nological processes, and can generally be replaced by it (although there
are considerations of what is appropriate in different registers).[10] The
distribution of the preverbal particle is, then, entirely straightforward.
We propose that it has just a single form in its lexical entry, *yr*. The
preconsonantal form *y* is derived from *yr* through simple allomorphy.
The proclitic form *'r* is also derived from *yr*, as a low-level vowel dele-
tion rule: this proclitic form is just an optional short form of the full
form *yr*.

Compare the enclitic *'r* form of the definite article. (i) It can be replaced
(e.g., under focus) by either *yr* or *y*, as appropriate in the phonetic con-
text; yet (ii) it is essentially obligatory (modulo the conditions for its non-
appearance outlined in Section 3.2 above); see, in particular, the facts in
Section 3.4.1 below; (iii) *'r* is not an optional contraction of *yr*, and can
never simply be replaced by *yr*: if it is replaced by one of the full forms
of the article, the phonetic environment may require *y*; and (iv) *'r* is not
derived from *yr*, since it occurs where *yr* cannot occur: for instance, *o'r
llyfr* vs. **o yr llyfr*. As distinct from the preverbal particle, the informa-
tion that the definite article has a enclitic form *'r* must be part of its lex-
ical entry, as seen in (15).

3.4.   *Interaction of cliticization with other alternations*

3.4.1.   *Enclitic 'r and other alternations.*   Our position is that the enclitic *'r* form of the definite article is selected wherever possible, and is chosen in place of either full form of the article, because it has the more specific structural description of the two forms — *'r* and *y(r)* — in the lexical entry for the article. Hence, the elsewhere condition ensures that the appearance of either of the other two forms is blocked; cf., Zwicky (1987: 220). In this section we give further evidence for this proposal.

As noted in Section 2, the enclitic form of the definite article is even chosen — crucially — when it could apparently be avoided altogether. A small number of functional morphemes, specifically prepositions or conjunctions, have alternants of the kind seen in (28):

(28)   gyda/gydag   'with'
       â/ag   'with'
       tua/tuag   'towards, about'
       na/nag   'than'
       a/ac   'and'
       na/nac   'neither/nor'

Normally, the V-final allomorph (the citation form) occurs before consonants, whilst the C-final allomorph occurs before vowels, as shown in (29a) and (29b):

(29)   a.   gyda gwên
          with smile
          'with a smile'
     b.   gydag eraill
          with   others

We might naturally also expect that *gydag* and the other C-final allomorphs of these function words would appear before the vowel-initial full forms of the definite article, *y* and *yr*. However, instead we find that the V-final allomorphs (*gyda*, *â* etc.) plus enclitic *'r* generally occur:

(30)   a.   gyda'r nod/*gydag y nod
          'with the aim'
     b.   gyda'r iaith/*gydag yr iaith
          'with the language'

For instance, a search of UK Web sites[11] finds 1420 instances of *gyda'r nod*, and none of **gydag y nod*; similarly, there are 438 instances of *gyda'r iaith*, but none of **gydag yr iaith*.

This is further illustrated in (31), with the alternation between *â*/*ag* 'with'; (a) shows the allomorph *ag* occurring as expected before a vowel, and (b) shows the enclitic form of the article, giving *â'r*, and blocking *ag y, as in (c).

(31)   a.   cysylltu *ag*   ymholiadau cwsmer
            contact with enquiries   customer
            'contacting customer services'
       b.   cysylltu *â'r*   cyngor   (2280 instances found)
            contact with-the council
            'contacting the council'
       c.   *cysylltu ag y cyngor   (0 instances found)

Similarly with the morpheme *a*/*ac* 'and': we find the allomorph *ac* before a vowel, as in *pen ac inc* 'pen and ink', but the obligatory occurrence of *a* + *'r*:

(32)   a.   a'r   diwrnod   (71 instances found)
            and-the day
       b.   *ac y diwrnod   (0 instances)
       c.   a'r   iaith   (818 instances found)
            and-the language
       d.   *ac yr iaith   (0 instances; though cf., [36])

Note that the behavior of the enclitic form of the definite article here contrasts neatly with that of the preverbal particle discussed in Section 3.3, and supports our conclusions in that section regarding the status of preverbal *y*/*yr* and its proclitic variant *'r*. We can see from (33) that the sequence '*ac y*' is perfectly well-formed, provided that it's the preverbal particle *y* and not the article:

(33)   a.   ac y   mae hyn yn   cadarnhau'r esboniad
            and PRT is   that PROG confirm-the explanation
            'and that confirms the explanation ...'
       b.   *a'r mae hyn ...

Obviously, if there is no enclitic form of the preverbal particle, as we argued, then the result in (b) is predictable. The form **a'r* can't occur here because preverbal *'r* only exists as a prevocalic proclitic: the lexicon contains no enclitic form of the preverbal particle *'r*, so no such form can attach to the conjunction *a*.[12]

Although there are certainly sporadic examples in which the C-final forms of the morphemes in (28) occur before the article, these typically involve the deliberate suppression of the *'r* form in one of the environments involving an intonational phrase boundary discussed in Section

3.2. For instance, we find examples such as (34) through (36), where the NP containing the article is a title:[13]

(34)   Fel *gydag Yr* Ochr Arall . . .
       as   with   the side   other
       'As with *The Other Side* . . .'
       (www.bbc.co.uk/cymru/celf/eisteddfod03/ackroyd_harvey.shtml)
(35)   gweithio *gydag y*   GIG Cymru
       work   with   the   Wales
       'working with the GIG (National Health Service) in Wales'
       (http://www.acc.cymru.gov.uk/hysbysiadau/2001/acc2001_5hw
       .htm)
(36)   dau adroddiad, sef   "TTG" *ac* "*Yr* Iaith   Gymraeg
       two report   namely   and the language Welsh
       a'i   Chymunedau"
       and-its communities
       'two reports, namely *TTG* and *The Welsh Language and its
       Communities*'
       (www.penrhyndeudraeth.com/cymraeg/minutes/20011217.htm)

