

Reconstructing last week's weather: Syntactic reconstruction and Brythonic free relatives

David Willis

Department of Linguistics, University of Cambridge

Lightfoot (2002) argues that syntactic reconstruction is rendered impossible by the lack of any analogue in syntax to the traditional notion of the phonological 'correspondence set' of the Comparative Method and by the radical discontinuity caused by reanalysis between successive grammars. Harris & Campbell, in various works, have defended the notion of 'syntactic pattern' as the analogue of the correspondence set, arguing that patterns can be compared across languages, with innovations being stripped away to reveal aspects of the protolanguage. In this article, I argue that syntactic reconstruction can be carried out while maintaining and indeed utilizing core notions in generative approaches to syntactic change such as the central role of reanalysis and child language acquisition and the distinction between the abstract grammatical system and the surface output of that system. Reanalysis itself is constrained by the fact that both pre- and post-reanalysis grammars must be acquirable on the basis of the same primary linguistic data. This imposes limits on the possible hypotheses that can be entertained ('local directionality') even in the absence of any crosslinguistic generalizations about patterns of change ('universal directionality'). This approach is then applied to aspects of the syntax of free relative clauses and negation in the early Brythonic Celtic languages (Welsh, Breton, Cornish), showing that non-trivial reconstructions can be achieved even where the daughter languages manifest significant differences.

Keywords: Syntactic reconstruction, free relatives, negation, reanalysis, grammaticalization, Brythonic Celtic languages.

1 INTRODUCTION

David Lightfoot has recently argued that the prospects for realist syntactic reconstructions are very limited because of the 'chaotic' nature of syntactic change. Outside of very narrowly defined limits (essentially where daughter languages show identity), he suggests that 'one can no more reconstruct the syntax of a proto-language than one can reconstruct last week's weather, and for the same reason: both reflect chaotic systems' (Lightfoot 2002: 135). Lightfoot makes the important observation that the central mechanism of syntactic change, abductive reanalysis, replaces one grammatical analysis with another, which may differ from its predecessor without limit. Consequently, he argues, the new grammar can tell us nothing about its immediate predecessor.

In this article, I will argue that, while the radical discontinuity of abductive reanalysis is an important aspect of syntactic change and reconstruction, it does not have the disastrous consequences for reconstruction that Lightfoot claims. It is accepted that children cannot compare their grammars with those of the adults around them and that, therefore, there can be no theory of the direct relationship between successive grammars. However, it does not follow from this that a grammar can differ in unlimited ways from its immediate predecessor. An innovative abduced grammar must be as good a fit for the primary linguistic data (PLD) of language acquisition as its predecessor, and this fact sharply constrains what hypotheses the historical linguist can entertain about an earlier grammar. I propose that this fact can be used in

syntactic reconstruction: the earlier reconstructed grammar must have produced an output very similar to that which must have led to the attested systems. Reanalysis proceeds via the availability of acquisitionally ambiguous sentence types and does not alter the output of the grammar. Whether or not a particular change manifests a significant trend towards unidirectionality across different language histories ('universal directionality'), these facts will further limit possible reconstructions across instances of historical reanalysis ('local directionality'). I will demonstrate the usefulness of this approach using data from free relatives in the early Brythonic languages (Middle Welsh, Middle Breton and Middle Cornish).

1.1 Historical background

A series of attempts were made in the 1970s to reconstruct Indo-European word order on the basis of typological generalizations. For instance, Lehmann (1974) showed that early Indo-European languages all had features typologically characteristic of SOV languages (modifier–head orderings such as adjective–noun order, genitive–noun order, standard–pivot–comparative order etc.). Tracing a development away from these features over the attested histories of the early languages, he reconstructed Proto-Indo-European as a typologically consistent SOV language. Other researchers (Friedrich 1975, Miller 1975) reached radically different conclusions, but used similar methods. These approaches, which assumed the protolanguage had typologically consistent syntax, were widely criticized at the time. Watkins (1976), for instance, criticized them for reducing syntax essentially to word order, and for then reducing word-order reconstruction to asserting implicational consistency of head–modifier order in a small number of head–modifier pairs in the protolanguage.

The legacy of the 1970s typological work has tarred all research in syntactic reconstruction with the same brush. However, more recent work is radically different in approach, and has focused not on broad-brush typological generalizations about languages, but rather on the development of the syntax of individual functional items and narrowly defined syntactic patterns, witness for instance the fine-grained reconstructions of various syntactic patterns across Finno-Ugric in Campbell (1990) or internal reconstruction of the syntactic origins of various verbal forms in Swahili in Givón (1999). It is this second approach that I pursue in this article. I argue that, even accepting the validity of several of Lightfoot's basic premises (namely the radical discontinuity in abductive reanalysis and the need to reconstruct abstract grammatical systems), progress can still be made. Successful reconstruction can be achieved by investigating microvariation in closely related varieties. Reanalyses are not completely unconstrained: a successful reanalysis manifested acquisitional ambiguity, the possibility of two different structural analyses for a particular subset of relevant sentences at the point of transition, and this fact can be used to 'reverse' reanalysis without any appeal to universal directionality of change. Directionality can be assessed at a purely local level. This approach also has the advantage that it does not require us to give up the internalized abstract grammar as the object of investigation: we investigate both surface patterns (E-language) and the abstract grammatical systems that generate them (I-language), reconstructing the diachronic interplay between the two.

This article is structured as follows. Traditional problems associated with syntactic reconstruction are outlined in section 2. Section 3 addresses possible solutions to these problems and methods for engaging in syntactic reconstruction, outlining the approach adopted here which focuses on reconstructing fragments of grammars and reanalyses rather than surface outputs alone. Section 4 applies this method to a data set consisting of free relatives in the three medieval Brythonic languages, Welsh, Breton and Cornish, leading to a reconstruction of the relevant area of the Common Brythonic grammar.

2 PROBLEMS

The problems in applying the Comparative Method to syntax are well-known and have been used to deny the possibility of non-trivial syntactic reconstruction in its entirety. This section sets out the four main problems, dubbed here the correspondence set problem, the directionality problem, the radical reanalysis problem and the transfer problem.

Phonological reconstruction relies on establishing correspondence sets: lexical items containing the same sound in a particular phonological environment in a set of languages hypothesized to be related. This is illustrated, in simplified form, adapted from Campbell (1998), for two correspondence sets in Romance in Table 1. We find a set of cognate lexical items containing a /k/ before a back vowel in correspondence set (1) in all the modern Romance languages listed. In correspondence set (2), before a front vowel, we find /k/ in Italian, Spanish and Portuguese, but /ʃ/ in French.

	ITALIAN	SPANISH	PORTUGUESE	FRENCH	
(1)	k	k	k	k	
	correre	correr	correr	courir	‘run’
	costare	costar	costar	coûter	‘cost’
(2)	k	k	k	ʃ	
	caro	caro	caro	cher	‘dear’
	capo	cabo	cabo	chef	‘head, top’

Table 1. Simplified correspondence sets for Proto-Romance */k/.

We assume that any sound change that gave rise to this variation across Romance applied to all lexical items in a given phonological environment (in this case, after a front vowel) (Regularity Hypothesis) (Labov 1981, Osthoff & Brugmann 1878). We hypothesize a sound change, namely /k/ > /ʃ/ before a front vowel in French, and therefore reconstruct each of the items in correspondence set (2) as originally containing an initial /k/. In establishing the historical development, we appeal to two factors: (a) economy: reconstruct a history with as few sound changes as possible; and (b) universal directionality: /k/ > /ʃ/ before a front vowel is a commonly attested and, above all, physiologically motivated change and hence ‘natural’, while /ʃ/ > /k/ before a front vowel is not. These considerations are crucial to the method since, otherwise, we would not be able to determine whether French or the other Romance languages best reflected the ancestral situation. The method would be able to demonstrate relatedness but would not reconstruct the shape of the given lexical items in the parent language.

This procedure cannot be straightforwardly applied to syntax. First, it is hard to know what the analogue of the correspondence set is in syntax (the ‘**correspondence set problem**’). The obvious answer is that it is collections of ‘cognate’ sentences containing a particular feature rather than sets of lexical items. However, sentences are not transmitted as whole units from generation to generation; rather an entire grammatical system is transmitted, and, if a particular sentence survives as a possible sentence of a language from one generation to the next, this is because the relevant aspect of the grammar has been transmitted intact, not because the sentence itself has. In this respect, sentences are different from lexical items, which are transmitted directly from generation to generation (although not phonological systems such as Optimality Theory constraint hierarchies). It has therefore been concluded that there are no correspondence sets in syntax (Lightfoot 2002: 120–1).

Secondly, while research on grammaticalization appears to offer an analogue to universal directionality in sound change, this too has been challenged: Lightfoot, in particular, argues that grammaticalization changes have purely local motivation and doubts the relevance of any general theory of grammaticalization to explain them. From this, he concludes that ‘we have no well-founded basis for claiming that languages or grammars change in one direction but not in another, no basis for postulating algorithms mapping one kind of grammar into another kind’ (Lightfoot 2002: 125–6). The result is that, even when presented with a series of alternative grammatical systems, it is hard or impossible to know which one to posit for the parent language (the ‘**directionality problem**’).

Related to the directionality problem is another problem related to the central mechanisms of change. On many interpretations, both generative and traditional, syntactic change is often mediated by radical reanalyses of structure, in which children abduce an entirely new grammatical structure for a particular construction, organizing other aspects of the language system (e.g. lexicon) into line with the analysis they adopt. Reconstruction amounts to working out what structure the language had before the reanalysis, and there is no trace of that structure in the new language: if children had been able to work out what the earlier structure was, they would not have introduced the new structure. Radical reanalysis breaks a language’s relationship to its past (in some particular area of the grammar), obscuring its genetic connections (the ‘**radical reanalysis problem**’). This seems to present formidable obstacles to reconstruction, although the suggestion that reanalysis leaves no trace of the former structure is not entirely true (see section 3.2.1 below).

