Chapter 12

Syntactic reconstruction in Celtic

Marieke Meelen
University of Cambridge

Reconstructing syntax has been the most challenging part of research into language history, phylogeny and proto-languages. Problems regarding directionality, transfer and in particular correspondence seemed impossible to address, until recent advances in syntactic theory. In this chapter, I discuss ways to address these issues, focusing on (generative) Minimalist approaches to syntactic reconstruction in particular, with concrete examples from Celtic languages. These methods combine the goal of reconstructing language history, with the analytical tools of formal grammar and cognitive science.

1 Introduction

The comparative historical study of languages, first seen in the work of Marcus Zuerius van Boxhorn and Sir William Jones in the 17th and 18th centuries, laid the foundation for breakthroughs in the reconstruction of a Indo-European proto-language. Historical linguistics textbooks praise the comparative method as the "central" method of reconstruction (Campbell 2013; Trask 1996; Fox 1995; Hock 1991).[1] There are, however, few clear definitions or step-by-step execution guides for the method (Fox 1995: 57; Durie & Ross 1996; Walkden 2014), and it focused

[1]The first visualizations of relations between languages go back to the 17th century with some Germanic varieties by Georg Stiernhielm in 1671 (Sutrop 2012). Other Indo-European languages followed much later (around 1800) in the work by Felix Gallet (Auroux 1990). These early diagrams generally showed networks, however, rather than rooted trees. Trees were popularized in the second half of the 19th century by Auguste Schleicher's 1853 *Stammbaum*, which had 'Celten' as the first ones to branch off from 'Slawogermanen', 'Lateiner', 'Griechen' and 'Arier' (see Schleicher 1853, but also his main work from 1861–1862, as well as Morrison 2016 and Pellard et al. forthcoming for a full history of phylogenetic relations).



Marieke Meelen. 2025. Syntactic reconstruction in Celtic. In Andrew Carnie, Diane Ohala, Dee Hunter, Samantha Prins, Michael Hammond & Luis Irizarry (eds.), *Foundational approaches to Celtic linguistics*, 417–467. Berlin: Language Science Press. DOI: 10.5281/zenodo.15654879

only on the reconstruction of phonology, morphology and lexical items. This is illustrated by the omission of syntax in Indo-European textbooks (e.g., Beekes 1995; Szemerényi 1996; Krasukhin 2017). Traditionally, Indo-Europeanists have been skeptical about extending this methodology to syntax (Meillet 1954; Jeffers 1976). Because there was no comprehensive syntactic theory in the early heyday of the comparative method, syntactic reconstruction was limited to the placement of individual grammatical items (Wackernagel 1892), or mood and modal categories (Thurneysen 1885 and Delbrück 1893). The fact that publications on PIE alignment systems (e.g., Uhlenbeck 1901; Pedersen 1907 and Vaillant 1936 vs Villar 1983; Rumsey 1987; Lehmann 1993, etc.) and "basic word order" (Dressler 1969; Lehmann 1974; Friedrich 1975; Miller 1975) reached different conclusions only reaffirmed the idea going back to Delbrück himself (Delbrück 1893: v-vi) 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). Proponents of more recent methods in grammaticalization and syntactic theory (see Section 2), however, have had a more positive outlook on the possibility of unveiling the syntax of earlier languages. In addition to recovering the syntactic history of a proto-language, reconstruction methods can also potentially tell us more about whether and how languages are related and where related languages originate (i.e., the time-lined topology, subgrouping and root of a phylogenetic reconstruction). Hartmann & Walkden (2024) test a number of grammatical factors in a wide range of languages and conclude that syntactic properties are very weak when it comes to encoding phylogenetic information specifically. However, we might also use reconstruction to learn more about how syntax in general or the word order of one specific language can change over time and why it has changed in one particular way (Ferraresi & Goldbach 2008: 9).

In this chapter, I shed light on this last goal of syntactic reconstruction – focusing on Celtic specifically. Despite the fact that the verb-initial word order in insular Celtic languages is an interesting outlier in the family, major publications on Indo-European syntax (e.g., Delbrück 1893; Lehmann 1974) do not discuss any examples from Welsh, Cornish, Breton or Irish. Important early comparative Celtic grammars like Lewis & Pedersen (1961) devote very few words to syntax and are often limited to the position of the verb only (Thurneysen 1927: 301; Lewis & Pedersen 1961; Watkins 1964; Meid 1968). It is not until later that we start seeing comparative syntactic research on the insular Celtic languages. Where work on Celtic syntax initially focused on synchrony (e.g., Harlow 1981; Hendrick 2000; McCloskey 1996; Carnie et al. 2000), in the 1990s, we start to see the first attempts in syntactic reconstruction (e.g., Koch 1991; MacCana 1991; Eska & Evans 1993;

Isaac 1993; Eska 1996 for the verbal complex and basic word order; Willis 1998; 2011 for V2 and free relatives; Newton 2006 for pre-Irish C-domains; Meelen 2016 for British preverbal particles). These studies are extremely useful, but do not yet cover all aspects of Celtic syntactic reconstruction.

I start this chapter by outlining some issues with syntactic reconstruction in general, with specific reference to some examples from Celtic. I discuss how a range of more recent methods have tried to address those problems (Section 2). I then zoom in on generative Minimalist approaches in particular (Section 2.3) in order to show that Celtic languages can teach us a lot about how to develop these methods further (see Sections 3.1 on Celtic and 3.2 on Celtic in contact). Finally, in Section 4, I claim that the reverse is also true, namely that Minimalist syntactic reconstruction can teach us more about the history of Celtic languages.

2 Methods of syntactic reconstruction

2.1 Problems with syntactic reconstruction

Some traditional methods for syntactic reconstruction exploit unidirectional tendencies of grammaticalization. Givón's (1971) adage that "today's morphology is yesterday's syntax" as well as similar work by Gildea (2000b) and Balles (2008), for example, rely on the assumption that morphology exhibits signs of fossilized syntactic items that used to be independent in earlier stages of the language. Similarly, fossilized lexical items whose derivation is transparent could be employed for reconstruction, e.g., present day Welsh *sef* from *ys* '(it) is' +*ef* 'it/this' (Meelen 2016) or *efallai* 'perhaps' from *ef* 'he' and *(g)allai* 'could' (Willis 2011: 14). As Willis (2011) observes, in a VSO language examples like *efallai* that point to SV word order are unmotivated unless in earlier stages of the language SV orders were possible. These studies focusing on individual grammatical items are limited in scope, however, because it is not clear which earlier stage can be reconstructed, nor can it be used for syntactic changes that do not involve grammaticalization.

While sound changes can be constrained by the limits of physical articulators, syntactic *directionality* constraints are harder to discover or measure. A change from unvoiced to voiced consonants between vowels is much more likely than the reverse, for example, but directionality factors of these kind are harder to predict and/or explain in syntax. There are no articulatory restrictions on changes in word order from Object-Verb (OV) to Verb-Object (VO), for example. Typological approaches, such as those by Greenberg (1963), Lehmann (1974), Friedrich (1975), Miller (1975), but also more recent ones like Von Mengden (2008) and Ferraresi &

Goldbach (2008), use implicational relationships between grammatical phenomena to achieve a similar result. Walkden (2009) rightly points out, however, that in order to make any predictions about the syntax, this method requires historical texts in which at least one half of the relevant implicational universal is present. Von Mengden (2008: 109) acknowledges this drawback. For proto-languages, we have no texts by definition and hence no starting point for this method. Walkden (2014: 65) notes, however, that Niyogi & Berwick's (2009: 101–127) model can straightforwardly yield directionality in syntax if it is combined with a cue-based learning algorithm such as that of Lightfoot (1999). As I show later on in Section 4.1, recent discoveries in theoretical syntax – such as parameter hierarchies – may be able to shed more light on this directionality issue as well.

In addition to this so-called "directionality problem", Lightfoot (1980) draws attention to the fact that grammars 'must be created afresh by each new language learner' (Lightfoot 1980: 37), which Willis (2011: 4) has called the "radical reanalysis problem": if children had been able to work out what the earlier structure was, they would not have introduced the new structure. Strictly speaking, this idea of change due to reanalysis in first language acquisition is not unique to syntax. As Harris & Campbell (1995: 372) and Walkden (2009: 23) have pointed out, the same issue arises for phonological reconstruction, where sound change can occur by way of discrete reanalyses (Ohala 1981). Furthermore, Roberts & Roussou (2003), among others, follow the traditional "poverty-of-stimulus" reasoning going back to Chomsky (1980), observing that most of the time, first-language acquisition is remarkably successful despite the lack of direct access to parental grammar. This should follow from Chomsky (2005)'s division of first, second and third factors: "grammars actually acquired on the basis of similar primary linguistic data (PLD) do not vary substantially from one another, otherwise the acquisition task becomes intractable" (Walkden 2014: 49). This extends Willis's (2011: 2) observation that innovative grammars constrain the range of possible hypotheses about earlier stages of the grammar, which thus aids syntactic reconstruction (see Section 2.1 below).

Furthermore, Willis (2011) reminds us that grammatical rules that are present in two or more languages could represent a continuous line of transmission *or* they could be transferred via language contact (the "transfer problem"). In the correspondence sets for phonological reconstruction, borrowings are lexical and much easier to identify and eliminate. Clackson (2017) illustrates this issue with Barðdal's (2013) work on the reconstruction of the construction 'woe is X.DAT' found in a number of Germanic and other Indo-European languages. He argues that the Gothic and Armenian instances of these constructions could also simply have been transferred from Greek through the common bible translations. In

Section 3.2 below, I demonstrate that although this is a serious issue, there are interesting opportunities related to syntactic transfer in reconstruction too.

Another issue that has been raised with syntactic reconstruction is the lack of arbitrariness. Harrison (2003) argues that the real success of the comparative method in terms of making genetic inferences relies on the fact that similarities in symbolic form-meaning pairings (i.e., lexical cognates) "cannot be attributed to nature, and are unlikely to be the result of chance" (Harrison 2003: 223). "Grammatical objects", on the other hand, do not provide any such evidence for genetic relatedness, because form-meaning pairs in syntax are not arbitrary. Where lexical items represent individual linguistic signs that reside in a lexicon, which is "a repository of linguistic unpredictability" (Harrison 2003: 223), grammatical objects are "templates, diagrams, or rules" that are iconic, and not symbolic signs. Iconic signs are complex and do not reflect arbitrary mappings between linguistic categories and the world, which is why they cannot establish genetic relatedness (Harrison 2003: 225). Transferring the comparative method is difficult due to the supposed lack of regularity in syntactic change compared to the famous *Ausnahmslosigkeit* ('exceptionlessness') of sound change (Osthoff & Brugmann 1878; Labov 1981 for the regularity hypothesis in sound change; Lightfoot 2002; Pires & Thomason 2008 for difficulties transferring this to syntax). All the above factors illustrate how difficult it is for syntactic patterns (or "templates, diagrams, or rules") to meet the condition in (1):

(1) Double Cognacy Condition (Walkden 2014: 50)

In order to form a correspondence set, the contexts in which postulated cognate sounds occur must themselves be cognate.

Since any potential higher units of syntax, such as phrases or sentences, are not transmitted from generation to generation in the same way lexical items are, there is no syntactic equivalent to the "cognate" we know from the comparative method of phonological reconstruction (Lightfoot 2002: 125–126). The problem of what to compare instead of cognates, the so-called "correspondence problem" (Watkins 1976: 312), is the most difficult to address. Apart from perhaps some (poetic) stock phrases and idioms, such as PIE *$eg^{wh}ent$ *$og^{wh}im$ 'he killed (the) snake' (Watkins 1995: 301) and *$k̑lewos$ *$n̥d^{h}g^{wh}itom$ 'unperishable glory' (Kuhn 1853: 467), there is no direct correspondence between sentences of one language and another. As syntactic theory has developed in recent years, however, there are now potential answers to search for appropriate alternative comparanda.

In the next section, I discuss some more recent methods of syntactic reconstruction that address this issue. For more detailed exposés on addressing prob-

lems with syntactic reconstruction, see, among others, Bowern (2008), Willis (2011), Walkden (2014), Viti (2015), Clackson (2017), and Gildea et al. (2020).