Setting aside these special contexts, it is clear that the encliticized form of the article takes precedence over a nonreduced form. Why though should selection of *'r* take precedence over the appearance of the C-final allomorphs of the words in (28)? As we outline in Section 4, our proposal is that this results from the interaction of the lexical representations of both the article and the words in (28), along with a specific view of lexical insertion. In the following section we examine how the alternating morphemes seen in this section behave in construction with other items that have an enclitic form, in order to see if the behavior of the article *'r* is typical or exceptional.

3.4.2.   *The* yn *morphemes and their enclitic forms.*   An obvious question is whether the fact that *'r* takes precedence over the other two forms of the article is related to the fact that *'r* is an enclitic. In this section we see that the two facts crucially CANNOT be related, since, as we outline, there are other enclitics (the *yn* morphemes with enclitic *'n*) which do NOT take precedence over full forms of the morpheme.

The alternation between full form and enclitic is not restricted to the definite article: a number of other functional morphemes in Welsh have similar alternations. For instance, there are two different morphemes which both have the (full) word form *yn* [ən] (which occurs following a consonant) and an enclitic variant *'n* which occurs with a vowel-final word as host, just like the definite article *'r*:[14]

(37)   progressive *yn* / *'n*[15]
    Mae *hi'n*   canu.   (cf., Mae Gwen *yn* canu 'Gwen is singing.')
    is   she-PROG sing
    'She's singing.'
(38)   predicative *yn* / *'n* before predicate adjectives/adverbs and nouns
    a.   Mae *hi'n*   hapus.   (cf., Mae Gwen *yn* hapus 'Gwen is
         is   she-PRED happy   happy.')
         'She's happy.'
    b.   Mae Gwen yn   *canu'n*   dda.
         is   G.   PROG sing-PRED good
         'Gwen sings well.'
    c.   Mae *o'n*   ddarlithydd prifysgol.
         is   he-PRED lecturer   university
         'He's a university lecturer.'

Since they both have enclitic variants, an obvious question is how do these two *yn* morphemes interact with morphemes such as *gyda/gydag*, listed in (28), which have C-final and V-final allomorphs? We showed in Section 3.4.1 that the enclitic form of the article, *'r*, takes precedence when in construction with forms such as those in (28), so that we find a vastly greater preponderance of *tua'r*, *gyda'r* etc. than we do *tuag y/yr*, *gydag y/yr*. As we noted, the V-final allomorph plus the encliticized article is the norm.

Not all the potentially relevant items have a syntax that is compatible with either *yn* morpheme, or else they co-occur so infrequently that no firm conclusions can be drawn. For instance, *gyda* collocates occasionally with *yn agos* 'PRED close', as in *gydag yn agos i fil* 'with close to a thousand': out of a handful of instances found on UK Web sites, all but one takes the form *gydag yn*, rather than *gyda'n*. It appears, then, that enclitic *'n* does not take precedence in this environment, but the available data are few.

Two of the remaining alternating morphemes in (28) give a better indication, since they do commonly co-occur with both the progressive marker *yn* and the predicate marker *yn*: the disjunction *na/nac* and the conjunction *a/ac*.

The logical possibilities are shown in (39):

(39)   a.   C-final form + yn (e.g., *nac yn*, *ac yn*)
      b.   V-final form + *'n* (e.g., *na'n*, *a'n*)

What we find is that contrary to the situation with enclitic *'r*, enclitic *'n* definitely does not take precedence: the full forms in (39a) are vastly preferred. Thomas (1996: 769) comments that the *'n* forms do NOT occur with *a* and *na*; we therefore expect examples such as those shown in (40):

(40) a. *nac yn*   anfodlon
      nor PRED unhappy
      'or unhappy'

   b. sicrhau nad ydym   yn   mynd yn   sâl *nac yn*   cael ein
      ensure NEG are:1PL PROG go   PRED ill nor PROG get 1PL
      gwenwyno
      poison
      'to ensure that we don't become ill or get poisoned'
      (www.consumereducation.org.uk/laws/cymraeg/hawliau/06
      .htm)

For instance, a search of UK Web sites finds 9 instances of the string *nac yn anfodlon*, as in (40a), but only one of the form with enclitic *'n*, i.e., *na'n anfodlon*; similarly, there are 94 instances of the string *nac yn cael*, as shown in (40b), but none at all of the equivalent cliticized form *na'n cael*. Occasional deviations from the norm can be found, as shown in (41), but these are quite rare:[16]

(41) a. Ni fydd   y   CRB yn   derbyn copi *na'n*   gweld
      NEG be:FUT the   PROG receive copy nor-PROG see
      y   cynnwys.
      the contents
      'The CRB will neither receive a copy nor see the contents.'
      (www.disclosure.gov.uk/welsh/docs/html_reg_matters/
      disclosurenews2/information.html)

   b. Nid wyf yn   wraig, yn   blentyn, yn   ŵyr   *na'n*
      NEG am PRED wife   PRED child   PRED grandson nor-PRED
      wyres   iddo.
      granddaughter to:3MSG
      'I am not his wife, child, grandson or granddaughter.'
      (demserv.powys.gov.uk/welsh/election/postalv/proxy.pdf)

It is clear from the facts surrounding *a* and *na* that enclitics in general do not always take precedence where there is an alternative C-final form[17], which means that the precedence of enclitic article *'r* does indeed appear to be a special characteristic of that particular morpheme. The precedence of *'r* does not, then, simply follow from the fact that it is an enclitic, thus lending further support to our proposal that the precedence of *'r* is encoded in the lexical entry of the definite article. This contrasts with the enclitic *'n* forms: unlike *'r*, these do NOT appear in the lexical entry for the *yn* morphemes, but are simply derived by a low-level phonetic process from the single underlying form *yn*. Hence, the elsewhere principle has no role to play in the selection of *'n* at all, which is why the *'n* forms do not take precedence over the full *yn* forms.