Discussion of a fourth problem, the ‘**transfer problem**’ will be delayed until further discussion of correspondence sets in syntactic reconstruction in section 3.

3 SOLUTIONS

3.1 Patterns as elements of a correspondence set

In various works (Campbell 1990, 2003, Campbell & Harris 2002, Campbell & Mithun 1980, Harris 1985, Harris & Campbell 1995), Alice Harris and Lyle Campbell have defended the possibility of non-trivial syntactic reconstruction. They suggest that ‘syntactic patterns’ can usefully be treated as the elements of the correspondence set. That is, they claim that the relevant type : token relationship in syntax is between syntactic patterns (types) and sentences (tokens), just as the relevant relationship in phonology is between the phoneme in a given phonological environment (e.g. /t/ in intervocalic position) and lexical items (e.g. *butter*, *better*, *eating*).

For instance, from observation of actual sentence patterns, Harris (1985: 37–58) abstracts the patterns of case marking for various Kartvelian languages in Table 2, where ‘Series I’ and ‘Series II’ refer to particular verbal categories (tense–mood–aspect combinations).

Laz	subject of transitive	subject of unergative	subject of unaccusative	direct object
Series I	NAR	NAR	NOM	NOM
Series II	NAR	NAR	NOM	NOM
Mingrelian	subject of transitive	subject of unergative	subject of unaccusative	direct object
Series I	NOM	NOM	NOM	DAT
Series II	NAR	NAR	NAR	NOM
Svan/Old Georgian	subject of transitive	subject of unergative	subject of unaccusative	direct object
Series I	NOM	NOM	NOM	DAT
Series II	NAR	NAR	NOM	NOM

Table 2. Case marking patterns in various Kartvelian languages.

Laz has an active-ergative case-marking system (distinguishing agentive subjects from non-agentive, thematic subjects and objects), with identical patterning in both series. Mingrelian has accusative case marking in both series (distinguishing subjects from objects), but uses different morphological cases to express that system. Svan and Old Georgian have an accusative case-marking system in Series I, but an active-ergative one in Series II. The languages can be compared within this subsystem, and we can ask what type of case system the parent language could have had that would have allowed both to emerge. For instance, Harris proposes that, for Series I, the system common to Mingrelian, Svan and Old Georgian is the ancestral one, with Laz having generalized nominative case marking from direct objects to subjects of unaccusatives by analogy with the active-ergative system that it uses in Series II. Notice that there is no appeal here to (universal) directionality: the claim that Laz has been innovative in Series I is not made on the basis of any claim that accusative systems tend to turn into active-ergative systems, but rather on the basis of very local factors, namely the existence of a possible extension of the pattern found in Laz in Series II (local directionality).

This procedure produces plausible reconstructions in general, and the particular claim about the innovative nature of the Laz system above seems entirely convincing. However, it is hard to see that syntactic patterns really are correspondence sets in syntax. The idea that change is regular (in the sense of the Neogrammarian Hypothesis) within a syntactic pattern is odd. Sound change is regular because it affects every member of the relevant correspondence set, but syntactic change does not affect every sentence that uses a particular pattern: sentences, unlike lexical items, are not learned by each new generation; only the grammatical rule is.

Furthermore, this understanding of correspondence sets endangers the two crucial functions of correspondence sets in sound change, namely to eliminate chance resemblance and to protect against and identify language contact. To begin with, syntactic reconstruction is normally carried out among languages whose genetic relatedness is already known. While some attempts have been made to demonstrate relatedness and perform subgrouping using syntactic parameters (Guardiano & Longobardi 2005, Longobardi & Guardiano 2009: 368–75, Roberts 2007), this is not the usual situation. Hence the possibility that a syntactic pattern is shared because of chance or for typological reasons is not a barrier to demonstrating relatedness, which is done by other means in any case. Nevertheless, where the number of possible patterns is limited, two languages may share the same general pattern by chance (that is, both have innovated the same pattern independently). This danger can be reduced if the two

languages share details of the patterns (e.g. exceptional language-specific rules) or if the crucial functional items in the relevant patterns are shared (cognate).

Correspondence sets in phonological reconstruction are also used to identify and eliminate lexical borrowings: a loaned item will only participate in sound changes that got underway after it was introduced into the language, and will therefore show irregular correspondences in cases where the expected correspondences are due to sound changes that took place before the item in question was borrowed. However, syntactic crosslinguistic comparisons will often contain only one item (pattern) in each language. Treating patterns as elements of a correspondence set in syntax means that we cannot eliminate the possibility that a particular syntactic pattern was transferred from one language to another (the ‘**transfer problem**’). We therefore need to be aware of the need to identify transfer by other means. Even if we suspect transfer, syntactic comparisons will provide no particular clue to the direction of transfer, which will have to be determined on external grounds.

Harris & Campbell’s approach creates a kind of vague analogue between syntax and phonology: in phonology, we use the lexical item as basis for forming correspondence sets, and not the phoneme alone, for the simple reason that phonemes are not transmitted from generation to generation via child language acquisition; lexical items are. The phoneme is already one level of abstraction away from the observed data. In syntax, Harris & Campbell abstract patterns from sentences in the same way. This procedure makes sense in a theory of grammar in which syntactic patterns are stored as units of the grammar (e.g. construction grammar), and Harris & Campbell seem to be assuming this type of model of grammar. Note also that phonological reconstruction also operates at a fairly limited level of abstraction: the deepest abstraction normally used is the phoneme in a particular environment in the daughter languages. The procedure does not reconstruct the phonemic status of an item in the parent language, only a speech sound which would have to be subjected to further analysis to establish its phonemic status within the phonological system of that language. Furthermore, historical linguists undertaking phonological reconstruction do not normally attempt to reconstruct a phonological grammar (e.g. within Lexical Phonology or Optimality Theory), but rather are largely concerned with establishing surface forms of the protolanguage. Harris & Campbell’s relatively surface-oriented approach is therefore not so different from standard practice in phonological reconstruction.

This procedure is, in fact, comparing grammatical systems, not correspondence sets, and we should be honest about this. Comparing grammatical systems may nevertheless be a useful enterprise, and this is a procedure that will be developed in the next section. While I am not unsympathetic to Harris & Campbell’s general approach, I shall argue that it is not necessary to abandon central tenets of generative grammar (namely the existence of an abstract grammatical system that projects syntactic structure from individual lexical items) to pursue syntactic reconstruction fruitfully.

3.2 Techniques for reconstructing from correspondence sets

3.2.1 Reconstruction of abstract grammatical systems

Syntactic reconstruction involves both syntactic patterns and abstract mental grammars that produce those patterns as their output. A given syntactic pattern may be produced by more than one possible grammar, so the relationship between the two is not trivial. At all stages of the enterprise, we must ask what grammatical system generated the outputs that we observe or postulate. As a methodology, we can observe patterns in the daughter languages, analyse those patterns as grammars, and ask what earlier outputs could have led ultimately to the historical

emergence of those grammars (via reanalysis and actualization/extension). From these reconstructed outputs we infer a reconstructed abstract grammatical system. This basic approach follows Pires & Thomason (2008) in requiring the process of syntactic reconstruction to involve the following three elements:

- (i) hypotheses about the grammars that generate the outputs of the daughter languages;
- (ii) hypotheses about the grammar internalized by speakers of the protolanguage;
- (iii) hypotheses about how the different grammars of the daughter languages could have developed from exposure to the output of the proposed protogrammar (Pires & Thomason 2008: 45)

The model of change assumed here follows a standard model of reanalysis and actualization / extension (Andersen 1973, Harris & Campbell 1995, Timberlake 1977). The basic schema is given in Figure 1. The language state that we are aiming to reconstruct is grammar A, on the basis of which output A was produced. The daughter languages may either replicate this situation successfully for a given aspect of syntax or may undergo reanalysis and actualization / extension. Reanalysis, by definition, does not change the patterns available; that is, by definition, output A and output B in Figure 1 are identical. A change in the syntactic patterns found occurs subsequently, as the new grammatical system is more open to certain types of innovations (generalizations / syntactic analogies) than the old one was. Observed innovations, then, occur only in output C.

Reanalysis requires acquisitional ambiguity: some subset of the examples of a particular construction must be amenable to two possible analyses by children acquiring the language (Timberlake 1977, Willis 1998: 41). That is, output A must have been amenable in principle to two different analyses, namely grammar A and grammar B. If we can establish grammar B, then we can ask what instances of acquisitional ambiguity there may have been, thereby limiting the possible forms that grammar A could have taken.