2.2 Addressing the correspondence problem

Alice Harris and Lyle Campbell suggested that syntactic "patterns" could solve the correspondence problem as patterns can be compared cross-linguistically (Campbell & Harris 2002). They illustrate this with the reconstruction of alignment systems in Kartvelian languages (Harris 2008), using language-internal facts. Similarly, Gildea (2000a) reconstructed the Proto-Cariban verbal system and Kikusawa (2002) focused on alignment systems in Proto-Central-Pacific. While the results of these particular reconstruction efforts may be right, Willis (2011) and Walkden (2014) note a number of issues, such as the lack of psychological reality of "patterns", the unrestricted nature of the method and that "patterns" do not appear to have a clear place in syntactic theory.

This specific issue could be resolved by an approach that uses syntactic patterns stored as units of grammar: Diachronic Construction Grammar (DCxG, see Barðdal & Eythórsson 2012; Smitherman 2015; Eythórsson & Barðdal 2016; Gildea et al. 2020). In DCxG, the "construction" is proposed as the equivalent to the cognate in the comparative method, as they can be compared cross-linguistically (at least partially equivalent to Saussurean arbitrary form-meaning pairs Goldberg 1995, Croft 2001: 18) and, they are hypothesized to be transmitted across generations (Traugott & Trousdale 2013; Barðdal & Gildea 2015). Crucially for syntactic reconstruction, constructions can be either lexical, e.g., *cheeseburger* or schematic, e.g., ditransitive order [SVOO], which adds the semantics of transfer to a sentence like *I'll bake you a pie* (Goldberg 1995). Currie (2013) employs DCxG and in particular the mechanism of "motivation", which, according to him, explains the gradual rise in frequency of Absolute Initial Verb (AIV) order in Early Modern Welsh. When it comes to reconstruction of syntax using DCxG, we find numerous recent attempts such as personal prefixes in Proto-Tupí-Guaraní (Gildea 2002: 322–323), dative experiencers in Germanic (Barðdal & Eythórsson 2012), the subject possessor case marker in the Viceitic (Chibchan) alienable possession construction (Pacchiarotti 2020) and a number of other contributions in Barðdal et al. (2020). Gildea et al. (2020) note that objections regarding constructions fail to meet the Double Cognacy Condition in (1) above (Walkden 2014: 50) – because sentences (i.e., the potential higher order units) they generate are not cognate – are in fact based on a "domain error". They argue that "it is a logical error to require the equivalent of "regular correspondences" when seeking to identify cognates in other domains of language, such as syntax" (Gildea et al.

2020: 18). If this is indeed true we can still call constructions (or "patterns" for that matter) "cognate" simply because they can descend by inheritance from a common ancestor.

Rejecting the need for the Double Cognacy Condition in syntax, however, also means that despite suggested "safeguards" to identify patterns (Harris & Campbell 1995), similarities due to chance, language contact and/or typological factors are hard to identify. Some proposed methods to mitigate these issues, such as avoiding chance similarities (Seržant 2015: 125–130 and Gildea et al. 2020) can only really work in cases of identity or strong similarity of the proposed cognates in the daughter languages. This, Willis (2011) observes, is a general weakness of the CxG approach to syntactic reconstruction: when there is structural reanalysis (and thus a potential lack of identity in one or more daughter languages), this type of reconstruction can only work if it can rely on a straightforward account of how constructions are reanalyzed, innovated or dissolved. Traugott & Trousdale (2013: 35–38)'s "neoanalysis" and "analogization" provide ideas in this direction. When it comes to identifying similarities due to language contact, Daniels (2017) acknowledges this is an important issue and aims to provide a method for doing exactly this. His method exploits the possibility of a third type of construction, which is "somewhere between lexical and semantic" (Daniels 2017: 382), e.g., as in [the X-er the Y-er] (*The more you read, the less you understand*) (Fillmore et al. 1988: 506–508). Daniels (2017) builds on earlier work by Ross (2015) and Seržant (2015), suggesting that in order to exclude contact, after hypothesizing correspondence sets, it is necessary to follow the next step in the traditional comparative method ensuring that the phonological material represented in the partially schematic constructions is in fact inherited. This then could work for those constructions that have at least some inherited lexical element, but it is hard to see how this method could be extended to purely schematic constructions. Even those with lexical elements cannot always provide a straightforward answer, as Clackson (2017) has shown with his evaluation of the *woe*-DAT construction. Finally, the lack of reference to a theory of language acquisition makes it harder to extend the reconstruction of similar constructions to the reconstruction of a grammar that reaches "the same level of analysis that one would expect for an attested language" (Willis 2011: 10). In the next section, I discuss recent generative approaches to syntactic reconstruction that have aimed to do exactly that.

2.3 Minimalist approaches to reconstruction

Using the Minimalist Program, Van Gelderen (2011) formalizes syntactic change through a feature cycle in which semantic features (e.g., a first-person pronoun) turn into interpretable formal features [iF:1s] and, finally, into uninterpretable features [uF] that can be probes. She connects common cases of syntactic reanalysis to labeling of phrases, which is seen as a driving factor of change (Van Gelderen 2022). Using grammatical (formal) features in syntactic comparison and change goes back to the Borer-Chomsky Conjecture (Baker 2008: 155), which states the following shown in (2):

(2) Borer-Chomsky Conjecture (BCC)

All parameters of variation are attributable to the features of the functional heads in the lexicon.

Walkden (2009, 2014) observes that this offers a solution to the correspondence problem in cases where functional items are overt. Cognacy of lexical items can be established in the usual way employing all the advantages of the traditional comparative method for reconstruction, which means there is little room for chance or contact-induced similarities.

Willis (2011) uses the example of British Celtic free relatives, i.e., the cognates of Modern Welsh *bynnag* '-ever', Middle Welsh *bynnac*, Middle Breton *pennac* and Middle Cornish *penagh*. Just like in English, all these languages allowed for those forms to follow a *wh*-word (cf. English *whoever*, *whatever*, etc.), but they differ in whether they allow for both strong and weak *wh*-forms preceding reinforcing elements, e.g., Middle Cornish *penag-**el*** 'who-ever-**all**', etc. He first describes the exact similarities and differences between the three languages and then asks what grammar we should reconstruct for the proto-language, before carefully laying out every stage of the abductive reanalyses and extensions. He starts from the output of Common British *py nag el* 'whoever goes' with the pronoun *py* with an interpretable *wh*-feature in SpecCP and the negator *nag* + the verb *el* incorporated into the C-head. At the next stage, the negator joins the *wh*-pronoun in SpecCP opening up a slot in the C-head which can then, after extension, be filled with the regular pre-verbal particle *a*, resulting in Middle Welsh *bynnag a el* 'whoever PCL goes' (see Willis 2011: 25–36 for a full analysis). The strength of Willis's reconstruction lies in the meticulous attention to detail in justifying every reconstructed stage and form. He not only tests to what extent we can use traditional techniques for reconstruction such as majority and economy, but also refines the notion of directionality distinguishing "local directionality"

from "universal directionality", thus imposing important constraints on the possible reconstructions: the first testing plausible reanalyses that are consistent with detailed aspects of the known daughter language systems, the second through "broad-brush rules of thumb" akin to that in phonological reconstruction (Willis 2011: 36). Finally, through describing and analyzing the microvariation between the daughter languages and carefully checking each stage of reanalyses and extensions, he tests which outputs could result from each stage of the reconstructed grammars, thus addressing the "radical reanalysis" problem. This method may be rather limited as it can generally only be applied to overt lexical items with functional features that have cognates in other languages, which means it also cannot work well for establishing reconstructed stages of languages that go very far back in time. However, the results are extremely reliable when each stage of the previous grammar and the various reanalyses and extensions are carefully checked. Further examples of exploiting true cognates with functional features can be found in Walkden (2009, 2014) for Germanic and Meelen (2017) for preverbal particles in British Celtic.

Another recent Minimalist attempt at syntactic reconstruction uses the Parametric Comparison Method (PCM) pioneered by Guardiano & Longobardi (2005) but extended significantly based on Roberts (2019)'s approach to microvariation and parameter hierarchies. The Parametric Comparison Method combines the goal of reconstructing language history with the analytical tools of formal grammar and cognitive science. This method ultimately goes back to the 1980s, with work on syntactic Principles and Parameters (P&P) of Universal Grammar (UG), which concerns language acquisition, language change, markedness, implicational relations, learnability theory and opportunities for cross-linguistic comparison. Chomsky (1980) draws on Jacob's (1977) ideas on biological diversification making a specific analogy to linguistics: "the problem of accounting for the growth of different languages... is not unlike the general problem of growth, or for that matter, speciation" (Chomsky 1980: 67). The P&P approach offered solutions for Plato's Problem (i.e., the poverty of stimulus in language acquisition) and made excellent predictions for typological variation. However, within a Minimalist approach to cross-linguistic variation, the idea that parameters are part of the genetic endowment, the innate UG, is difficult to maintain, which is exactly what Roberts (2019) aims to remedy since proposals that abandon the notion of parameters altogether cannot offer viable alternatives (for a full discussion, see Newmeyer 2005 and reviews and replies by Dryer 2007 and Walkden 2014, Berwick & Chomsky 2011, Boeckx 2015, Roberts 2019: 11–13).

Ledgeway & Roberts (2017) follow Chomsky (1995: 6) in the Borer-Chomsky Conjecture (BCC): all microparameters are properties of individual heads,

i.e., the formal features of functional heads in the syntactic structure. Cross-linguistic variation in syntax is variation in the values of these features (e.g., the C(omplementizer) head having C[+wh] vs C[−wh] to capture the distinction between *wh*-movement and *wh*-in-situ languages). This aligns Roberts' earlier work (Clark & Roberts 1993, Roberts 1997, Roberts & Roussou 2003, Roberts 2007) with Chomsky's (2005) proposal of the three factors of language design and makes it more concrete:

(3) Three factors relevant for parameters (Roberts 2019: 7)
   a. Factor 1: Underspecification of formal features in UG
   b. Factor 2: Trigger experience/what the child takes in (PLD)
   c. Factor 3: Feature economy (FE) and input generalization (IG)

Unlike earlier parametric theories, formal features are therefore not specified by UG, and parametric variation arises from the interaction of the three factors. Biberauer & Walkden (2015) and Biberauer (2019) add that children postulate formal features on functional heads in cases where the PLD intake departs from the simplest Saussurean form-meaning mapping such as cases of discontinuity and displacement (traditional "Move" or "Internal Merge"), multiple realization of elements (e.g., due to morphosyntactic agreement) and apparent non-realization of meaningful elements (i.e., silence). Similarly, we can recover a parameter setting for a particular language by looking at the overt realization of features, e.g., a participant, gender and/or number feature could be overtly realized through inflection or a clitic, as shown in the Fiorentino inflectional paradigm in (4):

(4) (e) parlo, tu parli, e/la parla, si parla, vu parlate, e/la parlano
'I/you/(s)he/we/you/they speak(s)' (Fiorentino; Brandi & Cordin 1989)

In addition to this straightforward spell-out (exponence, i.e., equivalent to F ↔ /p/ in Distributed Morphology), features could be "realized" in a more indirect way, e.g., when feature values are copied into functional heads via Agree with a defective goal leading to head-movement (Roberts 2010). When features are spelled out overtly, this approach comes close to the approach adopted by Walkden (2009, 2014) and Willis (2011) above, with the distinction that feature exponents in the microparametric approach do not necessarily meet the Double Cognacy Condition, unless their exponents are in fact "true lexical cognates" that can be independently verified as "inherited" through their sound correspondences and the Regularity Hypothesis. The main advantage of reconstruction based on "true lexical cognates" resulting from the Double Cognacy Condition is that it allows us to eliminate resemblances due to chance, contact or typological factors. Examining cross-linguistic variation based on one single microparameter

would therefore yield rather limited results, which will need to be independently checked very carefully in order to establish whether they represent an inherited feature or not. Chance resemblance is furthermore problematic as all parameter settings are binary and "horizontal" transfer due to contact may be harder to detect in syntax than it is for lexical items. The Parametric Comparison Method for examining cross-linguistic syntactic variation therefore relies on not one, but many parameters, which effectively nullifies the opportunity to find chance resemblances of every parameter setting between any pair of languages and also reduces the potential impact of contact factors (see Section 3.2 for how the PCM could actually exploit potential lack of contact factors too). Typological factors, finally, are actually built into the method in the form of implicational factors and parameter hierarchies to which I turn now.