At the beginning of this section we proposed that the distribution of the forms of the definite article is determined by the lexical entry for the article, which encodes the precedence of the enclitic *'r* form via the elsewhere condition. We also noted that the enclitic form of the article is subject to suppression through blocking by an intonational phrase boundary. An examination of other Welsh function words, such as the proclitic form of the article, a homophonous preverbal particle, progressive and predicative *yn*, has shown that these words behave differently from the definite article in relevant respects. This establishes that the properties of the definite article cannot be attributed to the general behavior of clitics in Welsh. Finally, we discussed a set of words which alternate between a vowel-final and a consonant-final form and the interaction of those words with both the definite article and with other clitics.

From this section we can conclude that, unlike clitic forms associated with other function words of Welsh, the enclitic article *'r* has a privileged status: it is not optional, it does not alternate freely with the full forms of the article, and it cannot be derived from either full form, since it replaces both of them (including the *r*-less *y* form). Most importantly, enclitic *'r* actually takes precedence over the nonreduced allomorphs, so that it has to occur wherever its environment is met.

In Section 4 we discuss the interaction between initial mutation and selection of the definite article and other function words, and also the question of why the *'r* form of the article takes precedence when in construction with morphemes such as *gyda*. This discussion will have a bearing on the final piece of the puzzle surrounding the definite article, namely, lexical insertion.

4. Lexical insertion and morphological interactions

In this section we address crucial aspects of the selection of the definite article, namely the question of lexical insertion, as well as how initial mutation interacts with the choice of the form of the article. We argue in Section 4.1 that selection of the correct form of the definite article follows the lexical insertion of content words during the syntactic derivation, and that this allows the definite article to be chosen straightforwardly on the basis of its surrounding phonological environment as encoded in its lexical entry. In addition to addressing the problem of the selection of the definite article, our proposal for a staggered lexical insertion also sheds light on aspects of the initial mutation system which are otherwise unaccounted for. In Section 4.2 we present our solution to the issues raised in Section 3.4.1, involving the interaction of the article with alternating

morphemes such as *gyda/gydag*, where the enclitic *'r* form of the article again takes precedence.

4.1.   *Lexical insertion and initial mutation*

In recent years it has been proposed in a variety of frameworks that rather than filling the slots available in a syntactic derivation in a single operation, lexical insertion in fact occurs in a staggered fashion, at various points in the derivation (cf., Halle and Marantz 1993; Marantz 1995; Jackendoff 1997; Emonds 2000, 2002). Basically, content words are inserted early in the derivation, whilst function words are inserted at a later stage.

   The summary below, adapted from Emonds (2002: 251–260), indicates the three levels of insertion he argues for:

 – First level of insertion (Deep Insertion): involves lexical items associated with a purely semantic feature *f*, i.e., open-class items and idioms, which must be inserted at the outset of derivations. This includes nouns, verbs, adjectives and possibly, some prepositions.
 – Second level of insertion (Syntactic Insertion): if an item contains no purely semantic feature *f*, but has "cognitive syntactic features F_i" which contribute to interpretation, e.g., *to* as a PATH marker, it is inserted during the syntactic derivation leading to LF.
 – Third level of insertion (Late Insertion): this involves items which lack both purely semantic features and also interpretable instances of F, or else which serve purely as a place-holder for predictable values of F. These are inserted at PF and are "absent during the derivation from underlying structure to LF". An example would be case-assigning *of* in *destruction of the city*.

We will assume the basic correctness of this in what follows. Given a staggered lexical insertion, nouns and other content words are subject to deep insertion, whereas the definite article is, we propose, inserted later in the derivation. With this idea in place, we are now in a position to address the dilemma concerning the interaction of initial mutation with the selection of the form of the article, briefly outlined in Section 2.

   Recall that the selection of *y/yr* depends on whether the following word is vowel-initial or consonant-initial. Although the canonical form of a word may be consonant-initial, a mutation environment can change this. Specifically, since an initial /g/ deletes under soft mutation, this may result in an underlyingly C-initial word becoming V-initial. One context in which soft mutation is triggered is on feminine singular nouns following

the definite article. The data in (5a) and (5b), repeated here as (42), illustrate:

(42) a. glasog   y   lasog
          'gizzard' [FEM. SG.] 'the gizzard'
      b. gardd   yr   ardd
          'garden' [FEM. SG.] 'the garden'

The paradox, then, is that selection of y/yr in (42) must be sensitive to the phonological shape of the feminine singular noun resulting from soft mutation, yet the noun only undergoes the mutation because it follows the definite article. The article is needed to trigger the mutation, but the correct form of the article cannot be selected until the mutation has applied. Given the assumption that content words are subject to "deep insertion", but that the definite article inserts later, we propose that there is a point in the derivation in which we have an empty Det position followed by the noun. Under this proposal, the mutation can then be triggered by the (as yet, unfilled) determiner position, thus for (42b):

(43) [__]$_{Det}$ [gardd]$_N$ → [__]$_{Det}$ [ardd]$_N$

The selection of the correct form of the determiner (in this case, the pre-vowel form *yr*) is straightforward. Since it can now "tell" whether the following word is C-initial or V-initial, the correct allomorph (*y* or *yr*) can be inserted as appropriate. (Obviously, if the context preceding the article slot is vowel-final, then *'r* will insert under the elsewhere principle as outlined in Section 3.)