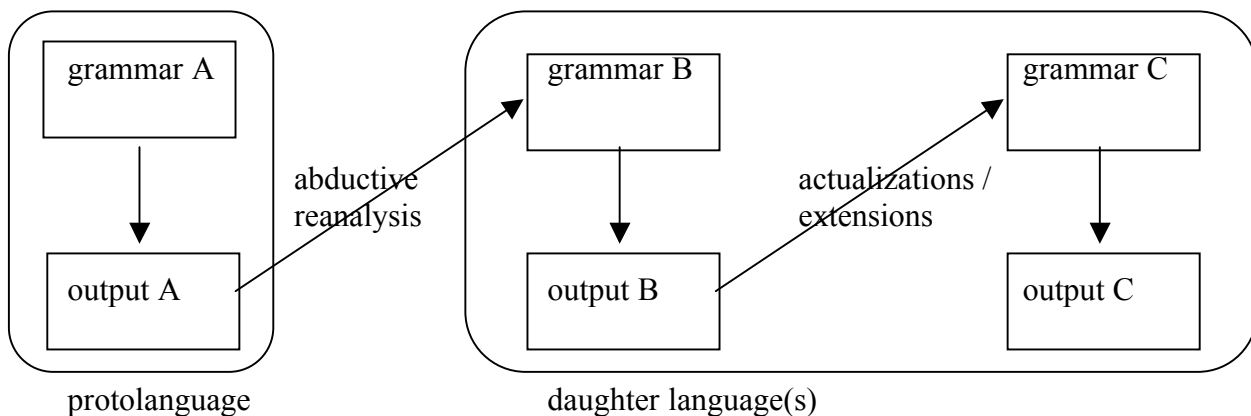


Figure 1. Model of syntactic change for reconstruction.

These two observations (reanalysis does not change the output and reanalysis needs acquisitional ambiguity) guide our methodology in carrying out syntactic reconstructions and impose restrictions on possible hypotheses. While there is a single grammar A and output A, this representing the protolanguage to be reconstructed, any given daughter language may successfully replicate this system or else undergo changes of the type given in Figure 1, leading to the emergence of output C. Different daughter languages may undergo different changes, hence we can have multiple versions of output C.

For any given daughter language, we have direct access to output C, which we can analyse to establish grammar C. Output A/B may be present in some other language in the group, or it may not (having disappeared in all languages). If it is present, then comparison of output A/B and output C reveals that C involves generalization from B, reflected in a change in the grammatical system between B and C. However, synchronically grammar B seems less well motivated than grammar C. This leads us to posit the earlier existence of grammar A, a different system for arriving at output A/B, one where output C is not a natural generalization. This is the basic approach that will be adopted in the reconstruction of Brythonic free relatives shortly. This will be supplemented by local directionality, to which we now turn (section 3.2.2), along with a series of tradition criteria (economy, archaisms, extension) (section 3.2.3).

3.2.2 *Use of directionality and ‘possible change’ in syntactic reconstruction*

Research in grammaticalization has apparently made syntactic reconstruction easier by allowing us to turn back the clock on grammaticalization paths. Given a language with a lexical verb ‘want’ and another with a similar-looking future auxiliary, grammaticalization tells us that the emergence of future markers from verbs of volition is a unidirectional pathway, and hence, that we can confidently reconstruct the lexical verb rather than the auxiliary for the parent language (Bybee & Pagliuca 1987, Bybee, Pagliuca & Perkins 1991, Bybee, Perkins & Pagliuca 1994). Such research supplies us with general pathways of change. I shall refer to this criterion as **universal directionality**: a change must have proceeded in a particular direction because it is a general property of language change that it proceeds only in this direction and not the reverse.

Grammaticalization is, however, not absolutely unidirectional (Campbell 2001, Newmeyer 2001, Norde 2009), and this may pose problems; see, for instance, Willis (2010) for an example of the potential problems posed for reconstruction by counterdirectional changes in the Slavonic conditional. However, recognition that grammaticalization is not unidirectional does not preclude us from using statistical tendencies about directionality or particular instances of change that do seem to be unidirectional. Universal directionality, then, is a useful tool in syntactic reconstruction, but not an infallible one. The same in fact applies in phonology: universal pathways of sound change vary according to how unidirectional they are, but we must always keep open the possibility that a particular constellation of circumstances led to an unexpected direction of change in a particular instance.

Lightfoot argues that ‘a distinction between possible and impossible changes is in principle a necessary prerequisite for reconstruction’ (Lightfoot 1979: 154), and concludes that syntactic reconstruction is impossible because there is no such distinction in syntax and therefore no directionality: ‘in syntax there are no formal constraints on possible changes independent of those which follow from a definition of a possible grammar’ (Lightfoot 1979: 155).

This approach confuses two things. On the one hand, Lightfoot is correct to argue that there can be no constraints imposed by the acquisition algorithm on possible changes. Children cannot limit the extent or nature of the difference between their grammar and that of the adults around them, because, if they could, that would mean they knew the exact form of the adult grammar, and, if children knew the exact form of the adult grammar, change would not arise in the first place. Hence any possible grammar can change into any other possible grammar. However, change is constrained in another way: both the adult grammar and children’s hypotheses about their language are constrained by the fact that they must be consistent with the Primary Linguistic Data upon which they were based. In reversing a reanalysis we are therefore looking for two analyses, both of which account for the data, one of which was adopted by children in preference to an earlier one used by the adults from which the acquired their language. This fact imposes quite narrow limitations on transitions, essentially requiring

us to state the nature of the acquisitional ambiguity at each point of reanalysis in our reconstruction. The upshot of this is that, at the macrolevel, Lightfoot's claims may well be true, at least for a significant amount of the domain we are investigating: an SVO language may become a V2 language or a V2 language may become an SVO language, or, to put it another way, the value of the V2 parameter may shift in apparently random fashion. However, at the microlevel, things are much more constrained.

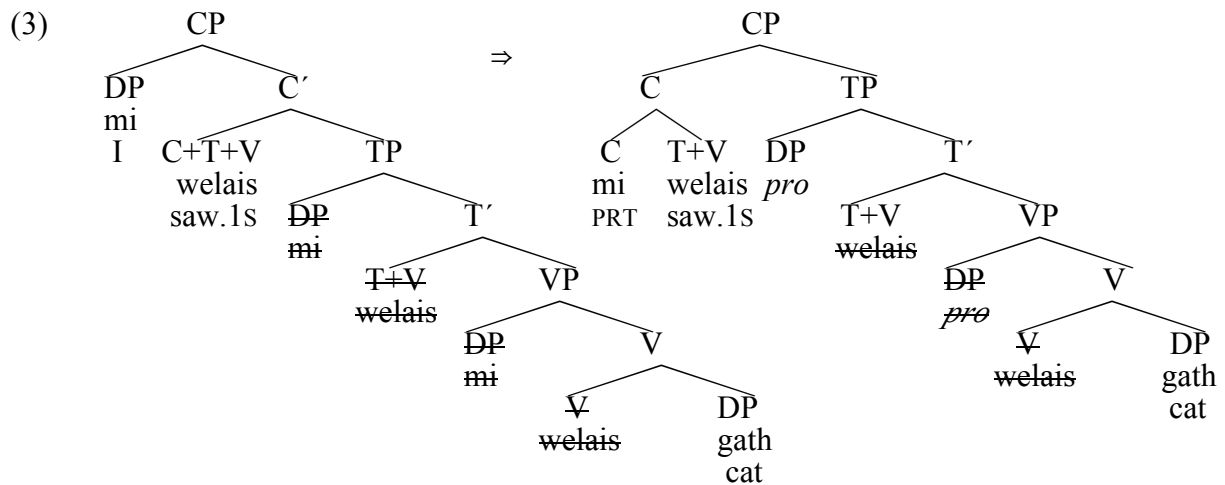
To give a concrete (and simplified) example, consider Brythonic word order. In main clauses, Modern Breton is a V2 language (Borsley & Kathol 2000), while Modern Welsh is VSO. On the basis of the modern languages, should we reconstruct Common Brythonic as V2 or as VSO in main clauses? At the macrolevel of argumentation, we have no basis for deciding. We have no evidence that the historical relationship between the two word order types manifests any kind of universal directionality, so either is a possible reconstruction. However, at the microlevel, we ask how Welsh could have undergone a transition from V2 to VSO and how Breton could have undergone a transition from VSO to V2: is there local directionality? Closer examination of Welsh phrase structure reveals a means by which Welsh could have developed VSO word from earlier V2. Modern Welsh allows initial verbs in main clauses to be preceded by *mi* or *fe*, particles that function as affirmative complementizers. These are identical in form to pronouns (first person singular *mi* and third person masculine singular *fe*), but clearly do not function as such. In (1), the verb is third person singular and the subject is *Steffan*, not *mi*.

- (1) Mi welodd Steffan gath.
 PRT see.PAST.3S Steffan cat
 'Steffan saw a cat.'

(Present-day Welsh)

These have no analogue in Breton. If Welsh main clauses were formerly verb-second, then the availability of sentences such as (1) can be understood as the result of a change in which a preverbal subject pronoun was reanalysed as an affirmative complementizer (Willis 1998):

- (2) [CP [Spec *mi*] [C *welais*] [TP ~~*mi*~~ [T ~~*welais*~~] [VP ~~*welais*~~ *ferch*]]]]]
 ⇒ [CP [Spec \emptyset] [C *mi+welais*] [TP *pro* [T *welais*] [VP ~~*welais*~~ *ferch*]]]]]



We can therefore reconstruct a V2 system for the parent language, a conclusion borne out by the actual historical records of the languages in question. We do not need to assume that pronoun > complementizer is a universal unidirectional path of development to conclude this.

Even if we permitted the reverse reanalysis to have occurred in Breton (reconstructing the particle *mi* in the parent language), it would not explain the creation of a generalized system of V2 in Breton. This approach then agrees with the observation that we need to ‘tell a plausible story about how grammatical objects in different languages developed from a single antecedent grammatical object’ (Harrison 2003: 225), elements of which are not guided by a narrow interpretation of the Comparative Method.

To conclude, it is consistent to accept that differences between successive grammars are not subject to any universal diachrony-specific constraints and to believe that the unidirectionality of grammaticalization is not exceptionless (a belief which in fact follows from the first claim), while at the same time believing that a local form of directionality has a role to play in reconstruction. For any given reconstruction problem, we can evaluate the relative likelihood or plausibility that the development took place in one direction rather than the other. This includes an assessment of whether factors known to play a role in acquisition, such as simplicity or perhaps markedness, cf. Roberts & Roussou (2003) and van Gelderen (2004), suggest that a particular direction of change is more likely than another.