Baker (2008) observed that on the microparametric view where variation lies in the feature make up of functional heads (i.e., the BCC) "there should be many mixed languages of different kinds, and relatively few pure languages of one kind or the other" (Baker 2008: 350). Macroparameters (Baker 1996; 2008; Bošković 2008; Huang 2015), on the other hand, were traditionally proposed to be directly associated with UG principles, e.g., the Polysynthesis Parameter (Baker 1996), which is a general notion of "argument visibility" with two possible options: syntactic configurations vs formation of complex words. A macroparametric view on cross-linguistic variation predicts that there is a rigid division of all languages into clear types (e.g., OV vs VO, polysynthesis or not, etc.). This means that every category in every language should pattern in one way or another. The empirical evidence, however, shows that instead we have a more bimodal distribution: when it comes to word order, for example, we find many languages clustering around one type or another, but there are also some exceptions to these principal patterns. This suggests that we need a combination of both micro- and macroparameters to adequately account for cross-linguistic variation. Based on Kayne (2005: 10) among others, Roberts (2019) argues that in addition to formal feature values of functional heads, there are Independent Factors (IF) "which can cause groups of heads – over which we can generalize with an appropriately formulated feature system – to act in concert" (Roberts 2019: 5). This essentially turns macroparameters into a surface effect of aggregates of microparameters acting in unison, creating a strict hierarchy of parameters, shown in Figure 1.

Biberauer & Roberts (2012, 2017) thus propose a typology of parameters where macroparameters share a value of a variant feature on ALL heads; mesoparameters share it on ALL heads *of a given class* (e.g., [+V]), whereas microparameters show that value on only a small, lexically-definable subclass of functional heads (e.g., subject clitics or modal auxiliaries) and, finally, nanoparameters have this

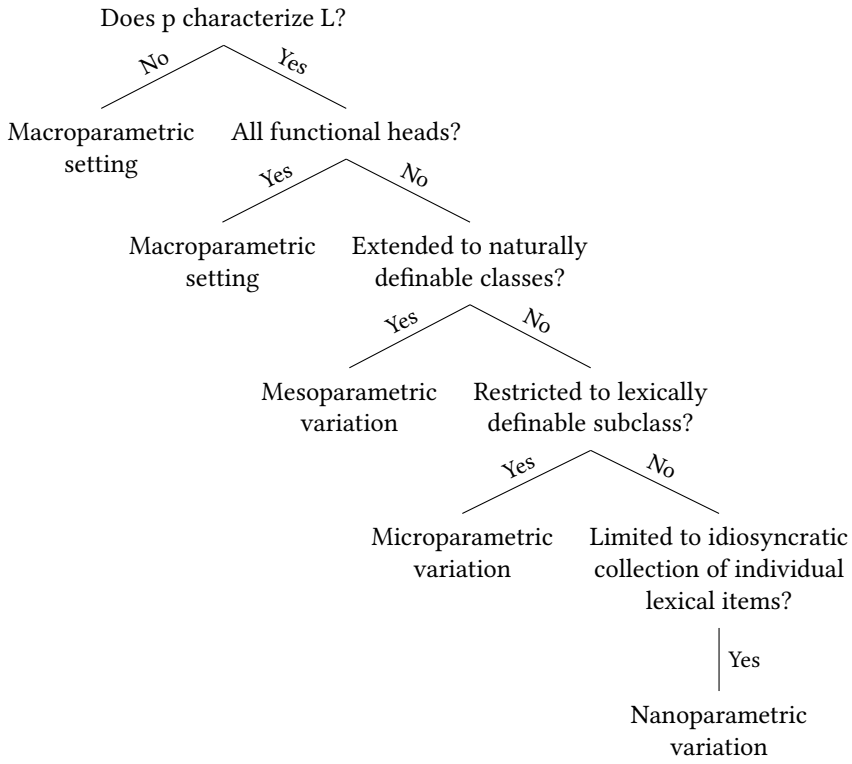


Figure 1: Parameter Hierarchies from Ledgeway & Roberts (2017: 595)

value specified to just one or more individual lexical items. This hierarchy makes specific predictions for syntactic typology, but also for language acquisition and change. Typologically, macroparameters are expected to be pervasive or universal, affecting a large number of all relevant functional heads in the grammar of a language. Macroparameters should be salient in the Primary Linguistic Data (PLD) and therefore easier to acquire and diachronically thus be more stable than microparameters. Although this is still work-in-progress, a number of hierarchies have already emerged from the work by the Cambridge ReCoS group and others based on empirical evidence from a wide range of languages (see Biberauer et al. 2014, Sheehan 2014, Van der Wal 2020, Ledgeway 2012, Ledgeway & Roberts 2017, Kinn 2017, Douglas 2017 and mainly Roberts 2019 and references therein).[2]

[2]See also https://recos-dtal.mmll.cam.ac.uk/Publications/Published.

This radical revolution in parametric theory, with well-established data-driven parameter hierarchies, gives the Parametric Comparison Method (PCM) a very solid ground for the reconstruction of language history. The PCM employs a parameter grid with binary values (+/–) for each discrete parameter in every language under investigation. This is based on a growing list of implicationally organized parameter values consisting (in its most recent iteration) of 94 nominal parameters tested on 69 languages (Longobardi 2017; Ceolin et al. 2020), as well as at least 87 verbal parameters for 39 languages (Roberts et al. 2022; Baker & Roberts forthcoming). These hierarchically organized implicational parameters differ significantly from earlier attempts to use a list of binary features to encode "language structure" such as Nichols (1992), Ringe et al. (2002), Dunn et al. (2005). The latter only look at straightforward surface patterns without taking into account what kind of grammar could generate these patterns. Similarities in surface patterns are interesting but do not necessarily indicate that languages are historically closely related, even if the pattern in question is typologically unusual. Jongeling (2000) draws our attention to this listing numerous surface patterns in Welsh that look remarkably similar to Biblical Hebrew and finds an even closer parallel in Phoenician, for example, with the unusual Determiner-Noun-Demonstrative order shown in (5):

(5) a. y llances hon [Welsh]
the girl this
'this girl'

b. ha- nnaᶜərā ha- zōt [Biblical Hebrew]
the girl the this
'this girl'

c. h gbr z'ɔ [Phoenician]
the man this
'this man' (Jongeling 2000: 81)

Similarly, Roberts (1998) notes that in Greek, for example, the same linear and concord pattern (Noun-Genitive-Adjective) also exists in Hebrew, but Longobardi et al. (2013) demonstrate that these similar surface patterns in fact reflect opposite values in two distinct parameters. In essence, the PCM here therefore requires the same amount of in-depth grammatical analysis as that shown by Willis's (2011) reconstruction of the various stages of free relatives in British: the question of how the surface patterns can be generated remains of crucial importance.

A simple example of an implicational relation in the parameters is shown by Longobardi & Guardiano (2009) as Nominal Parameter 23 (Pn23), which distinguishes languages that exhibit concord (at least) in Number with a subject nominal in predicative constructions. This parameter is contingent on the positive setting of two other parameters, namely whether there is a Number feature on N (i.e., Parameter 5 should be positive), which in turn can only be positive if there is a Number feature in the language in general (i.e., Parameter 2 should be positive). Longobardi & Guardiano (2009) note in their extensive Appendix listing all parameters in detail that this is the parameter that separates languages like English, German, Irish, Welsh and Farsi with uninflected predicative adjectives from the rest of Indo-European, as shown by examples of English and Irish, on the one hand (shown in 6), and Standard Italian and Triestino[3] (Northern Italian variety spoken in Trieste), on the other (shown in 7 and 8):

(6) a. Tá an fear beag. [masc. predicative]
be.3s the man small
'The man is small.'

b. Tá an bhean beag. [fem. predicative]
be.3s the man small
'The woman is small.'

c. Tá na daoine beag. [pl. predicative]
be.3s the people small
'The people are small.'

d. Na daoine beag**a**. [pl. attributive]
the people small.PL
'the small people'

(7) a. L'uomo è piccol**o**. [masc. predicative]
the.man be.3s small.SM
'The man is small.'

b. La donna è piccol**a**. [fem. predicative]
the woman be.3s small.SF
'The woman is small.'

c. Le persone sone piccol**e**. [pl. predicative]
the people are small.PL
'The people are small.'

[3]Many thanks to Christian Faggionato for these examples.

d. le persone piccol**e**. [pl. attributive]
the people small.PL
'the small people'

(8) a. L'omo xe pici**o**. [masc. predicative]
the.man is small.SM
'The man is small.'

b. La dona xe pici**a**. [fem. predicative]
the woman is small.SF
'The woman is small.'

c. Le persone xe pici**e**. [pl. predicative]
the people are small.PL
'The people are small.'

d. le persone pici**e** [pl. attributive]
the people small.PL
'the small people'

The PCM can be used to compare a small number of individual parameters (or even a single parameter) related to one particular grammatical feature, as is the case of Number in the DP with parameters 2, 5 and 23 above. These can also be compared in different diachronic stages of a single language to trace syntactic changes over time (see Section 4.1 below).

However, another potential strength of the PCM lies in its ability to compare a large number of parameters in many different languages to create a phylogeny that can tell us how languages are historically related to each other. One advantage is, for example, that this method can be applied to any set of languages, regardless of how similar or different they are, because the parameters are drawn from a finite, universal list. Their discrete, binary nature lends itself to precise calculations of differences between languages. There are two main methods for these types of phylogenetic computational comparisons: distance-based methods (e.g., Neighbor joining, FITCH/KITCH, UPGMA) or character-based methods (e.g., Maximum Parsimony, Bayesian analysis). The latter type of methods are based on optimizing the probability of a proposed tree given the observed data, which mainly consists of databases of lexical cognates such as Dyen et al. (1992). These methods have gained much popularity in computational historical linguistics in the last decades and are mainly used to show taxonomies as well as depth of splits of family branches (Holden 2002; Atkinson & Gray 2005; Bouckaert et al.

2012, etc.). Character-based methods crucially rely on the assumption that characters (in this case lexical cognates) are independent. The arbitrary form-meaning pairing of cognates makes them plausible candidates for this type of probabilistic modeling. However, as Jäger (2013), among others, has observed, this level of independence cannot be maintained when it comes to the regular sound changes: "It rests on the simplifying assumptions that mutations at different positions are stochastically independent and that mutation probabilities are constant across lineages" (Jäger 2013: 286.) More recent character-based studies aim to address earlier issues with this method in various ways, e.g., through finding methods to create better and more reliable input data (especially for lesser-studied language families where extensive cognate databases are still a desideratum, see List et al. 2016; 2017) or by adding specific priors (e.g., Rama 2018), weights or other additional information (e.g., Neureiter et al. 2022 attempting to include the possibility of horizontal transfer, i.e., language contact). Although these methods have achieved interesting results, with a precise representation of the history of a language family including potential dates of splits and information about how reliable each part of a given taxonomy is, Hammarström et al. (2019) and others have pointed out that some issues remain with forcing cognate classes to be binary characters (as is necessary for these models). Lexical cognates are not always perfect pairs of 1 form + 1 meaning, and often meanings have changed to the extent it is hard to establish cognacy. Finally, certain meanings (or lexical items) can be lost from the language entirely (Meelen et al. 2022).