   This proposal for an empty lexical slot receives crucial support from two other aspects of the Welsh mutation system, for which it allows a straightforward and perspicuous treatment. First, it is commonplace in Colloquial Welsh for the mutations triggered by function words to be maintained, even when the function word itself is absent. It therefore seems that for independent reasons we need an account in which slots, rather than their lexical occupants, trigger various mutations. Second, the slots occupied by these nonovert mutation triggers can, moreover, be demonstrated to exist, since each of them BLOCKS any other preceding item from triggering its own mutation. Mutations are strictly local, in the sense that only an immediately preceding trigger can affect the following word. If a slot which is not overtly filled then serves to block a mutation by what looks superficially like an immediately preceding OVERT trigger, then we have good evidence that the empty slot really exists.

   To illustrate the first point, the maintenance of mutation in the absence of an overt trigger, consider the data in (44) and (45), where the function

words that trigger specific mutations are shown in parentheses, and the immediately following item is shown to receive the mutation:

(44) Tybed (*a*)   *ddylech*   chi fynd.   (dylech + soft mutation)
      wonder INTERROG should:2PL you go
      'I wonder if you should go.'

(45) Beth wyt ti'n   (*ei*) *ddarllen*? (darllen + soft mutation)
      what are you-PROG 3MSG read:NONFIN
      'What are you reading?'

In (44), the interrogative marker *a* is a trigger for soft mutation, and the mutation occurs whether or not the slot for the trigger is overtly filled. In (45), the agreement proclitic *ei* (3MSG) is a trigger for soft mutation, which again occurs even if the agreement marker is absent. We assume that the actual lexical insertion of the relevant function words is not obligatory, since typically, these function words occur overtly in the more formal literary language, but not in the (essentially spoken) colloquial language. However, the mutations triggered by unfilled slots in general are robust, so that mutation occurs whether or not an overt trigger is present. In fact, a very prevalent feature of colloquial Welsh is the presence of mutations which do not have an overt trigger.

The second point, the blocking effect by a nonovert mutation trigger, is illustrated in (46). This string contains a trigger for another type of initial mutation: the morpheme *yn* 'in' triggers nasal mutation (nasal in standard Welsh; some speakers generalize soft mutation to this context).

(46) a. yn ein tŷ   ni
          in 1PL house us
          'in our house'
      b. yn tŷ ni
          'in our house'
      c. *yn nhŷ ni

In (a), the word immediately following *yn* does not have a mutable initial segment, so no mutation can be triggered by *yn*. And the 1PL marker *ein* is not a mutation trigger, so in (a) the noun tŷ 'house' receives no mutation. In (46b), though — a more informal variant — 1PL *ein* is absent, which means that the noun tŷ now appears to immediately follow the trigger for nasal mutation, *yn* 'in'. The noun does have a mutable initial consonant, so we might expect it to bear nasal mutation: *nhŷ*. However, despite the surface adjacency, no mutation occurs, as (c) makes clear. Given the presence of unfilled slots, argued for above, we can say that the position for the 1PL marker *ein* blocks the nasal mutation from occurring. Note that the string *yn nhŷ* is perfectly feasible in a different context,

as for instance in *yn Nhŷ'r Cyffredin* 'in the House of Commons', where there is no determiner slot blocking the mutation on the noun.

Thus, the assumption that certain positions in the derivation may remain unfilled sheds some light on other parts of the grammar, as well as solving the paradox concerning the form of the article and its interaction with mutation.

To summarize our discussion to this point, we argue that lexical insertion of content words occurs first in the derivation, but that slots for function words must be present throughout the derivation, whether or not they are filled overtly, since these play a crucial role in triggering — and blocking — initial mutation. Mutations are triggered by the function word slots,[18] some of which are already filled, others of which are to be filled (sometimes optionally) at a later stage in the derivation. Then, crucially FOLLOWING mutation, *yr* or *y* is selected to fill Det, depending on the phonological shape of the noun or other following word. This last step would, of course, yield *'r* in the case of a vowel-final word to the left of Det.

It seems clear that mutation must apply before any items which undergo late insertion appear in the derivation: this is why the form of the article is sensitive to the output of mutation. Given the proposal for a staggered lexical insertion, this in turn makes an interesting prediction. Logically, we would expect that there might be words with alternating C-final and V-final forms, where the choice of allomorph is again dependent on the following phonetic context, but which are inserted BEFORE mutation applies. The prediction then is that their shape should NOT be sensitive to the outcome of the mutation. In other words, along with *y/yr*, the shape of which is only determined by the post-mutation environment, there ought to be items whose shape is only determined by the premutation environment. And indeed, there are such items. The clausal negation markers *ni* (main clauses) and *na* (subordinate clauses) occur in those forms before consonants, but also have the allomorphs *nid* and *nad* before vowels. Both morphemes have the same behavior, but we illustrate with *nad*, which, unlike *ni*, occurs in the spoken language. First, we show the basic distribution of each allomorph, with *na* before a C-initial verb in (47) and *nad* before a V-initial verb in (48):

(47) Dywedodd Aled *na*   ddylai   Mair fynd.
      said:3SG   NEG should:3SG   go
      'Aled said that Mair ought not to go.'

(48) Dywedodd Aled *nad* aeth   Mair ddim.
      said:3SG   NEG went:3SG   NEG
      'Aled said that Mair did not go.'