3.2.3 *Other techniques*

Within such an approach, some familiar techniques of reconstruction play their usual role. The first of these is economy. As in phonological reconstruction, all other things being equal, we posit the reconstructed form which requires the smallest number of innovations in the daughter languages. Any reconstruction which requires the same development to occur in more than one daughter language or in more than one branch of the daughter languages (independent parallel development) is immediately suspect. A related concept is ‘majority rules’: the option found in the largest number of languages is reconstructed. However, this can only be of use where the subgrouping is uncertain. Once subgrouping is known, the more sophisticated concept of economy must be brought into play. This takes into consideration the fact that an option found in two distantly related parts of a family is more likely to be inherited and that a single innovation may lead to an option appearing in many daughter languages if one part of the family tree is more densely packed with languages than other parts.

In line with Meillet’s observation that we reconstruct on exceptions, not on rules (Meillet 1931), identification of archaisms allows us to evaluate hypotheses about the ancestral grammar against one another. In this context, archaisms are features (exceptional grammatical rules) that are motivated (generated by the regular pattern) only within an earlier grammar. Exceptions can often be explained as having once been motivated and therefore having once been part of a regular and therefore more widely applied grammatical rule. On the other hand, it is generally difficult to construct a convincing account for how an isolated exception could arise from nothing. Campbell (1990: 82–6) cites the example of Finnish infinitives. Finnish normally uses the ‘third infinitive’ after a verb of motion (‘Father went to cut hay’), but when the infinitive is ‘lie down’, we unexpectedly find the ‘first infinitive’ being used. This can be explained as a relic: the third infinitive is an innovation and replaced the first infinitive after verbs of motion, but some fixed environments retain the earlier pattern.

Finally, analogical extension of a pattern in one language to contexts where it is not found in another language must be eliminated. Finnish uses what looks like the negative of the perfect to express the negative of the preterite, while some other Balto-Finnic-Lapp languages something that looks straightforwardly like the negative of the preterite (Campbell 1990: 70–4). Finnish has extended use of the negative of the perfect (in a form that had become obsolete with the actual perfect) to the preterite. The problem here though is identifying that this is a case of analogical extension of a pattern, rather than an archaism, which would lead to the opposite reconstruction.

4 THE RECONSTRUCTION OF BRYTHONIC FREE RELATIVES

In the second half of this article, this general approach to syntactic reconstruction is applied to a problem concerning the reconstruction of patterns for free relatives in Brythonic. While all modern and medieval Brythonic languages have a free-relative marker cognate with Welsh *bynnag*, corresponding more or less to *-ever* in English *whoever* etc., the details of the syntactic rules for its use differ significantly from language to language. I will argue that the patterns of the daughter languages only make sense if a reanalysis is reconstructed for late Brythonic, and that both the details of that reanalysis (that is, both the ancestral grammar and the grammar resulting from that reanalysis) can be reconstructed along with a series of extensions that led to a better motivated system in some of the daughter languages. In doing so, I follow the basic model outline in section 3.2.1 above, while also making reference to other tools, such as identification of archaisms, plausible patterns of extension and identification of language contact, as discussed in sections 3.2.2 and 3.2.3.

4.1 The Brythonic subfamily

The Brythonic subfamily of Celtic consists of three languages with extensive attestation in the medieval period, namely Middle Welsh, Middle Breton and Middle Cornish. All are descended from a Common Brythonic language, itself unattested except via names in Latin sources and in one possible inscription discovered in Bath (Tomlin 1987). While once spoken in a single dialect continuum across the whole of England, Wales and southern Scotland, most varieties of the language were replaced by English and Scots from the fifth to the eighth centuries. A western variety gave rise to Welsh, while in the southwest, Cornish emerged, with migrants from and via this region creating a new language, Breton, in Brittany. In the northwest, another Brythonic language, Cumbric, survived into the Middle Ages, but remnants of it are too poor to be of any use for reconstruction. These relations are traditionally summarized in the family tree given in Figure 2. It is worth bearing in mind, however, that these languages originally formed a dialect continuum and essentially remained in this relationship (via sea routes) for a long time after they became physically separated from one another by land. Cumbric was most likely closer to Welsh than to the other varieties, and could be considered to form a western subgrouping with Welsh. Again, this makes little difference for current purposes.

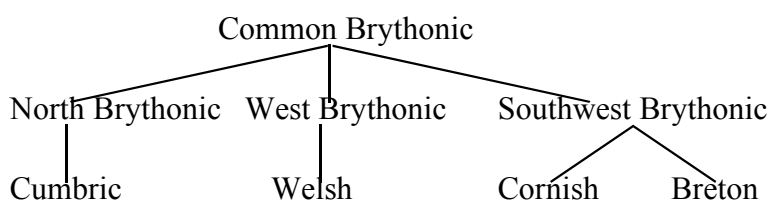


Figure 2. Subgrouping for the Brythonic branch of Celtic.

4.2 The data

Brythonic languages have a cognate set comprising Welsh *bynnag*, Breton *bennak*, Cornish *penag*, used to form free relatives.¹ All the languages allow this item to follow a *wh*-word, giving Welsh *pw y bynnag*, Breton *piv bennak* and Cornish *pyw penag* ‘whoever’:

- (4) **pw y bynnac** a 'm metrei i yuelly
 who ever REL 1S.ACC strike.IMP.F.SUBJ.3S me thus
 ‘whoever could strike me thus’ (*Pedeir Keinc y Mabinogi* 87.2) (Middle Welsh)
- (5) **Piou pennac** no cred nendeu guir seruicher da doue
 who ever NEG+3P believe.PRES.3SNEG+be.PRES.3S true servant to God
 ‘Whoever does not believe them is not a true servant of God’
 (*Middle-Breton Hours* 4) (Middle Breton)
- (6) **pyw penagh** a 'm gwellha vy
 who ever REL 1S.ACC see.PRES.SUBJ.3S me
 ‘whoever may see me’ (*Resurrexio Domini Nostri Jhesu Christi* 2384)
 (Middle Cornish)

Wh-phrases other than ‘who’ may precede:²

- (7) Canys **pa beth bynnac** ry gollych ti eno, minheu
 for which thing ever PERF lose.PRES.SUBJ.2S you there I
 a’e henillaf yty yma.
 PRT+3S.ACC win.PRES.1S to.you here
 ‘For whatever you might lose there, I shall win it [back] for you here.’
 (*Brut Dingestow* p. 77.1–2) (Middle Welsh)
- (8) A **pha borthua bynnac** y delhei y bydei
 and which harbour ever REL come.IMP.F.SUBJ.3S PRT be.COND.3S
 Edwin a llu ganthav yn y ludyas.
 Edwin and force with.3MS PROG 3MS.GEN impede.INF
 ‘And whichever harbour he came to, Edwin would be there with a force with him
 impeding him.’ (*Brut Dingestow* p. 195.21–2) (Middle Welsh)

¹ The following descriptive outline of the data for the medieval Brythonic languages is based on published grammars and exhaustive searching of a range of medieval texts. The grammars used are Evans (1964) for Middle Welsh, Hemon (1975) for Middle Breton and Lewis (1946) for Middle Cornish. The texts used are: for Middle Welsh, *Brut Dingestow*, *Pedeir Keinc y Mabinogi*, *Peredur* and *Ystoryaeu Seint Greal*; for Middle Breton, *Le breton de Gilles de Keranpuil*, *Le dialogue entre Arthur et Guinclaff*, *Doctrin an Christenien*, *Le mystère de sainte Barbe*, *La vie de Sainte Catherine*, and extracts from *Le Mirouer de la Mort* (lines 1–1200); for Middle Cornish, *Origo Mundi*, *Passio Christi*, *Life of Saint Meriasek*, *Bewnans Ke* and *Gwreans an Bys*. While the relevant items are attested in Old Welsh and Old Breton sources (Falileyev 2000: 131, Fleuriot 1964: 269), the evidence is too sparse to add anything significant to the picture given by the much more voluminous Middle Welsh and Middle Breton sources.

² In the Middle Welsh example in (8), the base form is *pa borthua bynnac*, with the shift from *pa* /pa/ to *pha* /fa/ (aspirate mutation) being triggered by the presence of the co-ordinator *a* ‘and’.

- (9) **pé quen bras pennac** vé an offancz
 how big ever be.PRES.SUBJ.3S the offence
 ‘however big the offence may be’
(Le breton de Gilles de Keranpuil 240.1–2) (Middle Breton)
- (10) **py le penag** y’s kyffyn
 what place ever REL+3S.ACC find.IMP.1S
 ‘Wherever a place I find it’ *(Passio Christi 1551) (Middle Cornish)*

All Celtic languages once had both a strong (independent) *wh*-element and a weak (dependent) one (Lewis & Pedersen 1937: 226–9). The strong one survives as Middle Welsh *pyw*, Middle Breton *piu*, Middle Cornish *pyw* ‘who’. The weak one is largely obsolete on its own in Middle Welsh, but there are a few examples where it is used alone to mean ‘what’ or ‘why’:

- (11) **Pa** derw ytti?
 what happen.PERF.3S to.you
 ‘What has happened to you?’ *(Pedeir Keinc y Mabinogi 67.21) (Middle Welsh)*
- (12) Ha uab, **py** liuy ti?
 VOC son why blush.PRES.2S you
 ‘Son, why are you blushing?’ *(Culhwch ac Olwen 54) (Middle Welsh)*

Otherwise it survives only in a range of phrases: *pa* X ‘which X’, *paham* ‘why’ (< *pa* ‘what’ + *am* ‘for’), *pyr* ‘why’ (< *pa* ‘what’ + *yr* ‘for’), *padiw* ‘to whom’ (< *pa* ‘who’ + *di* ‘to’), *pieu* ‘whose’ (< *pi* ‘to whom’ + *eu* ‘is’), *pw y enw* (< *py* ‘what’ + *yw* ‘is’ + *enw* ‘name’) ‘what is the name of’ etc.