Parameters, on the other hand, are binary by definition, but since they interact and are part of implicational hierarchies, they cannot form the basis of off-the-shelf methods as they would violate the independence of character assumption. The effect of these interdependent relations between parameters is not negligible: compared to total independence, the number of possible languages is reduced by seven orders of magnitude (with a set of 75 parameters) showing that implicational relationships between parameters play a key role (see Bortolussi et al. 2011 and Longobardi 2017 for tests of statistical significance of this and other parts of the PCM). Similarly, Ceolin et al. (2021) calculate that nearly 40% of the taxonomic input is neutralized in a dataset of 94 nominal parameters set for 58 languages. For this and other reasons (see McMahon & McMahon 2005 and others), the PCM uses distance-based methods to create phylogenies instead. As opposed to character-based algorithms, distance-based ones use pairwise distances between languages based on the identities and differences in the comparanda, i.e., each of the parameters set positively "+" or negatively "–". In addition to these binary "±" values, parameter settings could be redundant or irrelevant indicated by '0' as they are implied by a particular setting of other parameters as

we have seen in the interdependent relation of the parameters referring to the Number feature above. Finally, the positive setting of a parameter requires some specified empirical evidence (Crisma et al. 2020 and the discussion on the realization of features above), whereas negative settings represent default values. It follows then that neither pairs that contain '0', nor pairs that are both default (Language A "–" *and* Language B "–"), should be part of the distance calculation, which is exactly what the Jaccard distance, shown in (9), corresponds to (with N as the number of pairs of parameter settings in two languages A and B):[4]

(9) $$\Delta\text{Jaccard}(\text{Lang}_A, \text{Lang}_B) = \frac{[N_{-+} + N_{+-}]}{[N_{-+} + N_{+-} + N_{++}]}$$

Ceolin et al. (2020, 2021), Longobardi & Treves (2023) and others demonstrate further applications of the PCM ranging from the possibilities for the reconstruction of "deep language history" to neural implementations, i.e., how can binary parameters be conceivably implemented in the cortical circuitry in the human brain? However interesting these opportunities, I leave those aside in this chapter and instead, before moving on to more Celtic data, highlight one important observation that has direct implications for the syntactic history of a set of languages, namely, the "Anti-Babelic principle":

(10) Anti-Babelic principle (Guardiano & Longobardi 2005)

Similarities among languages can be due either to historical causes (common origin or, at least, secondary convergence) or to chance; differences can only be due to chance (no one ever made language diverge on purpose)

Although the strict version of the Anti-Babelic principle cannot hold (see, among others, examples in Thomason 2007), deliberate language change is certainly not the norm and there is ample evidence that syntactic features in particular, unlike phonetic features, are less likely to be consciously adjusted (see, for example, Smith 2000). The importance of this principle for the PCM is that it makes clear predictions about historically significant distributions of languages. If languages are known to be closely related, we expect a distance δ between 0 and 0.5, whereas the distances of other languages should tend towards 0.5 and there

[4]Note that in earlier presentations of the PCM, e.g., Longobardi & Guardiano (2009), Normalized Hamming distances were used leaving out the potential significance of negative settings as the default. As work on the PCM develops and more research goes into the exact parameters and their hierarchical relations, it is clear that Jaccard distances yield a better representation of the historical linguistic situation. See supplementary materials in Ceolin et al. (2021) for further discussion.

should be very few (if any) language pairs with a distance between 0.5 and 1. Baker & Roberts (forthcoming) decide, based on this and calculations by Bortolussi et al. (2011), that any value of $\delta < 0.20$ indicates probable genetic relatedness, whereas any value of $\delta > 0.40$ is likely to be random. They illustrate this with results from 87 verbal parameters in known language families where Celtic, Slavic language family and Romance languages have an average $\delta = 0.15$ and Sinitic and Semitic have an average $\delta = 0.18$ and $\delta = 0.19$, respectively. Larger families such as Germanic or all of the Indo-European languages in their sample ($n = 39$) have $\delta = 0.22$ and $\delta = 0.27$ respectively, whereas a comparison of all languages in their database yields $\delta = 0.37$. By comparison, if we focus on just 56 parameters of the nominal domain, Celtic languages are even more closely related with $\delta = 0.02$ (according to Longobardi 2017: 257).

In the next few sections, I first show why Celtic languages are so important for the development of these types of Minimalist syntactic reconstruction methods. Since there already are good examples in Celtic of using cognate functional items such as Willis (2011)'s reconstruction of British free relatives summarized, I focus here on parametric Minimalist methods only. For both the verbal and the nominal parametric distance results discussed above, only data from Modern Irish and Modern Welsh were taken into account, but here I extend this with data from older stages of these languages (Middle Welsh and) as well as occasional examples from other Celtic languages like Cornish and Breton. I then show how the PCM in particular can help advance research on Celtic languages. Finally, I discuss some strengths and limitations of both types of Minimalist approach, ultimately arguing that we need both for a more comprehensive syntactic reconstruction.

3 How Celtic can help Minimalist syntactic reconstruction

The Celtic branch of Indo-European forms an ideal testing ground for the PCM. First of all, it contains a number of languages, both with attested older stages and with languages that are still spoken in the present day. The latter gives us the opportunity to get native-speaker judgments on every possible parameter providing the kind of detailed syntactic information that is necessary for in-depth analyses. As for the older stages, Celtic has both inscriptions and manuscripts from a wide range of times and places. These show that while there is no doubt that languages like Lepontic, Gaulish, Irish and Welsh are related, they exhibit numerous differences – especially in the syntax of their oldest stages. Whereas some branches of Indo-European are relatively uniform in terms of, for example,

"basic word order" (e.g., verb-second in Germanic), Continental Celtic inscriptions show important variation from their Insular Celtic counterparts. These differences within a set of languages that have been grouped together through the traditional comparative method of phonological reconstruction provide us with the opportunity to enhance and refine the set of parametric questions we ask. Finally, from a purely geographical point of view the Celtic languages are of crucial importance in the reconstruction of Indo-European as they provide an excellent test case for what can happen at the forefront of Indo-European expansion (Ram-Prasad & Meelen 2022). In what follows, I show how certain aspects of Celtic syntax can help to refine the set of parameters used in the PCM, and how evidence from some Celtic languages like Middle Cornish can provide more insight into issues of syntactic transfer in contact situations.

3.1 Celtic languages refining the PCM

From a syntactic point of view, the Insular Celtic languages in particular are interesting since their verb-initial nature sets them apart from other Indo-European languages. Other syntactic features that are found in Celtic, but not in other Western European languages include inflected prepositions, shown in (11) with examples from Cornish and Breton, and the lack of a 'have'-like predicate to express possession in Welsh and Irish in particular, shown in (12).)

(11) warnav/-nas/-nodho/-nydhi/-nan/-nowgh/-nedhe [Middle Cornish]
warnon/-nout/-nañ/-ni/-nomp/-noc'h/-nezho [Breton]
'on me/you/him/her/us/you(PL)/them' (Toorians 2014 & Wmffre 1998: 30)

(12) a. Mae gen i feic/annwyd [Welsh]
is with.1s me bike/cold
'I have a bike/cold.'

b. Tá teach ag an mbean. [Irish]
is house with the woman
'The woman has a house.'

Modern Breton, however, does have a lexical verb 'have' namely *kaout/endevout* (shown in 13a), but this verb is historically derived from the verb *bezañ* 'be' + a prefixed personal pronoun. Breton can also indicate possession using a preposition *gant* 'with' (shown in 13b), which is cognate with Welsh *gan* 'with' above.

(13) a. Bremañ Azenor ha Iona o deus un ti [Breton]
now Azenor and Iona 3P have.PRES a house
'Azenor and Iona have a house now.' (Jouitteau 2005: 373)

b. Ganti e oa teir yar. [Breton]
with.3SF PRT be.IMPF.3S three hen
'She had three hens.' (Press 1986: 140)

However, these examples of syntactic variation within and beyond the Celtic language family are not covered by the currently-available sets of parameters in the nominal and the verbal domains (see Crisma et al. 2020 for 94 nominal parameters and Baker & Roberts forthcoming for 87 verbal parameters).

Furthermore, although insular Celtic languages show remarkable similarities in terms of syntax, especially in the nominal domain, there are also cases of (micro)variation. Another example of such variation that is currently not captured by the PCM framework concerns resumptive pronouns. Resumptive pronouns are generally not available in subject positions cross-linguistically, even in languages like Irish which usually allow resumptive pronouns in other positions, e.g., direct objects, shown in (14):

(14) a. an fear a raibh (*sé) breoite [Irish subject relative]
the man rel was he ill
Intended: 'the man who (he) was ill' (McCloskey 1990: 210)
b. an fear ar bhuail tú é [Irish object relative]
the man rel struck you him
'the man that you struck (him)' (McCloskey 1990: 206)

In Modern Irish, resumptives are banned in subject position only in the main clause of a *wh*-construction. This so-called "Highest Subject Restriction" is also found in Hebrew and Palestinian Arabic (Shlonsky 1992 and Alexopoulou 2006). In Modern Welsh, on the other hand, resumptive pronouns are not allowed in subject position in main clauses and in those embedded clauses where a gap is possible (Borsley et al. 2007: 108).

(15) a. y dyn gafodd (*e) 'r wobr [Welsh subject relative]
the man get.PAST.3S (he) the prize
'the man who got the prize'
b. y ffrwydrad glywais i (*e) wedyn [Welsh object relative]
the explosion hear.PAST.1S I (it) then
'the explosion I heard then' (Borsley et al. 2007: 119)

Welsh does allow resumptives in relatives based on possessives and objects of prepositions, as shown in (16):

(16) a. y myfyrwyr werthodd Ieuan y ceffyl iddyn **nhw** [Welsh]
the students sell.PAST.3S Ieuan the horse to.3P them
'the students who Ieuan sold the horse to'

b. y dyn welais i **ei** chwaer [Welsh]
the man see.PAST.1S I 3SM sister
'the man whose sister I saw' (Borsley et al. 2007: 121)

This Welsh pattern is also found in other languages, e.g., Czech (Toman 1998) and Berber and Turkish (Ouhalla 1993). McCloskey (1990) and Willis (2000), among others, have analyzed this as a constraint on binding with respect to pronouns in certain domains (the A'-Disjointness Requirement). More research on this microvariation in the distribution of resumptive pronouns in languages like Irish and Welsh can therefore clearly provide more comparanda for the PCM, which in turn will help it make better representations of differences and similarities between subbranches in particular, but also across language families.

When it comes to the nominal domain, the first set of large-scale parametric comparisons to create phylogenetic trees (Guardiano & Longobardi 2005) was based only on a set of 54 distinct parameters. As research on the PCM evolved, these parameters became more and more refined and further parametric distinctions were added. This was partly due to new discoveries in the field of nominal syntax, but also because more languages, especially non-Indo-European ones, were added to the PCM databases. Even within the most recent publications using a total of 94 nominal parameters and a total number of 69 languages from a wide range of language families, Celtic is only represented by Modern Irish and Welsh, however. It is therefore clear that more research on the syntax of Celtic languages, especially older stages as well as those not yet queried by the PCM, such as Scottish Gaelic and Modern Breton, can help improve Minimalist syntactic reconstruction.

3.2 Celtic languages in contact

It is easier to detect similarities due to language contact with Neogrammarian lexical cognates. There is no clear equivalent to the regularity of sound laws in syntactic change, but parameter hierarchies and the implicational relations found in theoretical syntax do provide some opportunities to distinguish inherited structure from transfer due to language contact. Ledgeway & Vincent (2025), for example, show that syntactic changes due to contact do not follow the step-by-step parameter hierarchy from macro- to meso-, micro- and nano- or vice versa. Recall from Section 2.3 that nanoparameters involve features on individual lexical

items, whereas microparameters refer to a lexically definable subclass of functional heads, e.g., auxiliary verbs. Mesoparameters deal with core functional categories or larger natural classes of functional heads and their effects. They differ from macroparameters as they involve formal features of some but not *all* functional heads acting together, such as head-movement in Germanic and Romance. When it comes to endogenous changes, we expect parameters to change step-by-step from one level to the next, as shown by Romance active participle agreement. Ledgeway & Vincent (2025) observe that parametric comparison of Romance and Greek varieties spoken in the south of Italy instead show abrupt and saltatory movements through parametric space indicating exogenous change, i.e., changes due to contact. The Romance language Aspromonte Calabrese, for example, which has been in contact with Greko (a variety of Greek spoken in Italy), has lost auxiliary selection completely, "jumping" from a nanoparametric setting with auxiliary selection sensitive to argument structure to a macroparametric setting found in Greko that has no auxiliary selection at all. Similarly, in Old Calabrese syncretic marking of internal arguments was restricted in a nanoparametric fashion to 1/2 person only. Endogenous changes in this and non-contact varieties in Romance led to the loss of this 1/2 person restriction, moving one step up the hierarchy. However, contact with Greko then led to a subsequent jump up two levels much closer (but not identical) to the level of Greko, showing that exogenous changes can lead to hybridism which is more like approximation than convergence, but still predicted through the strict distinctions made by parameters. Although the biggest leap or "catapult effect" from nano- to macroparametric setting is actually predicted for some forms on endogenous change as well (e.g., passives), Ledgeway & Vincent (2025) show that parameter hierarchies can still be used as a good heuristic for detecting contact-induced change.