We assume that the verbs, as content words, are subject to lexical insertion at the first level, as discussed above. Subsequently, functional items including conjunctions or complementizers such as *na* are inserted. Given that the initial segment of the verb is visible when *na* is inserted, the correct shape of the negation marker appears — *na* or *nad* as appropriate — in (47) and (48).

If the definite article is inserted at the third level, which is following the operation of mutation, then the evidence leads to the conclusion that *na* must be inserted at the latest at the second level, namely before mutation applies in the derivation. The conjunction/complementizer *na* is a trigger for soft mutation, which means that a canonically /g/-initial word can become vowel-initial following *na*: for instance, under soft mutation, the verb *gofynnodd* 'asked' becomes *ofynnodd*. But now note that the shape of the morpheme *na* itself — entirely unlike the *y/yr* alternation — is not sensitive to the postmutation form of the following word at all:

(49)   Dywedodd Aled *na*/**nad* ofynnodd Mair ddim.
      said:3SG   NEG   asked:3SG   NEG
      'Aled said that Mair did not ask.'

So, rather than the prevocalic form *nad*, in fact the preconsonantal form *na* occurs (and mutatis mutandis for *ni* and *nid*). All this behavior falls out from the assumption that different items are inserted at different levels in the course of the derivation. Crucially, the insertion of *na* (and selection of its correct form) must occur BEFORE mutation whilst the insertion of the definite article (and selection of its correct form) must occur AFTER mutation:

(50)   Proposed derivation for *na ofynnodd* 'did not ask'[19]
      syntactic structure   [Conj   V]
      i)   insertion of content word   __   gofynnodd
      ii)   insertion of conjunction   na   gofynnodd
      iii)   triggering of mutation   na   ofynnodd
(51)   Proposed derivation for *yr ardd* 'the garden'
      syntactic structure   [Det   N]
      i)   insertion of content word   __   gardd
      ii)   triggering of mutation   __   ardd
      iii)   insertion of article   yr   ardd

Note that once *na* is inserted, it crucially cannot "go back" following mutation to reselect the C-final allomorph *nad*.

The significance of the data discussed in this section is threefold. First, these data support the notion that a number of morphophonological alternations in Welsh cannot be seen as purely low-level phonetic processes:

the complex interactions with initial mutation demonstrate this point. Second, the data provide evidence for different levels of lexical insertion, and they indicate that this is not simply a distinction between content words vs. function words, but suggest (given the mutation facts) that even among function words, insertion must apply at different points in the derivation. And third, the data indicate that lexical slots for function words can remain empty until late in the derivation, or may even remain unfilled entirely, but nonetheless have a nontrivial presence, since these slots serve both to trigger and to block initial mutation.

4.2. *Lexical insertion and the precedence of the* 'r *enclitic article*

We are now in a position to address the remaining question concerning the form of the definite article, which was raised in Section 3.4.1. This involves the interaction of *'r* with the morphemes which have a V-final and a C-final allomorph, such as *gyda* and *gydag*, *â* and *ag*: see (28). We saw in that section that *'r* in general takes precedence, cliticizing to the V-final form, so that we find, for instance, forms like those in (52):

(52) a.   gyda'r nod/*gydag y nod
            'with the aim'
       b.   cysylltu â'r   cyngor/*cysylltu ag y cyngor
            contact with-the council
            'contacting the council'

The final consonant in forms such as *gydag* can be viewed as "latent", like the final consonants in French words such as *petit* 'small' and *dans* 'in': these are phonetically absent in isolation and in their citation form, *petit* [pəti], *dans* [dã], but the consonants may (in some cases must) show up when followed by a vowel-initial word, *petit ami* [pətitami] 'boyfriend', *dans un instant* [dãzœ̃nɛ̃stã] 'in an instant'; on the complexities of French latent consonants see Walker (2001). Since the particular final consonant associated with each word is unpredictable, it must be given in the lexical representation of that word. The consonants are latent in the sense that they appear in specific (morphophonological) contexts, but are absent in others. The Welsh words in (28) are entirely parallel: the citation form of the word is the vowel-final alternant, *gyda* etc.; the consonant-final version only appears before a vowel. We assume that the default forms, *gyda* etc., are the forms which are inserted into the derivation, unless a vowel-initial word immediately follows, in which case the *gydag* form will appear.

Conjunctions and prepositions such as *gyda* 'with', *â* 'with' and the remaining items in (28) must be inserted by the second level of insertion, as argued above, whilst the definite article is, as proposed above, subject to late insertion, i.e., at the third level. For argument's sake, we assume here that the preposition *gyda* is inserted at the second level (as well as the other alternating items), but it would not affect our point if they were subject to deep insertion. In (53), the V-final allomorph *gyda* is selected because the immediately following position — the determiner slot — is EMPTY, rather than being vowel-initial. When the article is inserted, it will follow a vowel that is already in place, and hence, the *'r* form will be chosen by the elsewhere principle:

(53)   Proposed derivation for *gyda'r nod*.

syntactic structure	[Prep	Det	N]
i) insertion of content word			nod
ii) insertion of preposition	gyda	__	nod
iii) insertion of article	gyda	'r	nod

Although we have assumed that the citation form, i.e., the V-final form *gyda* etc., is inserted into the derivation, note that this is not crucial: our analysis works equally well if the C-final form *gydag* etc. is inserted, providing that the final consonant deletes unless a vowel immediately follows. Since a vowel does not immediately follow *gyda(g)* in (53) — but instead, the empty determiner slot follows — then the /g/ would in any case delete, and again, the *'r* form of the article will be chosen under the elsewhere condition.