The weak *wh*-element is not found before *bynnag* in Middle Welsh or in Middle Breton, but it does occur in this environment in Middle Cornish, giving *pepenag*, which may mean ‘whoever’, ‘whatever’ or ‘wherever’:

- (13) **pepenag** vo a ’n Barth wyr
 who-ever be.PRES.SUBJ.3S of the part true
 ‘whoever is of the true part’ *(Passio Christi 2025) (Middle Cornish)*
- (14) **pe-penag** vo
 what-ever be.PRES.SUBJ.3S
 ‘[It is necessary to follow his will,] whatever it may be.’
(Origo Mundi 662) (Middle Cornish)

Cornish is also alone in allowing *penag* to occur without any preceding element at all to mean either ‘whoever’ or ‘whatever’:

- (15) **penag** a worthya ken du
 whoever REL worship.PRES.3S other god
 ‘whoever worships another god’ *(Life of Saint Meriasek 764) (Middle Cornish)*
- (16) **pynak** vo lettrys py lek
 whoever be.PRES.SUBJ.3S lettered or lay
 ‘whoever he may be, lettered or lay’ *(Passio Christi 681) (Middle Cornish)*
- (17) **panak** vo age deses
 whatever be.PRES.SUBJ.3S their disease
 ‘whatever their disease may be’ *(Life of Saint Meriasek 3104) (Middle Cornish)*

Both Cornish and Breton allow some reinforcing element to follow. In Cornish this is *-ol* ‘all’:

- (18) **penagel** nath car
 who-ever-all NEG+2S love.PRES.3S
 ‘whoever does not love you’ (Bewnans Ke 1792) (Middle Cornish)

In Breton, the suffix *-et* is added, but this is not attested until after the Middle Breton period (Hemon 1975: 144, 1976–: 222).

Middle Breton alone combines *bennak* with the indefinite article *un* to form indefinite noun phrases, such as *un lech pennac* ‘some place’:

- (19) en **un lech pennac** ... /ez crethenn ef-fe em tennet
 in a place some PRT believe.COND.1S PRT+be.COND.3S REFL pull.PP
 de hem cuzet
 to+3SF.GEN REFL hide.INF
 ‘I should think that she has slunk away **somewhere** to hide’
 (*Le mystère de sainte Barbe* 364) (Middle Breton)
- (20) Ret eu diff gouzout... /Diouz **un re pennac**... /Vn tra...
 necessary be.PRES.3S to.me know.INF from someone a thing
 ‘I must learn a thing from **somebody**.’
 (*Le mystère de Sainte Barbe* 107) (Middle Breton)
- (21) Rac na couezhemp en vn **fæczon pennac** dindan an temptation
 lest NEG fall.COND.1P in a way ever into the temptation
 ‘lest we should **in some way** fall into temptation’
 (*Le breton de Gilles de Keranpuil* 240.8) (Middle Breton)
- (22) A palamour ... ma-z dleont vn **guez bennac** bezaff reuniet ouz
 because COMP must.3P one time some be.INF reunite.PP to
 ho eneffou / gloriu
 3P.GEN souls glorious
 ‘because ... they must **at some time** be reunited with their glorious souls’
 (*Doctrin an Christenien* §24) (Middle Breton)

In Modern Breton, an entire series of indefinite pronouns is based on this pattern:

- (23) *un...bennak*-series
 person *unan bennak* / *un den bennak* ‘someone’ (plur. *ur re bennak*)
 thing *un dra bennak* ‘something’
 quantity *un...bennak* ‘some’
 place *ul lec’h bennak* / *un tu bennak* ‘somewhere’

Breton *bennak* is stressed on the final syllable, which is unexpected given the usual penultimate stress rule of the language. Welsh *bynnag* receives regular penultimate stress.

4.3 The reconstruction

The etymology of *bynnag* is understood in its general outline. It is clear that two elements are involved in the emergence of *bynnag*: *pa*, the weak, unstressed form of the general

interrogative pronoun ‘who, what’, plus a negative element *na(g)* (Lewis & Pedersen 1937: 231).³ But what grammar exactly should we posit for the parent language in this area?

4.3.1 Preliminaries to reconstruction

First, we can strip away the Cornish suffix *-ol*. This is clearly an innovation to the form of the lexical item. We reach this conclusion because (a) it is only found in Cornish, which occupies a middle position in the dialect continuum, hence it is more economical to posit that it is an innovation (economy); (b) some instances of Cornish *-ol* violate the metre of their texts, suggesting the texts were originally composed without *-ol*; and (c) universal-type elements are frequently added to free relatives, compare also their use in free-choice indefinite pronouns, where ‘always, ever’ is frequently added (Haspelmath 1997: 137–8) (universal directionality). Similarly, the historical record shows that Breton *-et* is a recent innovation, so this can be stripped away easily too.

Next, we can strip away the Breton *un ... bennak* construction. This is reasonably well attested in Middle Breton, so we cannot dismiss it out of hand. However, we need a story for how it arose, whether in the parent language or later. *Un ... bennak* can be treated as an innovation specific to Breton in which Welsh and Cornish have not participated, because: (a) the development of a free relative (‘Bring whoever you like’) to a free-choice indefinite (‘Bring whoever’ with omission of the relative clause) to a non-specific pronoun to a specific pronoun is known crosslinguistically, but not the reverse (Haspelmath 1997) (universal directionality); and (b) precisely this development occurred in French, slightly earlier than the earliest Middle Breton texts (transfer). Old French *quel ... que* ‘which(ever) X that’ developed from free relative marker ‘whatever, whichever’ along the same pathway to free-choice indefinite marker, attested in the fourteenth and fifteenth centuries, illustrated in its former use in (24), to the Modern French *quelque*-series (*quelque* ‘some’, *quelqu’un* ‘someone’, *quelque chose* ‘something’ etc.) (Foulet 1919).

- (24) Qui femme prend, de quelque taille, /Il ne puet faillir a bataille.
 who wife takes of any size he NEG can lack at battle
 ‘Anyone who takes a wife, of whatever / any size, he cannot be short of battles.’
 (Jean le Fèvre, *Les lamentations de Matheolus* l. ii, v. 3817–18)
 (Old French, c. 1371) (Foulet 1919: 227)

The Breton pattern seems likely then to be a case of transfer from French, in which case we can eliminate it from our reconstruction.

4.3.2 Reconstructing the reanalysis

This leaves us with three types of free relative:

- (a) free relatives with a strong *wh*-pronoun preceding *bynnag* in all the Brythonic languages (the Welsh *pwyl bynnag*-type);
- (b) free relatives with a weak *wh*-pronoun preceding in Middle Cornish (the *pepenag*-type);
- (c) free relatives with *pynag* alone in Middle Cornish.

³ Morris-Jones (1913: 294) suggests a connection with *pan* ‘when’ and *a(c)* ‘and’, but the logic of this seems weak. Hemon (1975: 144) describes the origin of Breton *bennak* as ‘obscure’. I shall therefore pursue Lewis & Pedersen’s suggestion as the only coherent proposal worth considering. Its coherence is reinforced by the possibility of basing detailed reconstructed reanalyses around it.

Since patterns (a) and (b) are essentially variants of one another, differing only in whether a strong or weak form of the *wh*-pronoun is used, we are reduced to considering two basic hypotheses with regard to the surface outputs of the grammar:

- (i) pattern (a)/(b) is original and Cornish innovated pattern (c), dropping the *wh*-pronoun;
- (ii) pattern (c) is original, but pattern (a)/(b) was innovated, adding the *wh*-pronoun; pattern (c) was lost everywhere except Cornish.

‘Majority rules’ would suggest treating the Cornish patterns in (c) as an innovation (although both hypotheses are equally economical provided pattern (a)/(b) is innovated in the latter stages of the parent language in hypothesis (ii)). However, the basis for the Cornish innovation, for instance, in terms of reanalysis, would be quite unclear: why would Cornish speakers innovate free relatives without a *wh*-pronoun on the basis of primary linguistic data that contained a *wh*-pronoun?