We can explore the issue of transfer further through investigating a group of closely-related languages, where we know some have undergone more contact-induced change than others. The British Celtic languages form an excellent case study, since both Breton and Cornish are clearly heavily influenced by French and English respectively, with which they were/are in close contact. For Cornish, Padel (2021) notes that "Cornish shows extensive influence from English" (256), even in the earliest texts going back to the 15th century, because Cornwall by that time had been an administrative county of England for five centuries already. Further evidence comes from the widespread use of Anglo-Saxon forenames in the area as well as the numerous loanwords that are fully assimilated into Cornish found in the *Old Cornish Vocabulary*, dating from the later twelfth century (Graves 1962). Humphreys (1992) finds similar patterns for Breton, which

is heavily influenced by French already in the medieval period as it ceased to be the language of the ruling class in the 12th century.

When it comes to syntax, there are clear differences between Breton and Cornish, on the one hand – sharing certain features with French and English, respectively – and Welsh, on the other. This poses the question of whether these differences are due to language contact. Where Irish and Welsh lack a lexical verb 'have', as shown in Section 3.1 above, both Breton and Cornish have innovated these through the grammaticalization of an infixed object pronoun and a form of the verb 'be', as shown in (17a) for Middle and (17b) for Modern Breton:

(17) a. mil uileny diouz tut an bro ouz bezo huy
1000 insult from people the land 2PL be.FUT.3S you
'you'll have 1000 insults from the people of the land' (B 176.766)

b. Ur c'hi ac'h eus? - N 'em eus ki ebet.
Q dog have.2S - NEG have.1S dog none
'Do you have a dog? - I haven't got a dog.' (Press & Ar Bihan 2004: 57)

Although this grammaticalization could potentially be reconstructed for Proto-British if occasional instances in early poetry show this innovation was incipient in Welsh as well (Lewis & Piette 1966: 50–52), the increased level of contact with languages that did have lexical 'have' like English and French could have accelerated this grammaticalization in Breton and Cornish. The subsequent developments of using the verb 'have' to form a perfect is directly modeled on French (for Breton, cf. Willis forthcoming). For Cornish, Williams (2011: 328) cites one example shown in (18), which could be influenced by Middle English. However, Padel (2021: 258) argues that Cornish never developed an analytic perfect with the verb 'to have', and (18) shows a stative verb, so this subsequent change may only have occurred in Breton.

(18) flehys a m bes denethys (Middle Cornish)
children PRT 1S have.3S beget.PPP
'I have begotten children' (CW 1979)

Another example of grammaticalization is the development of the Cornish volitional verb *mynnes* 'to wish' to a future marker. This is crosslinguistically common (e.g., Danish *ville*, Greek *tha* < *thelo ina* 'I wish that', Bulgarian *šte* < 'I want that'), but not found in Welsh or Breton. Old English *wile* 'want, wish' (> PDE future *will*) unlike other PDE auxiliaries that only grammaticalized later, could already have future meaning before the Middle English period (Warner 1993).

This early grammaticalization in English makes it a good candidate for contact-induced, or at least contact-accelerated grammaticalization in Cornish.

Similarly, if we compare the word order in declarative matrix clauses across the medieval British languages, we can see a clear V2 pattern, which can be reconstructed to Proto-British (alongside V1) (Meelen 2016, Willis forthcoming). In Middle Cornish, however, there are many more exceptions to strict V2 with subject pronouns following initial constituents, shown by the following examples from Eska & Bruch (2021: 151, 155):

(19) a. an dragan me a ra guan [Middle Cornish]
the dragon I PRT make.PRS.3S weak
'I shall make the dragon weak.' (BM 232.4004)

b. the 'th scoforn wharre yehes sur me a re
to your ear soon health surely I PRT give.PRS.3S
I shall give, surely, healing to your ear soon.' (PC 312.1150–1151)

These could be due to poetic meter (Eska & Bruch 2021). However, the examples with unstressed subject pronouns are also found in neighboring Old English and southern Middle English (Kroch et al. 2000), language contact is very likely to have influenced this Cornish innovation as well (Willis forthcoming).

An example of an innovation in Middle Breton and Cornish that is very clearly modeled on French and English, respectively, is the passive. Similar to Irish, voice alternations in Brittonic can be expressed through impersonal verbal morphology, which is present across all three medieval languages, e.g., Middle Cornish *gylwyr* 'is called', *gyllyr* 'one can', Middle Breton *mireur* 'is watched', *guilir* 'one sees' and Middle Welsh *byddir* 'one may be', *lladawr* 'will be killed'. When it comes to specific passive features, such as the use of 'by'-phrases, there is variation with Middle Breton and Welsh lacking by-phrases with impersonal morphology, where this may be found in Middle Cornish, as shown in (20). The use of the by-phrase in later Welsh and Breton appears to be a parallel innovation due to contact with English and French, respectively (Willis forthcoming).

(20) maythogh keris gans lues [Middle Cornish]
that love.IMP by many
'that you are loved by many' (BM 272)

Examples like the above in Cornish with *-is/-ys* can be ambiguous, however, between imperfect impersonal morphology and past participles with left-out auxiliary 'be'. Periphrastic passives with 'be' + past participle are also found in Breton,

which just like in Cornish is due to language contact (Lewis & Piette 1966: 49–50, Le Roux 1957: 125–126).

(21) a. prak gans pup na vethaf lethys [Middle Cornish]
why by all NEG be.1S slay.PPP
'why by all I may not be slain' (O.595-96)

b. ha pa ez vezo laqueat mat euez [Middle Breton]
and when PRT be.FUT.3S put.PPP good attention
'and when good attention has been (will be) paid...' (Cathell 5)

Middle Welsh, on the other hand, innovates a *cael*-passive, which is typologically similar to the English *get*-passive, although this was probably an internal innovation (Willis forthcoming). Baker & Roberts (forthcoming) call the verbal parameter that captures the distinction between periphrastic and non-periphrastic passives Pv38 (VSV) 'Voice spread to V', which essentially asks whether voice is realized both on V and separately. The subtle distinctions between different types of periphrastic constructions exhibited by British Celtic languages are not captured by any specific parameter yet, however. Note that opportunities for further extension of the parameters in the passive domain could be based on the suggestion that the *by*-phrase doubles the passive argument *-en* (Baker et al. 1989: 223), which is compatible with the observation that there is a distinction between transitive and unaccusative readings of, for example, 'cost' in Welsh when they are used in the impersonal form, as shown in the Modern Welsh examples in (22) (see Arman 2015: 133–136 and Roberts 2019: 425fn5 for discussion):

(22) a. costiwyd y CD yn ddeg-punt gan y cwmni.
cost.PAST.IMP the CD PRED ten-pounds by the company
'The CD was costed at 10 GBP by the company.'

b. *Costiwyd (yn) deg-punt (gan y CD).
cost.PAST.IMP PRED ten-pounds by the CD

All of the examples in this section so far do show that when we know enough about the history of languages and their historical contact situations, this can inform our syntactic reconstructions. British Celtic languages form an excellent case study for this since we have medieval data for all three and despite the fact that they are closely related, we do see clear syntactic differences as well. All of these differences can be due to transfer from French (in Breton) and English (in Cornish), because we know about the socio-historical situation in which these

languages came into contact. Furthermore, we know enough about the historical stages of both French and English to be able to answer questions regarding syntactic parameters like Pv38 on periphrastic passives, which is set positively in both Middle English and Middle French.

The PCM offers another possible way to explore cases of (potential) syntactic transfer. Since loanwords are essentially cast aside before applying the comparative method to reconstruct language families, the resulting tree topologies are supposed to show vertical inherited lineages only, neglecting data that could indicate language contact. This means that even in large-scale Bayesian applications of phylogenetic reconstruction based on large cognate databases, historical cases of language contact (or, 'horizontal transfer') can only be detected using additional measures (as done by, for example, Neureiter et al. 2022). Since there is no such filtering out of contact features in the first stage of syntactic reconstruction using the PCM, this means the resulting topologies of the reconstructed trees would include similarities due to transfer in their vertical branches if there were any. Therefore, if we compare the traditional phonological and syntactic tree topologies, we would then expect them to differ exactly in those cases where language contact played a role. Since syntactic transfer is often argued to be more difficult than lexical borrowing, we may not see as many differences, but still, family trees and in particular subfamilies would be predicted to look slightly different. Although other factors could play a role here too, this prediction is certainly borne out in both the nominal and the verbal sets of parameters, as shown by Baker & Roberts (forthcoming) in the case of Germanic in particular. Such instances of divergence between phonological and syntactic reconstructions force us to zoom into the exact cases of identity and difference in syntactic features, thus providing us with a detailed research program. In the next section I discuss more examples of how Minimalist syntactic reconstruction, and the PCM through the use of parameter hierarchies in particular, can help us gain more insight into (Celtic) language history.

4 How Minimalist syntactic reconstruction can help Celtic

Although we do have older attestations of some modern Celtic languages like Irish, Breton and Welsh, evidence is still scarce compared to other branches of the Indo-European language family. The first Old Irish and Old Welsh texts are of a much later date than the first attestations we have in Greek, Latin or Indo-Iranian, which is one of the reasons why Celtic has always lagged behind in providing evidence for Proto-Indo-European reconstruction. In addition, various

(stages of) Celtic languages are not even attested at all, such as the language spoken by the first Celtic speakers moving to the British Isles or any of the modern languages that could have developed from later stages of Gaulish, Celtiberian or Lepontic. Finally, extant evidence for any of the Continental Celtic languages is scarce, and mainly consists of short inscriptions, which severely restricts the possibilities of getting a comprehensive picture of the overall grammatical system of any of these languages. Issues like these are further complicated by ambiguous and sometimes contradictory evidence from historical phonology. Some scholars (e.g. Kortlandt 2007; Schrijver 1997 and more recently Weiss 2022) have postulated an Italo-Celtic branch of Indo-European, based on certain similarities in phonology and morphology such as sigmatic future with reduplication, substitution of **-bhos* for PIE dat.PL **-mus* and the lack of innovation of an elaborate middle voice. More recent Bayesian approaches based on lexical cognates often place Celtic and Romance languages close together as well, e.g., Atkinson & Gray (2005: 215) and Neureiter et al. (2022: 7). However, even Kortlandt (2007) who presents a concrete reconstruction of Proto-Italo-Celtic, acknowledges that "specific Italo-Celtic innovations are few" (Kortlandt 2007: 157), which has led others (e.g., Watkins 1966) to reject this Italo-Celtic hypothesis at all originally (see Weiss forthcoming, however, for shared morphosyntactic features of preverbs and adpositions). Similarly, intermediate subgroups within Celtic itself are not always easy to establish. Insular Celtic languages like Welsh and Irish exhibit a number of similarities (Schrijver 1995: 463–465), but in some respects the British Celtic languages appear much closer to Gaulish (e.g., as a P-Celtic language) leading scholars like Sims-Williams (2003) to hypothesize that "British" could be inherited from both Gallo-Brittonic as well as Insular Celtic.

In both of these cases comprehensive methods for syntactic reconstruction like the PCM offer opportunities to fill the gaps in our linguistic history. This is not just apparent for phylogenetic topology on a larger scale, but also for specific gaps in our knowledge of syntactic features of Celtic languages on a smaller scale. In this final section, I first focus on syntactic changes in the history of some insular Celtic languages specifically, to create a better picture of the grammars of intermediate stages like Proto-British, but also to identify gaps in our knowledge and opportunities for future research. Finally, I discuss examples of syntactic features shared by Celtic and Romance, as well as where they differ based on a comparison of modern languages like present-day Irish and Welsh, on the one hand, and French, Spanish and both Northern and Southern Italian varieties, on the other, which leads to a final discussion of Celtic as a subfamily of Proto-Indo-European in general.

4.1 Reconstructing intermediate stages

In Section 3.2, we saw examples of syntactic innovations in the Voice system of British Celtic languages: both Middle Cornish and Breton already show examples of the *be*-passive, whereas Late Middle Welsh innovated the *cael*-passive, which behaves more like the English *get*-passive. Old and Middle Irish, on the other hand, did not have periphrastic passives, which would leave the verbal parameter Pv38 (VSV) in the negative setting as there is no evidence for a Voice feature being realized on both the verb and separately with an auxiliary. Present-day Irish may have innovated a periphrastic passive too, but this is so far limited to the perfective passive only (McCloskey 1996: 254–255). Since the periphrastic passives are a later innovation in Cornish, Breton and especially Welsh (all of which still have morphological passives and impersonals), we should reconstruct the negative setting of this parameter for Proto-British as well as Proto-Insular-Celtic (if this is an intermediate stage we want to reconstruct at all). The potential positive setting for present-day Irish depends on the exact interpretation of the verb 'be' as a possible auxiliary, which is debatable and it may mean this parameter is still set to negative even today. Note that this discussion reveals two important points about the PCM. First, the restriction to perfective passives in present-day Irish could lead us to refine the passive parameter hierarchy to allow for this type of microvariation. Second, it was the parametric questions and the comparison of the cross-linguistic settings, which highlighted this gap in our knowledge about Irish to begin with.[5]

A further example of systematic comparison driven by the PCM can be found in the domain of negation. Baker & Roberts (forthcoming) posit Pv83 (NMU) as the parameter for "Multiple Negation", which essentially asks whether a language permits negative concord. Modern Irish and Welsh differ in this respect, as shown in (23) and (24), with only the latter overtly realizing this feature thus rendering a positive setting.