Of course, if there is no determiner, then it will not have any slot at all, and typically the preposition then immediately precedes a noun, or other open class item, which is already inserted in the derivation. Since the initial segment of such an item is already known, then the correct allomorph (*gyda/gydag* etc.) can be selected straightforwardly. Thus we find, for instance, *gyda gwên* 'with a smile' but *gydag eraill* 'with others'. If the preposition etc. is a mutation trigger, then mutation will occur on any item with a mutable initial consonant which immediately follows the preposition.

An alternative possible derivation is that some lexical items which immediately follow the *gyda* type of lexemes may be inserted at the same level as they are. If this is the case, then the correct allomorph must also be able to be selected when lexical items are inserted simultaneously.[20] For instance, it might be the case that numerals are inserted at the same level as these prepositions and conjunctions, and indeed, we find the same allomorphic variation before V-initial and C-initial words: *gydag un* 'with one' vs. *gyda tri* 'with three', for instance.

In this section we have shown how the elsewhere principle interacts with a staggered lexical insertion to ensure that the correct form of the definite article is selected. It was the elsewhere principle which we proposed in Section 3 to be the source of the precedence of *'r* over *y* and *yr*, but it also crucially handles the occurrence of *gyda'r* rather than **gydag y(r)*, since, IN CONJUNCTION WITH late insertion, it is elsewhere that ensures that *'r* appears after *gyda*. Late insertion of the article is crucial to the account of the appearance of *gyda'r*, but we have convincing evidence for late insertion, coming from the mutation contrasts discussed above in Section 4.1. And late insertion also handles the choice between the *y* and *yr* forms of the article in the postmutation context.

What we have now is a unified account of all the facts surrounding the choice of the article. This is not an account with a single strand — since two mechanisms are required — but both mechanisms play an integrated role in our analysis. We have thus addressed all three of the issues concerning the definite article which were introduced in Section 2: the precedence of the enclitic *'r* form over the two full forms of the article; the precedence of *'r* in forms such as *gyda'r* rather than *gydag y(r)*; and the mutation paradox concerning the choice of *y* or *yr*.

5. Conclusion

We have examined in detail the distribution and properties of the three surface forms of the Welsh definite article, showing that an apparently simple arrangement of three allomorphs turns out in fact to be more complex. The definite article displays an interesting and intricate set of interactions with other forms in the grammar, and often contrasts with the behavior of other small functional elements which look superficially similar.

Our proposal for a solution to the various issues that arise involves the use of the elsewhere principle, as well as a version of trilevel lexical insertion. This machinery is neither novel nor untested, but it should be noted that our use of the elsewhere principle is morphological, choosing between two entries in the lexicon, rather than phonological, which would entail choosing between three surface forms, something we have argued to be untenable in the case of the definite article.

Our analysis also provides a solution to some issues concerning the Welsh mutation system that were previously unaccounted for, as well as untangling some intricate questions with regard to the interactions of various functional items in the language. Small details can often shed light

on larger issues, and we hope that these results might suggest avenues for further work at the interfaces of grammar.

Notes

1. This article has benefitted from the many valuable comments provided on numerous earlier drafts by the following people: Bob Borsley, Andrew Carstairs-McCarthy, Joe Emonds, Tonio Green, Bob Morris Jones, Andrew Spencer, Greg Stump, Nigel Vincent, David Willis, and Arnold Zwicky; none of the above is responsible for or necessarily agrees with our analysis. Thanks also to two reviewers for *Linguistics*, and to Lewis Davies and Bob Morris Jones for help with the Welsh data. Errors remaining are our own. Correspondence address: School of English Literature, Language and Linguistics, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, U.K. E-mail: s.j.hannahs@ncl.ac.uk, maggie.tallerman@ncl.ac.uk.
2. For completeness, we should add that again the *'r* form takes precedence, so that when a vowel-final word precedes the article we get *o'r lasog* 'from the gizzard', *o'r ardd* 'from the garden', as well as *o'r glo* 'from the coal'.
3. An anonymous reviewer notes that (7b) is one of a small set of exclamations which take this form, and which do not reflect a productive process in the language; presumably, these are lexicalized expressions. Other examples with a proclitic include *'rargian fawr!* 'good grief!' (<*yr argian*) and *'nenw Tad* (<*yn enw Tad*), literally 'in the name of the Father'.
4. Andrew Carstairs-McCarthy suggests to us that maybe *'r* does not contain any material suitable for a syllabic nucleus, but in fact [r] can be a nucleus in Welsh; cf., bisyllabic *Lloegr* [ɬɔɨgr] 'England'.
5. Thanks to Bob Morris Jones for the example in (9).
6. Note though that the set of conjoined phrases in (12) does contain an instance of *'r*, in *a'r*; the critical factor appears to be that this does not fall at the start of the list.
7. For completeness, we should add that there is one environment in which *'r* sometimes fails to appear, but which does not seem to involve an I-boundary. This involves the expressions *yr un* 'the same' and *yr unig* 'the only', typically (though not exclusively) when they appear at the start of NP/AP predicates in copula constructions. Examples where cliticization does occur, such as (i) through (iii), are easily found:

   (i)   Mae e'r   un fath.
         is   he-the one type
         'It's the same.'
   (ii) Roedd hi'r   un mor frwdfrydig.
         was   she-the one so   enthusiastic
         'She was just as enthusiastic.'
   (iii) y   mae hi'r   unig un o'r   cymeriadau
         PRT is   she-the only one of-the characters
         'she's the only one of the characters'
         (www.bbc.co.uk/cymru/adloniant/ffilm/010728jpark.shtml)

We also find occasional instances where it does not, as illustrated in (iv) through (vii):

(iv) pan   oedd *o* *yr* un oed a ni
when was   he the one age as we
'when he was the same age as us'
(www.cynnal.co.uk/llythrennedd/Plant/silff/awdurmaw.htm)

(v) Bydd *hi* *yr* un mor hawdd.
be:FUT she the one as   easy
'It will be just as easy.'