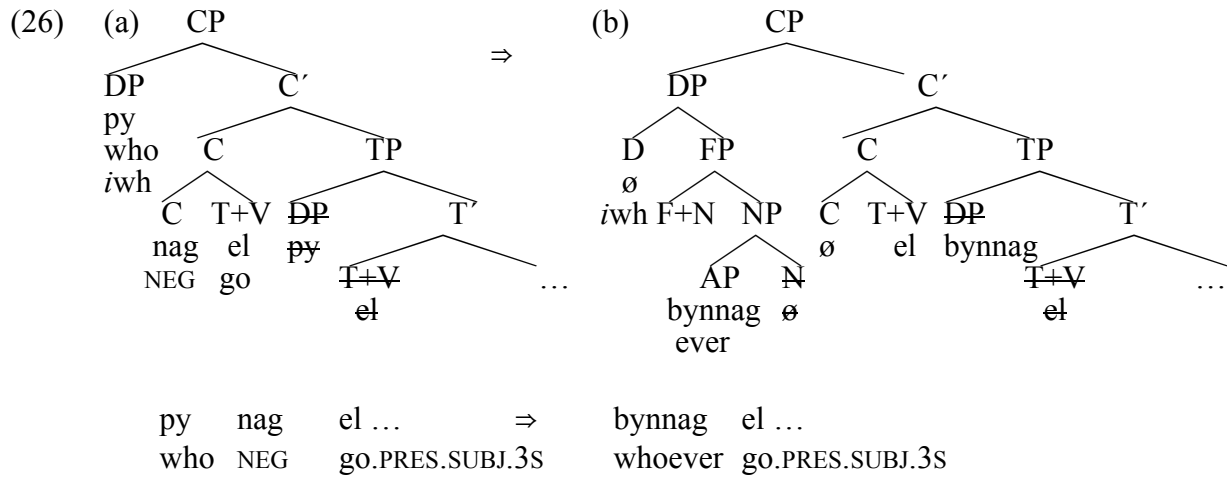
Hypothesis (ii) raises the following analogous question: why would speakers of Brythonic innovate the *pw y bynnag*-type if their language allowed only the *bynnag*-type? The obvious answer is that it seemed natural for a free relative pronoun to contain a *wh*-pronoun and, crucially, speakers did not perceive *bynnag* to be a *wh*-pronoun, that is, it was not a *wh*-pronoun in their grammar. Using the unmotivated element *bynnag* in free relatives seemed odd, and the degree of motivation of the grammar could be increased by introducing *wh*-pronouns into this context. To put it more concretely, non-specific free-relatives with *bynnag* had the same structure as specific free relatives, such as that in (25), but lacked the *wh*-pronoun found in the latter. This could be remedied by introducing a *wh*-pronoun into *bynnag* free-relatives.

- (25) A sef a wnaeth y uorwyn ryuedu [pvy ae galwassei].
 and FOC PRT do.PAST.3S the maiden marvel.INF who REL+3FS.ACC call.PLUPERF.3S
 ‘And the maiden marvelled at who had called her.’
 (*Brut Dingestow* f. 305.11–12) (Middle Welsh)

However, this suggestion requires the language to have a poorly motivated system which is made more transparent via an innovation (an extension rather than a reanalysis). It also raises another question, namely, how had this unmotivated system arisen in the first place? Why use a completely non-transparent free-relative marker like *bynnag* in the first place? Reanalysis often leads to the loss of transparency in a system (loss of structure), and in this case we can suggest that the lack of motivation was the result of a reanalysis that changed an earlier grammatical system in which *bynnag* made sense, that is, a system in which *bynnag* was treated as having internal syntactic structure and as containing a *wh*-pronoun.

We therefore posit, during the Common Brythonic period, the reanalysis illustrated in (26) (using a hypothetical example clause *py nag el* ‘whoever may go’ with Middle Welsh forms for the purposes of exemplification):⁴

⁴ In (26), it is assumed that clause-initial particles such as the negative *nag* are complementizers (head of C) and that inflected verbs raise to adjoin to these particles passing through T on the way. For discussion, see Borsley & Roberts (1996). The position of moved elements is indicated by strikethrough on the original copy.



In (26), *py nag* ‘who/what not’ is reanalysed as a single, unanalysable item ‘whoever, whatever’. This reanalysis obscures the fact that *py nag* contains a *wh*-pronoun, creating pressure for the creation of new free relative pronouns with full forms of *wh*-pronouns added to the front, that is, the forms which appear as Middle Welsh *pwy bynnag*, Middle Breton *piu pennac* and Middle Cornish *pyw penag* ‘whoever’ etc.

In terms of Figure 1, adapted to the current situation in Figure 3, the appearance of *bynnag* in Output A was motivated with the two elements in *bynnag* having their original functions, with *py* an interrogative pronoun heading a free relative clause introduced by the negative particle.

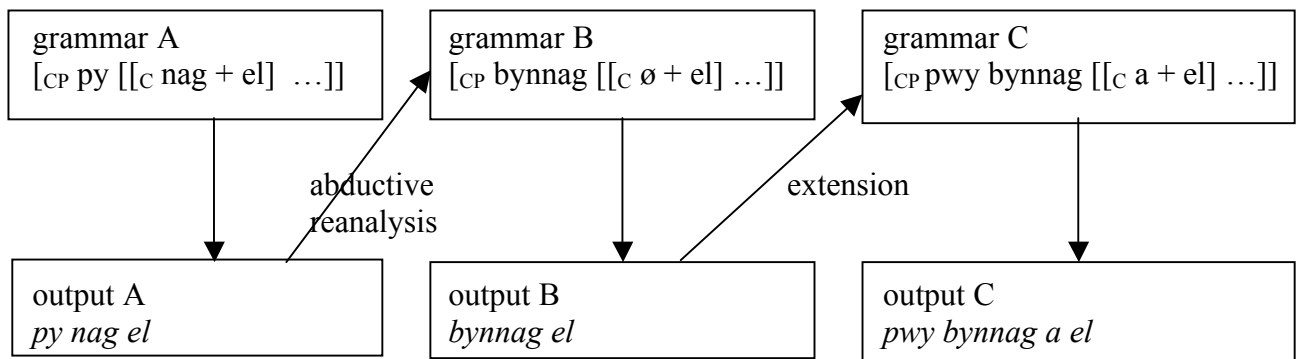
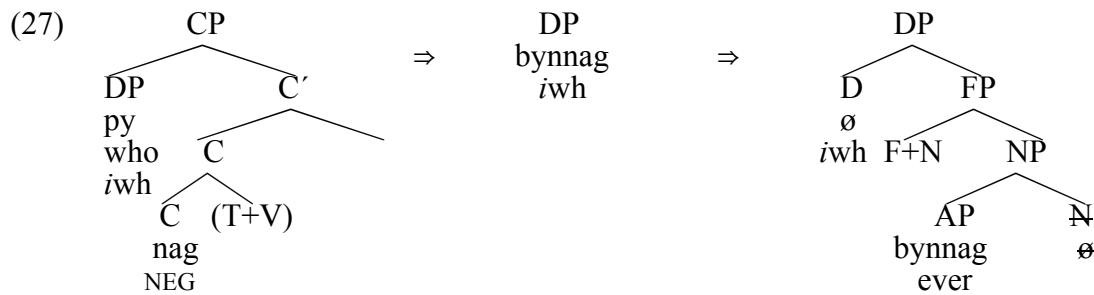


Figure 3. Reanalysis and extension of Brythonic free relatives.

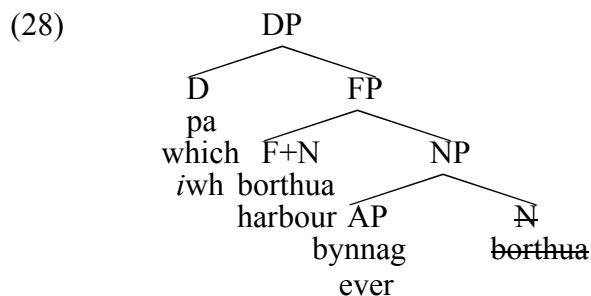
The free relative clause is in the subjunctive (but see evidence in section 4.3.4 below that this was not compulsory). As is common in irrealis contexts in many languages, it contains pleonastic negation, cf. Russian *kto by to ni byl* ‘anyone’ < ‘whoever it might (not) be’ (Haspelmath 1997: 136, Lewis & Pedersen 1937: 231). Compare also pleonastic negation in other contexts in various languages, such as, French *Je doute, qu’il ne soit là* ‘I doubt that he’ll be there’ (Rowlett 1998: 26–7), or German *Was es nicht alles gibt!* ‘(Marvel at) what was(n’t) there’.

In formal terms, a new type of *wh*-noun phrase is introduced into the grammar via the introduction a new item, *bynnag*, into the lexicon. This reflects the two-stage reanalysis given in (27).



In the first stage of the reanalysis, the word boundary between *py* and *nag* fails to be acquired, leading to the postulation of a distinct (new) lexical item, *bynnag* (previously only the *wh*-element *py* and the negative particle *nag* needed to be listed). This item is treated as a *wh*-pronoun (marked with an interpretable *wh*-feature). At the second stage, acquisition of *bynnag* as a *wh*-pronoun fails, and it is treated as an adjunct to a null *wh*-phrase. This second stage of reanalysis may have been triggered by the declining frequency of the weak *wh*-pronoun *py*. Like the early version of *bynnag*, *py* was neutral with respect to ontological category (it could mean ‘who(m)’, ‘what’, ‘where’ or ‘why’), but its strong pronoun successors increasingly distinguished category (e.g. Middle Welsh *pwy* ‘who’, *(pa) beth* ‘what’, *pa le* ‘where’, *paham* ‘why’). The failure of *bynnag* to distinguish category may have led to an extension, the creation of sequences such as *pwy bynnag* ‘whoever’, which suggested a new status for *bynnag* as an adverbial element. This is reflected in the final analysis for *bynnag* in (27). Here it adjoins within the *wh*-phrase, allowing other elements to appear. In the rightmost structure in (27), I assume a null D-element bears the interpretable *wh*-feature, turning the phrase syntactically into a *wh*-phrase. A null head noun undergoes raising to an intermediate functional head position (given the neutral label F), as can be assumed for all noun phrases in Brythonic (given noun–adjective word order).