(23) Níor chreideas é [Modern Irish]
NEG believe.PAST.1S 3SM.ACC
'I did not believe him.' (Nolan 2018)

(24) a. D-oes neb yn ennill. [Modern Welsh]
NEG-be.3S no.one PROG win.VN

[5]Many thanks to Jim McCloskey and Ian Roberts for sharing their discussion notes, which brought this to my attention.

b. *Mae neb yn ennill.
be.3s no.one prog win.VN
'No one is winning.' (Willis 2013)

Middle Welsh did not have negative concord, on the other hand, predominantly showing Stage I of Jespersen's Cycle with preverbal negation only, shown in (25a). Middle Breton already appears to be in Stage II showing both preverbal negation as well as postverbal *ket*, shown in (25b), which has been spreading within the language as a Breton-specific innovation probably due to French (Hemon 1975: 281–286; Poppe 1995: 103–105; Willis 2006, forthcoming).

(25) a. **ny** wnn i pwy wyt-ti [Middle Welsh]
NEG know.pres.1s I who be.pres.2s-you
'I don't know who you are.' (PKM 2.22-3)

b. Ne chome **ket** pell e wenneien en e c'hodell [Middle Breton]
NEG stay.3s **ket** long 3SM pennies in 3SM pocket
'His pennies did not stay long in his pockets.' (MAV p.29)

Cornish, just like Irish in this case, only allows a preverbal marker *ny* or *neng/nyngs* (before forms of *bos* 'be' and *mones* 'go'), with or without preverbal subject, as shown in (26).

(26) a. **Ny** lowenhaf in ow dythyow neffra lam [Cornish]
NEG rejoice.1s in 1s days never now
'I shall not rejoice ever now in my days.' (BK 2929–2930)

b. **Nyng** ew ow thowl
NEG be.3s 1s intention
'It is not my intention.' (BK 293)

c. My **ny** won. / My a wor
I NEG know.1s / I PTC know.3s
'I don't know.' (PA 121a / OM 185)

These examples reveal another curious fact about Cornish, namely that verbs can exhibit agreement with pronominal subjects that precede the negation (Lewis 1956: 48–49(§47)), which is impossible in their positive counterparts. Pronouns cannot be overtly expressed in Cornish, or most other Celtic languages, if there is subject-verb agreement (i.e., the so-called "complementarity principle", Stump 1984: 292, Borsley & Stephens 1989; Doron 1988). However, this principle clearly

does not hold for subject-initial negative clauses in Cornish. It also does not hold for Welsh, although in post-verbal position, only the weaker forms of the pronouns (echo pronouns) can be used, as shown in (27a). In Middle Welsh, strong pronouns can be found with agreement, but crucially only when they occur in preverbal position, as shown in (27b). When it comes to nominal subject DPs, they can never agree with verbs in number (shown in 27c), which Roberts (2019: 388–389) relates to the 3rd-person forms specifically.

(27) a. dywedais i/*fi [Modern Welsh]
say.PAST.1S I.WEAK/STRONG
'I said'

b. mi a wnn pwy wyt ti ac ny chyuarchaf i [Middle Welsh]
I.STRONG PTC know.1S who are.2S you and NEG greet.1S I.WEAK
'I know who you are, but I will not greet you.' (Pwyll p1rc2l18)

c. *Canon / Canodd y plant. [Modern Welsh]
sing.PAST.3PL / sing.PAST.3S the children.PL
'The children sang.'

These subtle distinctions among the Celtic languages in terms of agreement and the link to negation in Cornish in particular reveal gaps in the grammars of the intermediate proto-languages we would like to reconstruct, showing us exactly where further research is necessary. When it comes to the use of multiple negation, however, it is clear that negative concord was an innovation in Welsh and Breton only. This cannot be reconstructed to Proto-British as they each exhibit different postverbal negative markers (Welsh *ddim* vs. Breton *ket*), which are not lexical cognates. They do share the same syntactic features, however, but Willis (2006) has shown that the changes in these features can be traced exactly through the history of Welsh. For Proto-British, as well as Proto-Insular-Celtic, we therefore reconstruct the negative value for the Multiple Negation parameter, which is already found in Cornish, Irish, as well as Old and Middle Welsh. Note that this is also in line with the negative setting of this parameter in other Indo-European languages like Latin, Vedic and Hittite, which brings us to comparing Celtic parameters beyond the British Isles. Since evidence for Continental Celtic languages is scarce, the task of reconstructing its syntax seems daunting, especially from the perspective of reconstruction methods that rely on overt (lexical) cognacy of functional items as there are very few (if any) such items that can be found in both Continental and Insular Celtic languages. The PCM offers an

opportunity here too since it not only provides a systematic approach seeking concrete answers to a fixed set of parametric questions in the nominal and clausal domains, it also allows for reconstruction based on more covert realizations, such as movement or Internal Merge due to requirements of specific features on functional heads. In addition, the strong hierarchical structure with built-in implicational parameters means that a number of parameter values can be inferred from the positive or negative setting of others. Combining these strengths of the PCM will provide experts on Continental Celtic with the opportunity to reconstruct a grammar of the proto-language that will cover more aspects of syntax than was ever possible before.

4.2 Syntactic similarities in Italo-Celtic

In this final section, I show examples of syntactic features that can be reconstructed for Celtic, Romance and/or both to shed light on the "Italo-Celtic" hypothesis from a new angle of Minimalist syntax. Obvious similarities are found for fundamental (macro)parameters, such as using bound morphemes (Pn1 FMG "grammaticalized morphology"), agreement (Pn2 FGA "grammaticalized agreement") and Case (Pn3 FGK "grammaticalized Case"), which tend to distinguish Indo-European languages from languages like Mandarin, Japanese and Wolof. Much more interesting for our purposes are the similarities and differences found within the Indo-European family and languages these came into contact with in the course of history, such as Basque, Hungarian and a range of Semitic languages.

When focusing on Celtic, Romance and Germanic in particular, there are various parameters with positive settings shared with some (or all) languages in those three subfamilies, such as grammaticalized person, number and gender in both the nominal and verbal domains. In addition, all three subfamilies share features to do with the Categorization Requirement (CR, cf. Panagiotidis 2015), which states that every c-selecting and c-selected functional head H must be uniquely identified in relation to an EP (Extended Projection). It is built on the assumption that non-categorizing heads do not have intrinsic V- or N-features, but instead bear FF [Cat:__]. The higher heads in the EP then are categorially identified as belonging to that EP. Roberts (2019: 164–166) argues that UG makes available three mechanisms to satisfy the Categorization Requirement, yielding three types with Type I satisfying CR through roll-up, labeling all the functional categories as V or N giving rise to harmonic head-finality. Type II, on the other hand, satisfies CR through head-movement yielding head-initial languages, and, Type III satisfies CR by lexicalizing heads in the EP. The "purest" versions of these three types are rare and there is a possibility to combine types for different parts

of the syntax. Celtic, Romance as well as English combine Types II and III, but other West Germanic languages can also feature Type I, e.g., German featuring all three types: roll up of VP around v, head movement as well as lexicalizing heads. Roberts (2019: 163) argues that both Welsh and English allow intrinsic labeling followed by head-movement, as shown, for example, by the inflected auxiliary *oedd* in example (28) raising to a high functional position (e.g., some C head) when there are free lower aspectual heads, which could be considered some sort of "*be*-insertion" (cf. English dummy "*do*-support"):

(28) Oedd y bachgen wedi bod yn ymlad [Welsh]
be.3S.PAST the boy PERF be.INF PROG fight.INF
'The boy had been fighting.'

Syntactic features that are even more interesting for investigating Celtic as a subfamily are those that distinguish the order of adjectives and nouns, for example, NM1 "N under M1 Adjectives". Given the cross-linguistically canonical sequence of structured adjectives, Ceolin et al. (2021: 41) say that "Manner1 adjectives" (i.e., those describing quality or size as opposed to shape or color, which are 'Manner2') can surface to the left of the noun in languages like Italian, French, Spanish, Germanic, Slavic language family and Standard Greek. In Welsh, Irish, Farsi or some Romance dialects of Italy, however, they cannot, as shown in (29):

(29) a. un bel nuovo vestito blu [Italian]
a nice new dress blue

b. ffrog las newydd neis [Modern Welsh]
dress blue new nice
'a nice new blue dress'

Although these word order facts are relatively clear, agreement with adjectives in Celtic is not as straightforward as it seems, which becomes apparent when asking these detailed parametric questions, especially for older stages of the Celtic languages. Borsley et al. (2007) claim that "[I]n predicative position, adjectives never agree in gender or number with their noun" (2007: 179), citing example (30):

(30) Mae ei lygaid yn las / *leision. [Modern Welsh]
be.3S 3SM eyes PRED blue.GENERAL / blue.PL
'His eyes are blue.'

In spoken Welsh corpora, however, it is possible to find examples, as shown in (31):

(31) roeddynt yn fychain [Modern Welsh]
were.3PL PRED small.PL
'they were small' (CEG corpus; Meelen & Nurmio 2020)

Meelen & Nurmio (2020) show that agreement in earlier stages of Welsh was different, as shown in (32), which is backed up by a corpus study of Middle and Early Modern Welsh with 17 out of 50 predicative adjectives (34%) showing agreement. Since 13 out of the 17 examples stem from the Bible translation, some further research is necessary here before we can definitively show a parametric change, however.

(32) Mae y cyflenwadau o wenith Lloegr a Thramor yn fychain
be.3S the supplies of wheat English and overseas PRED small.PL
'the supplies of English and overseas wheat are small' (Papurau Newydd: Yr Amserau 12 Nov 1851)

Other examples of parameters that put Welsh and Irish close to the majority of Romance, but also close to Germanic languages are Grammaticalized Aspect (GRA) and Mood (GRM) and Mood- and Voice-checking (MCV and VCV) in the verbal domain as well as Grammaticalized Specified Quantity (DGR) in the nominal domain (encoding definiteness).

There are, however, also various parameters that show differences between Celtic and Romance languages, especially when we zoom in to the micro- and nano-levels. Roberts (2019) argues that verb-initial languages like Modern Welsh and Irish copy φ-features (i.e., Person/Gender/Number features) of the verbal suffix in one of the other agreeing heads (enclisis), just like Northern Italian Dialects (NIDs), where this is shown by subject (pro)clitics. Unlike NIDs, the Welsh weak/clitic pronouns actually combine both Person and Number features, creating a microparametric distinction. Other Romance varieties differ even more from Celtic in this respect, since they do not exhibit agreeing subject clitics (either enclitic or proclitic) altogether. Similarly, Welsh and Irish differ from languages like French and Brazilian Portuguese by optionally allowing the suppression of φ-features on T, making them into Partial Null-Subject languages (Roberts 2019: 608). Other Romance languages go one step further, not having obligatory EPP features on T at all, yielding them as Canonical Null-Subject languages with fully specified Person-features found on both T and D heads.

In the nominal domain, Celtic languages pattern with Semitic and Icelandic freely admitting bare singular count indefinite arguments (nominal Pn19 "Weak Specified Quantity", CGR), e.g., Hebrew *kelev*, Welsh *ci* '**a** dog', where Romance

and many Germanic languages require an indefinite article, e.g. Dutch *een hond*, French *un chien* and Italian *un cane* 'a dog'. Similarly parameter Pn47 (GFL) groups Celtic languages together with Greek, most Slavic languages as well as German and Icelandic in which a non-adpositional, non-iterable Genitive Case appears to the right of canonically ordered adjectives, as shown in (33) from Ceolin et al. (2021: 31):

(33) portread hardd y plentyn [Modern Welsh]
portrait beautiful the child
'the child's beautiful portrait'

Finally, there are also parameters that set Welsh and Irish apart from not just Romance, but most of Indo-European, such as Pn55 "plural spread from cardinal quantifiers" (PSC). In Welsh, for example, singular nouns are used after cardinal quantifiers, shown in (34), rendering a negative setting for this parameter, similar to Farsi, Uralic and Turkic languages (but unlike English or the rest of Indo-European).