(vi) cyfarfodydd lle   roedd *hi* *yr* unig siaradwr Cymraeg
meetings   where was   she the only speaker Welsh
'meetings where she was the only Welsh speaker'
(www.conwy.gov.uk/C_MINUTES/c_panels/c_corporate_affairs/c_minutes/
c_pdf/060300ci.pdf)

(vii) Mae llawer o bobl   yn   *teimlo yr* un fath.
is   many of people PROG feel   the one type
'Many people feel the same way.'

However, an anonymous reviewer for *Linguistics* comments that actually "spontaneous speech would produce *'r* here", and suggests that the fact of these examples being written may affect the perception of what is correct. In any case, the specific environment does not seem to be one that is generally associated with a lack of cliticization, since other function words with clitic variants (see Section 3.4) normally cliticize in this context. Another possible reason for the suppression of cliticization in some of these instances may again be the likelihood of misparsing. For instance, the strings *o'r un*, *o'r unig* (see (iv)) are technically ambiguous between 'he-the one', 'he-the only', and 'of-the one', 'of-the only'; however, cliticization of the article essentially forces the latter readings, perhaps because the preposition *o* forms a constituent with the following material.

8. In main clauses, the preverbal particle *y/yr* is restricted to occurring with two tenses of the verb *bod* 'be', namely the present (i) and the imperfect (ii), and cannot occur with other finite verbs.

(i) *Y* *mae* pob cwrs yn   cynnwys theori ac ymarfer.
PRT is   every course PROG include   theory and practice
'Every course includes both theory and practice.'

(ii) *Yr* *oedd* yn   amlwg.
PRT was   PRED obvious
'It was obvious.'

The preverbal particle in these contexts does occur in both formal and informal varieties of Welsh.

9. These examples are nonstandard: in more formal varieties of Welsh, embedded CPs with a focussed XP must begin with a complementizer such as *mai* or *taw* (cf., [19]). Some speakers react against the absence of the complementizer in examples such as (20) and (21). Since the NP containing the article is focussed there is also an obvious tendency to suppress the cliticization for just that reason (see Section 3.2).

10. Here we essentially concur with the analysis of Zwicky (1985: 300–302) who also considers preverbal *y/yr* and *'r*, and concludes that the full forms are independent words whilst *'r* is a simple clitic variant.

11. We note in the remainder of this section further findings from a search of UK Web sites, discounting irrelevant or obviously incorrect data whenever possible. The results

should be taken as indicative, rather than absolutely unassailable, and they have been confirmed as far as possible by checking with native speakers.

12. As an added complication, the *ac* allomorph can in fact occur before all forms of the preverbal particle, not only before *y* as in (33), and *yr*, as we would expect, but also, more surprisingly, before the contracted (procliticized) form *'r*. So we find both *ac yr oedd* and *ac 'roedd* for 'and there was', as well as the less standard *a 'roedd*. For histor- ical reasons the *ac* allomorph also occurs before various consonant-initial forms, so that we also find *ac mae* 'and is', *ac nid oes* 'and is not', and so on. The *ac* forms seem to be stylistically preferred in these contexts; cf., Thomas (1996: 769).

13. We note that occasional examples of *tuag y/yr* (i.e., rather than *tua'r*) can be found, together with sporadic examples of C-final variants of the other morphemes in (28) + *y/yr*. Although we cannot offer any particular explanation for this, we empha- size that *tua'r* is still the vastly preferred form and the far more prevalent form: it has around 2420 instances in UK Web sites, as opposed to around 14 of *tuag y* and around 5 of *tuag yr*. The same is true for all the morphemes of (28): the *'r* form of the article is the norm, co-occurring with the V-final allomorphs.

14. Two further homophonous *yn* morphemes exist which we will not discuss. Welsh dia- lects have the informal *yn* 'my', which does have an enclitic form *'n*, as in *lle mae'n het i?* 'where is my hat?'; there is also an informal variant *'y*, as well as the formal *fy*, as the 1SG possessive form. A fourth morpheme is *yn* 'in', which does not, at least in prescrip- tive terms, have a clitic variant *'n*, though this does occur occasionally in informal us- age. These two morphemes both trigger nasal mutation, whilst progressive *yn* triggers no mutation and predicative *yn* triggers soft mutation; it is uncontroversial that all four morphemes are distinct.

15. An anonymous review for *Linguistics* asks whether the cliticization occurring with the progressive aspect marker is analogous to the contraction occurring with the perfect as- pect marker *wedi* (>*'di*), as in *Mae hi'di canu* 'She has sung'. However, in fact this is a different phenomenon: the contracted form *'di* cannot be considered to be an enclitic at all, as is clear from the fact that it can occur in absolute initial position, as in (i):

(i) 'Di   colli'n   llais.
    PERF lose-1SG voice
    '(I've) lost my voice.'

16. We can be sure that this lack of cliticization does not, for instance, have anything to do with the fact that *a* and *na* are co-ordinating conjunctions, since the conjunction *neu* 'or' can host both of the enclitic *'n* morphemes seen in (37) and (38):

 (i) Ydyn   nhw'n   canu *neu'n*   dawnsio?
     are:2PL they-PROG sing or-PROG dance
     'Are they singing or dancing?'
(ii) yn   frawd *neu'n*   chwaer
     PRED brother or-PRED sister
     'brother or sister'

Neu is V-final and does not have a C-final allomorph, but since Welsh does not consis- tently avoid hiatus, the alternative *neu yn* is an acceptable form. The enclitic *'n* forms are more common than the alternative *neu yn*, though both forms occur fairly freely. For instance, a search of UK Web sites finds 124 instances of the string *neu'n defnyddio* 'or PROG using', but only 29 of the string *neu yn defnyddio*, with no cliticization.