There are two immediate results of this. First, the grammar expands by generating *bynnag*-phrases with a lexical element within the *wh*-phrase, such as *pa borthua bynnag* ‘whichever harbour’ (‘which harbour ever’) in Middle Welsh in (8) above, and parallel forms in the other daughter languages:



A second result concerns the particles in C. Since the ancestral construction involves pleonastic negation, we do not expect to find an affirmative relative particle in it. In all Brythonic languages there are separate relative particles in the affirmative and in the negative. As a result of the reanalysis, *nag* is no longer analysed as a negative relative marker, but is instead an integral part of *bynnag*. This leaves C empty in (26)b above, which makes the *bynnag*-clause exceptional. Affirmative relative clauses contain a relative particle, in Middle Welsh either *a* or *y(d)*. We therefore expect extension of the patterns normally found in affirmative relative clauses to free relatives. This is the pattern actually found in all the medieval Brythonic languages:

- (29) A phwy bynhac a vynho ennill clot
 and who ever REL want.PRES.SUBJ.3S win.INF praise
 ‘and whoever wants to earn praise’ (Peredur 57.28) (Middle Welsh)
- (30) piou pennac an exerczo
 who ever REL+3MS.ACC practise.FUT.3S
 ‘whoever practises it’ (Middle-Breton Hours 12) (Middle Breton)
- (31) penagel a gows er ow fyn
 whoever REL speak.PRES.3S against.me
 ‘whoever speaks against me’ (Bewnans Ke 3077) (Middle Cornish)

4.3.3 The status of weak and strong *wh*-pronouns in the protolanguage

While Welsh and Breton only ever use stressbearing *wh*-pronouns to create new free relative pronouns, Cornish uses both stressbearing *wh*-pronouns and unstressed *py* to do this, thereby creating both *pepenag* and *pyw penag* for ‘whoever’. The latter option is, however, rather rare. How has this state of affairs come about?

We know that unstressed *wh*-forms became restricted to adnominal use in all the Brythonic languages (Welsh *pa dŷ*, Breton *pe di* ‘which house’) and became frozen in other contexts where they were found (e.g. with a preposition such as Welsh *pam* < *paham* ‘why’ < *am* ‘for’ or Breton *perak* ‘why’ < *rak* ‘on account of’ etc.) (Lewis & Pedersen 1937: 226–9). We must suppose that their distribution was wider in Common Brythonic and contracted in late Common Brythonic and in the daughter languages. This leaves us with a range of possible hypotheses:

- (i) the complex free relative pronouns were innovated while weak *wh*-pronouns were the norm in Cornish, but only once strong ones had generalized in Welsh or Breton; this may indicate that they were innovated earlier in Cornish or (more likely given that Middle Cornish seems to retain the weak *wh*-pronouns better than Welsh or Breton) that Cornish use of weak *wh*-pronouns was simply more conservative;
- (ii) the complex free relative pronouns were innovated in Late Common Brythonic and both strong and weak *wh*-pronouns were possible; only strong ones survived in Welsh and Breton; predominantly weak ones survived in Cornish.

According to hypothesis (i), the absence of a form such as ***pybynnac* in Middle Welsh and Middle Breton is explained by the fact that the weak *wh*-pronoun *py* was already obsolete as a free pronoun in those languages when *wh*-pronouns were introduced into *bynnag*-phrases. According to hypothesis (ii), these developments, although they show some degree of differentiation between the daughter languages, are ascribed to the period of the parent language. That is, both the ancestor of Middle Welsh *pwy bynnac* / Middle Breton *piu pennac* / Middle Cornish *pyw penag* and the ancestor of Middle Cornish *pepenag* arose in the parent language. The latter remained morphologically transparent as a sequence of *pe*- ‘*wh*-pronoun’ and *penag* ‘-ever’ in the early stages of the daughter languages, and hence fell out of use in Welsh and Breton as the unstressed form of the *wh*-pronoun fell out of use. In Cornish, the loss of the unstressed form of the *wh*-pronoun fell out of use more slowly. Consequently, *pepenag* grammaticalized as a single unit before the loss of the *wh*-pronoun *pe*, thereby surviving into the attested period.

It is hard to evaluate these hypotheses against one another. Universal directionality (‘strong forms replace weak forms’) points against hypothesis (ii), since it requires Cornish to increase the frequency of a weak form (*pepenag*) over that of a strong form (*pyw penag*) over time. This then points towards hypothesis (i) as correct, but is hardly conclusive.

4.3.4 Negative particles in the protolanguage

Consider next another question: why is there a word-final velar, either /k/ or /g/, in all of the languages? The reconstruction suggests the negative marker used was *na(c)*. This makes no sense in Middle Welsh. Middle Welsh has three negative markers: *ny(t)* in main clauses and in relative clauses; *na(t)* in non-relative embedded clauses; and *na(c)* in certain modal clauses, specifically imperatives and optatives, and also in responses to questions.⁵ All are clause-initial, immediately preceding the verb. The consonant in parentheses appears when the particle appears before a vowel. *Wh*-words are followed by relative-clause syntax, which requires *ny(t)* in Middle Welsh:

- (32) *ny* *wydynt* *pwy* *a* *oed* *yn* *ev* *herbyn*. *na* *phwy* *nyd* *oed*.
 NEG know.IMPF.3P who REL be.IMPF.3S against.3P nor who NEG.REL be.IMPF.3S
 ‘they didn’t know who was against them and who was not.’
 (*Brut y Brenhinedd* Cotton Cleopatra B.v. 91.6) (Middle Welsh)

On the basis of Middle Welsh, then, we would expect *bynnag* to be based on *ny(t)*, not on *na(c)*. However, Breton and Cornish are more promising. Middle Breton has two negative particles: *ne(nd)* (Old Breton *nit*) in main clauses and *na(c)* in all embedded clauses and in imperatives and optatives (Hemon 1975: 282). In relative clauses and after *wh*-words, we find *na(c)*:

- (33) *an* *nep* *na* *eu* *discret*
 the anyone NEG be.PRES.3S discreet
 ‘anyone who is not discreet’ (*Mirouer de la Mort* 1200) (Middle Breton)

Middle Cornish is parallel, using *ny(ns)* in main clauses and *na(g)* in embedded clauses and imperatives (Lewis 1946). Again, the relevant form in a relative clause and after a *wh*-word is *na(g)*:

- (34) *cusyll* *nag* *o* *vas*
 advice NEG be.IMPF.3S good
 ‘advice that was not good’ (*Pascon agan Arluth* 31.3) (Middle Cornish)

These patterns are summarized in Figure 4.

	Middle Welsh	Middle Breton	Middle Cornish
main clauses	<i>ny(t)</i>	<i>ne(nd)</i>	<i>ny(ns)</i>
embedded clauses	<i>na(t)</i>	<i>na(c)</i>	<i>na(g)</i>
relative clauses	<i>ny(t)</i>	<i>na(c)</i>	<i>na(g)</i>
imperatives	<i>na(c)</i>	<i>na(c)</i>	<i>na(g)</i>

Figure 4. Clause-initial negative particles in the medieval Brythonic languages.

We can conclude that *bynnag* must have grammaticalized in a language which used *na(c)* in relative clauses. This has the side effect of leading us to reconstruct *na(c)* in negative relative

⁵ Negative interrogatives are excluded from discussion here as they differ in more fundamental ways in their syntax across the early Brythonic languages.

clauses in the parent language, with Middle Welsh extending *ny(t)* into relative clauses from main clauses (local directionality).

We also need to explain why the form with a velar consonant generalizes irrespective of whether a vowel or consonant follows, that is, why we do not end up with *bynna(g)* instead of *bynnaɡ*. This would suggest that, in its formative period, *bynnaɡ* was used habitually preceding a verb that began with a vowel. The only realistic prospects are the indicative forms of the verb ‘be’ in the present and imperfect tenses (Middle Welsh *yw* and *oed* respectively). Some Welsh dialects are reported with forms of *bynnaɡ* without a final velar, namely *benna* or *bynna* (Morris-Jones 1913: 293, Thomas 1950–2002: 364). Morris-Jones treats these as innovative, since they are attested late, but we cannot exclude the possibility that they are archaisms reflecting an earlier ancestral situation with an alternating form *bynna(g)*.

4.3.5 Residual issues

Finally, this reconstruction explains the irregular stress pattern of Breton *bennak*: exceptional word-final stress is found because *bennak* was originally two words. Stress was on the negative element as it in some other cases of these patterns, cf. the stress pattern of Russian *któ by to ní byl* ‘whoever’ with stress on the negative marker *ni*.

5 CONCLUSION

In this article I have argued that the absence of any direct counterpart of correspondence sets in syntax need not be an insuperable barrier to syntactic reconstruction. Maintaining a distinction between an abstract grammatical system (I-language) and the surface manifestations of that system (E-language), we can analyse the grammatical systems of the daughter languages and investigate what sequences of reanalysis and extension could have given rise to that microvariation. Directionality exists at two levels: the more reliable ‘local directionality’, which constrains reconstructions to those hypotheses which include plausible reanalyses consistent with detailed aspects of the known daughter language systems; and the less reliable ‘universal directionality’, which guides syntactic reconstruction through broad-brush rules of thumb, just as it does in phonological reconstruction. This approach to syntactic reconstruction can be embedded within a standard generative model of syntactic change without the need to abandon central distinctions such as that between the grammatical system and the surface output. Reconstruction does not meet a brick wall at historical points where reanalysis cuts us off from the past. Coupled with other more traditional notions such as economy, use of archaism and reversal of analogy, this method has proven fruitful in reconstructing parts of the grammar of the Brythonic ancestral language. While these reconstructions do not touch on large-scale word-order parameters such as those favoured in the 1970s, the reconstructions that they provide are not trivial, nor are they limited to cases of identity or cases where minor distorting data can be pruned away to reach identity.

Texts cited

Middle Breton

Ernault, Émile, ed. 1887. *La vie de sainte Catherine*. *Revue Celtique* 8, 76–95.

Ernault, Émile, ed. 1888. *Le mystère de sainte Barbe*. Paris: Librairie du Collège de France.