(34) un ffordd / tri ffordd [Modern Welsh]
one way / three way
'one way / three ways'

If we combine the results from 87 verbal (Baker & Roberts forthcoming) and 94 nominal parameters (Ceolin et al. 2021), the first observation is that languages that are usually grouped together into subfamilies using the comparative method of phonological reconstruction also form subfamilies based on syntactic features. Within Indo-European, we have a Slavic branch with Russian, Polish, Serbo-Croatian and Slovenian; a Germanic branch with Norwegian, Danish, German, Dutch, English, Afrikaans and Icelandic; a Romance branch with Spanish, Portuguese, Italian and French; and Modern Irish and Modern Welsh are grouped together representing Celtic. The Hamming distance between those two modern Celtic languages, which counts both positive and negative settings that are similar as "identities", is 0.039; the Jaccard distance, which ignores identical settings if they are both negative, is 0.167.[6] In general, the phylogenetic tree topologies

[6]Note that Baker & Roberts (forthcoming) only provide Hamming distances, whereas Ceolin et al. (2021) change their previous Hamming metric to Jaccard to be able to distinguish negative similarities from positive ones. Ceolin et al. (2021) report no difference at all, i.e., $\delta = 0.0$ between Irish and Welsh. Although the Insular Celtic DP indeed shows striking similarities, it is worth noting that 100% similarity in 94 distinct parameters is very unusual even for languages that are very closely related. In fact it is the only language pair in their dataset that shows 0 differences for relevant comparanda.

based on Hamming and Jaccard distances of all parameters combined are quite similar, although the Jaccard distances are about 0.13–0.2 larger. Note that with either metric, Irish and Welsh clearly are below the $\delta < 0.20$ threshold Bortolussi et al. (2011) suggested for probable relatedness.

If we compare both Irish and Welsh combined to other subfamilies, we see the average distance distributions in Table 1. Celtic, as represented by Irish and Welsh, on average shares more syntactic features with the Slavic languages than with Romance: both average Hamming and Jaccard distances are lower for Slavic than for Romance. If the parameters would solely be used for the purposes of detecting historical relatedness, this is a somewhat surprising result since family trees based on the traditional phonological comparative method would often posit a combined Italo-Celtic subgroup (Weiss 2022). Weiss (forthcoming), for example, shows that shared innovations and retentions in the system of preverbs and adpositions provide additional evidence for this closer relation between the Celtic and Romance languages (see also Section 4.2).

Table 1: Hamming and Jaccard distances for averaged subfamilies

	Hamming	Jaccard
Welsh vs Romance	0.204	0.364
Welsh vs Germanic	0.218	0.395
Irish vs Romance	0.196	0.358
Irish vs Germanic	0.227	0.429
Celtic vs Romance	0.200	0.361
Celtic vs Germanic	0.227	0.412
Celtic vs Slavic	0.185	0.323
Celtic vs Uralic	0.262	0.468
Celtic vs Semitic	0.233	0.400

This highlights an important point of interpretation and application of using the PCM for a variety of purposes. First, the results of detecting phylogenetic signals and establishing historical relations between subgroups should not necessarily be expected to be the same not only because lexical and phonological features on the one hand and grammatical features on the other are representing different parts of the language, but also because potential evidence for horizontal transfer (i.e., transfer due to contact, see Section 3.2) is systematically excluded for lexical/phonological reconstruction, but not for the grammatical features in

the PCM. Second, the number and range of languages and features representing a subgroup is important in any statistical reconstruction method (whether Bayesian or not). The Celtic subgroup is only represented by two languages in the PCM and as they are less-studied some features in the nominal and verbal domain still have some question marks. The Romance languages, on the other hand, are all studied in a high amount of detail and are represented even in its smallest iteration by at least five languages (French, Italian, Spanish, Portuguese and Romanian). Slavic languages currently fall somewhere in between, with four languages represented (Slovenian, Serbo-Croatian, Russian and Polish), some of which are well-studied from a syntactic point of view. Finally, it is worth bearing in mind that only the nominal and verbal parameters are part of the dataset yielding the above results. When clausal parameters that cover syntactic features of the C-domain are included, the overall results may look very different. Since Longobardi & Guardiano (2009) note that the nominal parameters are more likely to be robust over time, it is worthwhile zooming in on potential changes within the Celtic subfamily in the verbal domain. Modern Irish and Welsh show a Jaccard $\delta = 0.263$, which is much higher than the average of both nominal and verbal parameters since Ceolin et al. (2021) report 0 difference in the nominal domain. If we compare Middle Welsh to Old Irish, on the other hand, there are fewer syntactic distinctions with Jaccard $\delta = 0.171$. Between Middle and Modern Welsh verbal parameters Jaccard $\delta = 0.268$, whereas Irish appears to have changed fewer verbal parameters showing Jaccard $\delta = 0.229$.

Overall, these numbers are preliminary. As the number of parameters increases, we can expect to see a refinement of the analysis. Despite this, it is clear that subfamilies can be easily identified already and we can identify gaps in our knowledge of the old and modern Celtic languages. Even if the PCM eventually were to yield different results related to phylogenetic signal and subgrouping, we can see that it is a fruitful exercise for reconstructing the grammatical system of proto-languages.

When comparing parameter settings in both the modern and the medieval Celtic languages, we can get a much better picture of the syntactic developments going back to Proto-British and Proto-Celtic. For every reconstructed parameter whose values are different from those in the modern daughter languages, we should carefully address some of the remaining issues of syntactic reconstruction discussed in Section 2, e.g., the radical-reanalysis and transfer problems. The latter was already discussed in Section 3.2 and requires knowledge about the historical contact situation, as well as the application of methods to detect syntactic transfer in particular, such as those described by Bowern (2008) and Daniels (2017). The radical-reanalysis problem can only be addressed applying

the careful step-by-step methodology presented by Willis (2006), which involves a critical analysis of how every stage of reanalyses and extensions could have led to the syntactic patterns we find in attested stages of all the daughter languages. This is more difficult when we are forced to reconstruct negative values of parameters, since they do not have overt realizations, forcing us to rely more on implicational generalizations known from (Minimalist) theoretical syntax such as those worked out in detailed parameter hierarchies. A large number of parameters, however, (at least around a third for the verbal domain[7]) actually ask about direct PF realizations such as inflectional morphology, which can potentially be reconstructed using the traditional comparative method if they are true cognates, as done by Walkden (2011, 2014) and Willis (2011).

5 Conclusion

In this chapter, I have shown that recent advances in Minimalist syntactic theory enable us to tackle the long-standing challenges of syntactic reconstruction. While various methods of syntactic reconstruction have been put forward in the last decades, Minimalist approaches that rely on the Borer-Chomsky Conjecture (BCC) have been particularly successful in addressing problems of directionality, transfer and radical reanalysis. The Parametric Comparison Method (PCM) has some additional benefits of providing a comprehensive overview of the syntax of a proto-language because of the discrete nature of parameters, at the same time avoiding the need to find potentially improbably phenomena (e.g., sound correspondences in restricted contexts) and reliance on vague cognacy judgments (e.g., due to semantic changes). The PCM is most powerful when it is combined with the careful step-by-step evaluation of the reanalyses and extensions in all reconstructed stages shown by Willis (2011) and methods for detecting language contact in the domain of syntax (Bowern 2008, Daniels 2017).

As I showed in Section 3.1, there are various ways in which the parametric questions can be improved and extended, but the language family trees based on nominal or verbal domains (or both) show remarkable resemblances to the tree topologies based on lexical cognates using the traditional comparative method. Celtic languages can play an important role in the further development of the PCM due to their unique properties in syntactic features, even though only Modern Welsh and Irish have been taken into account so far. This knowledge can help us refine and extend the PCM, on the one hand, but on the other, the PCM can help our understanding of historical syntax by providing a systematic set of questions that can shape a clearly-defined research program for the Celtic languages.

[7] Jim Baker, p.c.

References

Alexopoulou, Theodora. 2006. Resumption in relative clauses. *Natural Language and Linguistic Theory* 24(1). 57–111. DOI: 10.1007/s11049-005-0898-2.

Arman, Laura. 2015. *The Welsh impersonal construction*. University of Manchester. (Doctoral dissertation).

Atkinson, Quentin D. & Russell D. Gray. 2005. Curious parallels and curious connections—phylogenetic thinking in biology and historical linguistics. *Systematic Biology* 54. 513–26. DOI: 10.1080/10635150590950317.

Auroux, Sylvain. 1990. Representation and the place of linguistic change before comparative grammar. In Tullio De Mauro & Lia Formigari (eds.), *Leibniz, Humboldt, and the origins of comparativism*, 213–238. Amsterdam: John Bejamins. DOI: 10.1075/sihols.49.15aur.

Baker, Jim & Ian Roberts. Forthcoming. Extending parametric comparison: Preliminary results. *Journal of Historical Syntax*. DOI: 10.33774/coe-2021-6lzqq.

Baker, Mark. 1996. *The polysynthesis parameter*. Oxford: Oxford University Press.

Baker, Mark. 2008. The macroparameter in a microparametric world. In Theresa Biberauer (ed.), *The limits of syntactic variation*, vol. 132, 351–373. Amsterdam: John Benjamins.

Baker, Mark, Kyle Johnson & Ian Roberts. 1989. Passive arguments raised. *Linguistic Inquiry* 20(2). 219–251.

Balles, Irene. 2008. Principles of syntactic reconstruction and "morphology as paleosyntax". In Gisella Ferraresi & Maria Goldbach (eds.), *Principles of syntactic reconstruction*, 161–186. Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.08bal.

Barðdal, Jóhanna. 2013. Syntax and syntactic reconstruction. In *The Routledge handbook of historical linguistics*, 343–373. Routledge.

Barðdal, Jóhanna & Thórhallur Eythórsson. 2012. "Hungering and lusting for women and fleshly delicacies": Reconstructing grammatical relations for Proto-Germanic. *Transactions of the Philological Society* 110(3). 363–393. DOI: 10.1111/j.1467-968x.2012.01318.x.

Barðdal, Jóhanna & Spike Gildea. 2015. Diachronic construction grammar: Epistemological context, basic assumptions and historical implications. In Jóhanna Barðdal, Elena Smirnova, Lotte Sommerer & Spike Gildea (eds.), *Diachronic construction grammar*, 1–50. Amsterdam: John Benjamins.

Barðdal, Jóhanna, Spike Gildea & Eugenio Lujan. 2020. *Reconstructing syntax*. Leiden: Brill.

Beekes, Robert S. P. 1995. *Comparative Indo-European linguistics: An introduction*. Amsterdam: John Benjamins.

Berwick, Robert C. & Noam Chomsky. 2011. The biolinguistic program: The current state of its development. In Anna-Maria di Sciullo & Cedric Boeckx (eds.), *The biolinguistic enterprise*, 19–41. Oxford: Oxford University Press.

Biberauer, Theresa. 2019. Factors 2 and 3: Towards a principled approach. *Catalan Journal of Linguistics*. 45–88. DOI: 10.5565/rev/catjl.219.

Biberauer, Theresa, Anders Holmberg, Ian Roberts & Michelle Sheehan. 2014. Complexity in comparative syntax: The view from modern parametric theory. In Frederick J Newmeyer & Laurel B Preston (eds.), *Measuring grammatical complexity*, 103–127. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199685301.003.0006.

Biberauer, Theresa & Ian Roberts. 2012. *The significance of what hasn't happened.* Lisbon: Presentation at Diachronic Generative Syntax 14 conference, Lisbon.

Biberauer, Theresa & Ian Roberts. 2017. Parameter setting. In *Cambridge handbook of historical syntax*, 134–62. Cambridge: Cambridge University Press.

Biberauer, Theresa & George Walkden (eds.). 2015. *Syntax over time: Lexical, morphological, and information-structural interactions.* Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199687923.001.0001.