17. If it were the case that enclitic forms always took precedence over full forms, we might invoke as an explanation something along the lines of Emonds' (1994: 162) "principle

of economy of derivation": "[t]he most economic realization of a given deep structure minimizes insertion of free morphemes". The principle is specifically intended to account for the preference for inflectional morphology and simple clitics over full words, e.g., *bigger* rather than *more big, but would seem to be perfectly adaptable to cases of encliticization such as those under discussion.

18. Depending on one's conception both of morphophonology and of the lexicon, this could be seen as an active process of lenition or, alternatively, as selection from a constellation of related shapes in the lexicon. We will take no position on these alternatives here.

19. We take no position here as to whether the verb is inserted in its fully inflected form, or whether the actual inflection occurs later in the derivation.

20. This also addresses a question raised by a reviewer for *Linguistics* concerning the point at which *o* 'of' is inserted in Welsh, in phrases such as (i):

(i) disgrifiad   o'r   dyn/*disgrifiad o y dyn
    description of-the man
    'the description of the man'

In the case of the English counterpart, it is usual to assume that *of* is inserted late in the derivation, as noted at the start of Section 4.1. Clearly, the *'r* form of the article takes precedence over the full form here too, so the issue is whether the *o* has to be already in place before the article is inserted, in order to determine the form of the article, or whether *o* can also be subject to late insertion. If we allow that items inserted simultaneously into the derivation can nonetheless interact to determine the correct allomorph, then the elsewhere principle will still ensure that *'r* is selected in (i), given that its environment is met, and this is the case whether *o* and *'r* are both inserted at the second level or both subject to late insertion. As before, the slot for *o* must be present throughout, since it is a trigger for soft mutation.

References

Carstairs, Andrew (1988). Phonologically conditioned suppletion. *Yearbook of Morphology 1988*, 67–94.

Coates, Richard (1987). On a class of solutions to a phonological dilemma. Cognitive Science Research Reports, University of Sussex.

Doherty, Berlie (1993). *Annwyl Neb* [Dear nobody], Emilie Huws (trans.). Llandysul: Gomer.

Emonds, Joseph E. (1994). Two principles of economy. In *Paths towards Universal Grammar*, Guglielmo Cinque, Jan Koster, Jean-Yves Pollock, Luigi Rizzi, and Raffaella Zanuttini (eds.), 155–172. Washington, DC: Georgetown University Press.

—(2000). *Lexicon and Grammar: The English Syntacticon*. Berlin and New York: Mouton de Gruyter.

—(2002). A common basis for syntax and morphology: tri-level lexical insertion. In *Many Morphologies*, Paul Boucher (ed.), 235–262. Somerville, MA: Cascadilla Press.

Halle, Morris; and Marantz, Alec (1993). Distributed morphology and the pieces of inflection. In *The View from Building 20*, Kenneth Hale and Samuel J. Keyser (eds.), 111–176. Cambridge, MA: MIT Press.

Hughes, T. Rowland (1943). *O Law i Law*. Llandysul: Gomer.

Inkelas, Sharon; and Zec, Draga (1990). *The Phonology-Syntax Connection*. Chicago, IL: CSLI Publications.

Jackendoff, Ray (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

Kager, René (1999). *Optimality Theory*. Cambridge: Cambridge University Press.

Kager, René; van der Hulst, Harry; and Zonneveld, Wim (eds.) (1999). *The Prosody-Morphology Interface*. Cambridge: Cambridge University Press.

Kiparsky, Paul (1973). 'Elsewhere' in phonology. In *A Festschrift for Morris Halle*, Stephen R. Anderson and Paul Kiparsky (eds.), 93–106. New York: Holt, Rinehart, and Winston.

—(1982). From cyclic phonology to lexical phonology. In *The Structure of Phonological Representations*, Harry van der Hulst and Norval Smith (eds.), 131–175. Dordrecht: Foris.

Klavans, Judith L. (1985). The independence of syntax and phonology in cliticization. *Language* 61, 95–120.

Lapointe, Steven G.; Brentari, Diane K.; and Farrell, Patrick M. (eds.) (1998). *Morphology and its Relation to Phonology and Syntax*. Stanford, CA: CSLI Publications.

Marantz, Alec (1995). 'Cat' as a phrasal idiom: consequences of late insertion in distributed morphology. Unpublished manuscript, MIT.

Morgan, T. J. (1952). *Y Treigladau a'u Cystrawen*. Cardiff: Gwasg Prifysgol Cymru.

Nespor, Marina; and Vogel, Irene (1986). *Prosodic Phonology*. Dordrecht: Foris.

Prince, Alan; and Smolensky, Paul (1993). Optimality theory: constraint interaction in generative grammar. RuCCs Technical Report #2, Rutgers University Center for Cognitive Science, Piscataway, NJ. (ROA 537–0802)

Thomas, Peter Wynn (1996). *Gramadeg y Gymraeg*. Cardiff: Gwasg Prifysgol Cymru.

Thorne, David A. (1993). *A Comprehensive Welsh Grammar / Gramadeg Cymraeg Cynhwysfawr*. Oxford: Blackwell.

Walker, Douglas (2001). *French Sound Structure*. Calgary: University of Calgary Press.

Zwicky, Arnold (1985). Clitics and particles. *Language* 61, 283–305.

—(1987). French prepositions: no peeking. *Phonology Yearbook* 4, 211–227.