Ernault, Émile, ed. 1914. *Mirouer de la Mort*. Paris: H. Champion.

Ernault, Émile, ed. 1928–30. *Le breton de Gilles de Keranpuil*. *Revue Celtique* 45, 202–71; 47, 72–159.

Hemon, Roparz, ed. 1977. *Doctrin an Christenien*. Dublin: Dublin Institute for Advanced Studies.

Largillière, René, ed. 1929. Le dialogue entre Arthur et Guinclaff. *Annales de Bretagne* 38, 627–74.

Stokes, Whitley, ed. 1876. *Middle-Breton Hours*. Calcutta: n.p.

Middle Cornish

Norris, Edwin, ed. 1859. *The Ancient Cornish Drama*. Oxford: Oxford University Press [Contains *Origo Mundi*, *Passio Christi* and *Resurrexio Domini Nostri Jhesu Christi*].

Stokes, Whitley, ed. 1860–1. Pascon agan Arluth. Appendix to the *Transactions of the Philological Society* 1860–1.

Stokes, Whitley, ed. 1864. *Gwreans an Bys. The Creation of the World, A Cornish Mystery*. London: Williams and Norgate.

Stokes, Whitley, ed. 1872. *The Life of Saint Meriasek, Bishop and Confessor*. London: Trübner and Co.

Thomas, Graham C. G. & Williams, Nicholas, eds. 2007. *Bewnans Ke: The life of St Kea*. Exeter: University of Exeter Press.

Middle Welsh

Bromwich, Rachel & Evans, D. Simon, ed. 1992. *Culhwch ac Olwen. An edition and study of the oldest Arthurian tale*. Cardiff: University of Wales Press.

Goetinck, Glenys W., ed. 1976. *Historia Peredur vab Efwrawc*. Cardiff: Gwasg Prifysgol Cymru.

Jones, Thomas, ed. 1992. *Ystoryaeu Seint Greal*. Cardiff: Gwasg Prifysgol Cymru.

Lewis, Henry, ed. 1942. *Brut Dingestow*. Llandysul: J. D. Lewis a'i Feibion.

Parry, John Jay, ed. 1937. *Brut y Brenhinedd: Cotton Cleopatra version*. Cambridge, Mass.: Mediaeval Academy of America.

Williams, Ifor, ed. 1930. *Pedeir Keinc y Mabinogi*. Cardiff: Gwasg Prifysgol Cymru.

References

Andersen, Henning. 1973. Abductive and deductive change. *Language* 49, 765–93.

Borsley, Robert D. & Andreas Kathol. 2000. Breton as a V2 language. *Linguistics* 38, 665–710.

Borsley, Robert D. & Ian Roberts (eds.). 1996. *The syntax of the Celtic languages*. Cambridge: Cambridge University Press.

Bybee, Joan L. & William Pagliuca. 1987. The evolution of future meaning. *Papers from the 7th International Conference on Historical Linguistics*, 109–22. Amsterdam: John Benjamins.

Bybee, Joan L., William Pagliuca & Revere D. Perkins. 1991. Back to the future. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, ii.17–58. Amsterdam: John Benjamins.

Bybee, Joan L., Revere D. Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.

Campbell, Lyle. 1990. Syntactic reconstruction and Finno-Ugric. In Henning Andersen & Konrad Koerner (eds.), *Historical Linguistics 1987: Papers from the 8th International Conference on Historical Linguistics (8. ICHL) (Lille, 31 August – 4 September 1987)*, 51–94. Amsterdam: John Benjamins.

Campbell, Lyle. 1998. *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press.

Campbell, Lyle. 2001. What's wrong with grammaticalization? *Language Sciences* 23, 113–61.

- Campbell, Lyle. 2003. Beyond the comparative method? In Barry Blake & Kate Burridge (eds.), *Historical linguistics 2001: Selected papers from the 15th International Conference on Historical Linguistics, Melbourne, 13-17 August 2001*. Amsterdam: John Benjamins.
- Campbell, Lyle & Alice C. Harris. 2002. Syntactic reconstruction and demythologizing 'Myths and the prehistory of grammars'. *Journal of Linguistics* 38, 599–618.
- Campbell, Lyle & Marianne Mithun. 1980. The priorities and pitfalls of syntactic reconstruction. *Folia Linguistica Historica* 1, 19–40.
- Evans, D. Simon. 1964. *Grammar of Middle Welsh*. Dublin: Dublin Institute for Advanced Studies.
- Falileyev, Alexander. 2000. *Etymological glossary of Old Welsh*. Tübingen: Max Niemeyer Verlag.
- Fleuriot, Léon. 1964. *Le vieux breton: Éléments d'une grammaire* (Collection Linguistique Publiée par la Société de Linguistique de Paris 63). Paris: Librairie C. Klincksieck.
- Foulet, Lucien. 1919. Études de syntaxe française: I Quelque. *Romania* 45, 220–49.
- Friedrich, Paul. 1975. *Proto-Indo-European syntax* (Journal of Indo-European Studies, Monograph 1). Butte, Montana: Montana College of Mineral Science and Technology.
- Givón, Talmy. 1999. Internal reconstruction: As method, as theory. In Spike Glidea (ed.), *Reconstructing grammar: Comparative linguistics and grammaticalization*, 107–59. Amsterdam: John Benjamins.
- Guardiano, Cristina & Giuseppe Longobardi. 2005. Parametric comparison and language taxonomy. In Montse Batllori, Maria-Lluïsa Hernanz, Carme Picallo & Francesc Roca (eds.), *Grammaticalization and parametric variation*, 149–74. Oxford: Oxford University Press.
- Harris, Alice C. 1985. *Diachronic syntax: The Kartvelian case*. Orlando: Academic Press.
- Harris, Alice C. & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
- Harrison, Shelly. 2003. On the limits of the Comparative Method. In Brian D. Joseph & Richard D. Janda (eds.), *Handbook of historical linguistics*, 213–43. Oxford: Blackwell.
- Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Clarendon Press.
- Hemon, Roparz. 1975. *A historical morphology and syntax of Breton*. Dublin: Institute for Advanced Studies.
- Hemon, Roparz. 1976–. *Geriadur istorel ar brezhoneg*. Quimper: Preder.
- Labov, William. 1981. Resolving the Neogrammarian controversy. *Language* 57, 267–308.
- Lehmann, Winfred P. 1974. *Proto-Indo-European syntax*. Austin: University of Texas Press.
- Lewis, Henry. 1946. *Llawlyfr Cernyweg Canol*. Cardiff: Gwasg Prifysgol Cymru.
- Lewis, Henry & Holger Pedersen. 1937. *A concise comparative Celtic grammar*. Göttingen: Vandenhoeck & Ruprecht.
- Lightfoot, David W. 1979. *Principles of diachronic syntax*. Cambridge: Cambridge University Press.
- Lightfoot, David W. 2002. Myths and the prehistory of grammars. *Journal of Linguistics* 38, 113–36.
- Longobardi, Giuseppe & Cristina Guardiano. 2009. Evidence for syntax as a signal of historical relatedness. *Lingua* 119, 1679–706.
- Meillet, Antoine. 1931. Caractère secondaire du type thématique indo-européen. *Bulletin de la Société de linguistique de Paris* 32, 194–203.
- Miller, D. Gary. 1975. Indo-European: VSO, SOV, SVO, or all three? *Lingua* 37, 31–52.
- Morris-Jones, John. 1913. *A Welsh grammar*. Oxford: Oxford University Press.

- Newmeyer, Frederick J. 2001. Deconstructing grammaticalization. *Language Sciences* 23, 187–220.
- Norde, Muriel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.
- Osthoff, Hermann & Karl Brugmann. 1878. *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*, vol. i. Leipzig: S. Hirzel.
- Pires, Acrisio & Sarah G. Thomason. 2008. How much syntactic reconstruction is possible? In Gisella Ferraresi & Maria Goldbach (eds.), *Principles of syntactic reconstruction*, 27–72. Amsterdam: John Benjamins.
- Roberts, Ian. 2007. *Diachronic syntax*. Oxford: Oxford University Press.
- Roberts, Ian & Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.
- Rowlett, Paul. 1998. *Sentential negation in French*. Cambridge: Cambridge University Press.
- Thomas, R. J. (ed.) 1950–2002. *Geiriadur Prifysgol Cymru: A dictionary of the Welsh language*. Cardiff: Gwasg Prifysgol Cymru.
- Timberlake, Alan. 1977. Reanalysis and actualization in syntactic change. In Charles N. Li (ed.), *Mechanisms of syntactic change*, 141–77. Austin: University of Texas Press.
- Tomlin, R. S. O. 1987. Was ancient British Celtic ever a written language? Two texts from Roman Bath. *Bulletin of the Board of Celtic Studies* 34, 18–25.
- Van Gelderen, Elly. 2004. *Grammaticalization as economy*. Amsterdam: John Benjamins.
- Watkins, Calvert. 1976. Towards Proto-Indo-European syntax: Problems and pseudo-problems. In Sanford B. Steever, Carol A. Walker & Salikoko S. Mufwene (eds.), *Proceedings of the Chicago Linguistics Society: Papers from the parasession on diachronic syntax*, 305–26. Chicago: Chicago Linguistics Society.
- Willis, David. 1998. *Syntactic change in Welsh: A study of the loss of verb-second*. Oxford: Clarendon Press.
- Willis, David. 2010. Degrammaticalization and obsolescent morphology: Evidence from Slavonic. In Katerina Stathi, Elke Gehweiler & Ekkehard König (eds.), *Grammaticalization: Current views and issues*. Amsterdam: Benjamins.