Boeckx, Cedric. 2015. Some thoughts on biolinguistics. *Veritas (Porto Alegre)* 60(2). 207–221. DOI: 10.15448/1984-6746.2015.2.22006.

Borsley, Robert D. & Janig Stephens. 1989. Agreement and the position of subjects in Breton. *Natural Language & Linguistic Theory* 7(3). 407–427.

Borsley, Robert D., Maggie Tallerman & David Willis. 2007. *The syntax of Welsh.* Cambridge: Cambridge University Press.

Bortolussi, Luca, Andrea Sgarro, Giuseppe Longobardi & Cristina Guardiano. 2011. How many possible languages are there? *Biology, computation and linguistics* 228. 168–179. DOI: 10.3233/978-1-60750-762-8-168.

Bošković, Željko. 2008. What will you have, DP or NP? In *Proceedings ofNELS*, vol. 37:1, 101–14. New Jersey.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard & Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097). 957–960. DOI: 10.1126/science.1219669.

Bowern, Claire. 2008. Syntactic change and syntactic borrowing in generative grammar. In Gisella Ferraresi & Maria Goldbach (eds.), *Principles of syntactic reconstruction*, 187–216. Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.09bow.

Brandi, Luciana & Patrizia Cordin. 1989. Two Italian dialects and the null subject parameter. In Oxvaldo Jaeggli & Kenneth J. Safir (eds.), *The Null Subject Parameter*, 111–42. Dordrecht: Kluwer.

Campbell, Lyle. 2013. *Historical linguistics: An introduction*. 3rd edn. Edinburgh: Edinburgh University Press.

Campbell, Lyle & Alice C. Harris. 2002. Syntactic reconstruction and demythologizing "Myths and the prehistory of grammars". *Journal of Linguistics* 38. 599–618. DOI: 10.1017/s0022226702001706.

Carnie, Andrew, Heidi Harley & Elizabeth Pyatt. 2000. VSO order as raising out of ip? In Andrew Carnie & Eithne Guilfoyle (eds.), *The syntax of verb initial languages*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780195132229.003.0003.

Ceolin, Andrea, Cristina Guardiano, Monica Alexandrina Irimia & Giuseppe Longobardi. 2020. Formal syntax and deep history. *Frontiers in Psychology* 11(488871). DOI: 10.3389/fpsyg.2020.488871.

Ceolin, Andrea, Cristina Guardiano, Giuseppe Longobardi, Monica Alexandrina Irimia, Luca Bortolussi & Andrea Sgarro. 2021. At the boundaries of syntactic prehistory. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376(1824). rstb.2020.0197, 20200197. DOI: 10.1098/rstb.2020.0197.

Chomsky, Noam. 1980. On binding. *Linguistic Inquiry* 11(1). 1–46.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge: MIT Press.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36(1). 1–22. DOI: 10.1162/0024389052993655.

Clackson, James. 2017. The comparative method and comparative reconstruction. In Adam Ledgeway & Ian Roberts (eds.), *The Cambridge handbook of historical syntax*, 189–206. Cambridge: Cambridge University Press. DOI: 10.1017/9781107279070.010.

Clark, Robin & Ian Roberts. 1993. A computational model of language learnability and change. *Linguistic Inquiry* 24(2). 299–345. DOI: 10.4324/9781315310572.

Crisma, Paola, Cristina Guardiano & Giuseppe Longobardi. 2020. Syntactic diversity and language learnability. *Studi e Saggi Linguistici* 58(2). DOI: 10.4454/ssl.v58i2.265.

Croft, William. 2001. *Explaining language change: An evolutionary approach*. Oxford: Oxford University Press.

Currie, Oliver. 2013. Gradual change and continual variation: The history of a verb-initial construction in Welsh. In Anna Giacalone Ramat, Caterina Mauri & Piera Molinelli (eds.), *Studies in language companion series*, vol. 133, 43–78. Amsterdam: John Benjamins. DOI: 10.1075/slcs.133.03cur.

Daniels, Don. 2017. A method for mitigating the problem of borrowing in syntactic reconstruction. *Studies in Language* 41(3). 577–614. DOI: 10.1075/sl.41.3.02dan.

Delbrück, Berthold. 1893. *Vergleichende Syntax der indogermanischen Sprachen.* Strasbourg: E. K. J. Trübner.

Doron, Edit. 1988. On the complementarity of subject and subject-verb agreement. In Michael Barlow & Charles Ferguson (eds.), *Agreement in natural language: Approaches, theories, descriptions*, 201–218. Stanford: CSLI Stanford.

Douglas, Jamie. 2017. Unifying the that-trace and anti-that-trace effects. *Glossa* 2(1):60). 1–28. DOI: 10.5334/gjgl.312.

Dressler, Wolfgang. 1969. Eine textsyntaktische Regel der idg. Wortstellung: (Zur Anfangsstellung des Prädikatverbums). *Zeitschrift für vergleichende Sprachforschung* 83(1). 1–25.

Dryer, Matthew. 2007. Review of Frederick J. Newmeyer. possible and probable languages: A generative perspective on linguistic typology. *Journal of Linguistics* 43. 244–52. DOI: 10.1017/s0022226706254586.

Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley & Stephen C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309(5743). 2072–2075. DOI: 10.1126/science.1114615.

Durie, Ross & Malcolm Ross. 1996. *The Comparative method reviewed: Regularity and irregularity in language change.* Oxford: Oxford University Press.

Dyen, Isadore, Joseph Kruskal & Paul Black. 1992. An Indo-European classification: A lexicostatistical experiment. *Transactions of the Philological Society* 82. 1–132.

Eska, Joseph F. 1996. On syntax and phonology within the Early Irish verbal complex. *Diachronica* 13. 225–257. DOI: 10.1075/dia.13.2.03esk.

Eska, Joseph F. & Benjamin Bruch. 2021. Remarks on pragmatic fronting and poetic overdetermination in Middle Cornish. *North American Journal of Celtic Studies* 5. 131–193. DOI: 10.1353/cel.2021.0014.

Eska, Joseph F. & D. Ellis Evans. 1993. Continental Celtic. In Martin Ball & Nicole Müller (eds.), *The Celtic languages*, 26–63. London: Routledge.

Eythórsson, Thórhallur & Jóhanna Barðdal. 2016. Syntactic Reconstruction in Indo-European: State of the Art. *Veleia* 33. 83–102. DOI: 10.1387/veleia.16823.

Ferraresi, Gisella & Maria Goldbach. 2008. *Principles of syntactic reconstruction.* Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.

Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language* 64. 501–538. DOI: 10.2307/414531.

Fox, Anthony. 1995. *Linguistic reconstruction: an introduction to theory and method*. Oxford: Oxford University Press.
Friedrich, Paul. 1975. *Proto-Indo-European Syntax*. Washington: Institute for the Study of Man.
Gildea, Spike. 2000a. On the genesis of the verb phrase in Cariban languages. In Spike Gildea (ed.), *Reconstructing grammar: Comparative linguistics and grammaticalization*, 65–106. Amsterdam: John Benjamins.
Gildea, Spike (ed.). 2000b. *Reconstructing grammar*. Amsterdam: John Benjamins.
Gildea, Spike. 2002. Reconstructing Pre-Proto-Tupi-Guarani main clause grammar. In Ana Suelly Cabral & Aryon Rodrigues (eds.), *Atas do encontro internacional de pesquisadores de línguas indígenas*, 315–326. Universidade Federal do Para Belem.
Gildea, Spike, Eugenio R. Luján & Jóhanna Barðdal. 2020. The curious case of reconstruction in syntax. In *Reconstructing syntax*, 1–44. Leiden: Brill.
Givón, Talmy. 1971. Historical syntax and synchronic morphology: An archaeologist's field trip. *Chicago Linguistic Society* 7. 394–415.
Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
Graves, Eugene Van Tassel. 1962. *The Old Cornish Vocabulary*. Ann Arbor: Columbia University.
Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language* 2. 73–113.
Guardiano, Cristina & Giuseppe Longobardi. 2005. Parametric comparison and language taxonomy. In *Grammaticalization and parametric variation*, 149–174. Oxford: Oxford University Press.
Hammarström, Harald, Philipp Rönchen, Erik Elgh & Tilo Wiklund. 2019. On computational historical linguistics in the 21st century. *Theoretical Linguistics* 45(3-4). 233–245. DOI: 10.1515/tl-2019-0015.
Harlow, Stephen. 1981. Government and relativization in Celtic. In Frank Heny (ed.), *Binding and filtering*, 213–254. MIT Press.
Harris, Alice C. 2008. Reconstruction in syntax: Reconstruction of patterns. In Gisella Ferraresi & Maria Goldbach (eds.), *Principles of syntactic reconstruction*, 73–95. Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.05har.
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*, 213–243. Oxford: Blackwell.

Hartmann, Frederik & George Walkden. 2024. The strength of the phylogenetic signal in syntactic data. *Glossa* 9(1). DOI: 10.16995/glossa.10598.

Hemon, Roparz. 1975. *A historical morphology and syntax of Breton*, vol. 3. Dublin: Dublin Institute for Advanced Studies.

Hendrick, Randall. 2000. Celtic initials. In Andrew Carnie & Eithne Guilfoyle (eds.), *The syntax of verb initial languages*, 13–37. Oxford: Oxford University Press. DOI: 10.1093/oso/9780195132229.003.0002.

Hock, Hans Henrich. 1991. *Principles of historical linguistics*. 2nd edn. New York: Mouton De Gruyter.

Holden, Claire. 2002. Bantu language trees reflect the spread of farming across sub-Saharan Africa: A maximum-parsimony analysis. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269(1493). 793–799. DOI: 10.1098/rspb.2002.1955.

Huang, C. T. James. 2015. On syntactic analyticity and parametric theory. In Y.-H. Audrey Li, Andrew Simpson & W.-T. Dylan Tsai (eds.), *Chinese syntax in a cross-linguistic perspective*, 1–47. Oxford: Oxford University Press.

Humphreys, Humphrey L. 1992. The Breton Language. In Glanville Price (ed.), *The Celtic connection*, 245–275. Buckinghamshire: Rowman & Littlefield.

Isaac, Graham R. 1993. Issues in the reconstruction and analysis of Insular Celtic syntax and phonology. *Ériu* 44. 1–32.

Jacob, Francois. 1977. Darwinism reconsidered. *Le Monde, Translated in WALS Atlas (January 1976)*.

Jäger, Gerhard. 2013. Phylogenetic inference from word lists using weighted alignment with empirically determined weights. *Language Dynamics and Change* 3(2). 245–291. DOI: 10.1163/22105832-13030204.

Jeffers, Robert J. 1976. Syntactic change and syntactic reconstruction. In William M. Christie (ed.), *Current progress in historical linguistics: Proceedings of the second international conference on historical linguistics, Tucson, Arizona 12–16 January 1976*, 1–16. Amsterdam: North Holland.

Jongeling, Karel. 2000. *Comparing Welsh and Hebrew*. Leiden: Center for Non-Western Studies.

Jouitteau, Mélanie. 2005. *La syntaxe comparée du Breton*. Université de Nantes. (Doctoral dissertation).

Kayne, Richard S. 2005. Antisymmetry and Japanese. In Richard S. Kayne (ed.), *Movement and silence*, 215–40. Oxford: Oxford University Press.

Kikusawa, Ritsuko. 2002. *Proto Central Pacific ergativity: Its reconstruction and development in the Fijian, Rotuman and Polynesian languages*. Canberra: Pacific Linguistics.

Kinn, Kari. 2017. Null arguments in Old Norwegian: Interaction between pronouns and the functional categories of the clause. *Linguistic Variation* 17(2). 309–346. DOI: 10.1075/lv.15018.kin.

Koch, John T. 1991. On the prehistory of Brittonic syntax. In James Fife & Erich Poppe (eds.), *Studies in Brythonic word order*, 1–43. Amsterdam: John Benjamins.

Kortlandt, Frederik. 2007. *Italo-Celtic origins and prehistoric development of the Irish language*. Leiden: Editions Rodopi BV.

Krasukhin, Konstantin. 2017. Was Proto-Indo-European an active language? (towards the typology of PIE). *Voprosy Jazykoznanija* (2). Number: 2, 49–76. DOI: 10.31857/S0373658X0000966-1.

Kroch, Anthony, Ann Taylor & Donald Ringe. 2000. The Middle English verb-second constraint. *Textual parameters in older languages*. 353–392. DOI: 10 . 1075/cilt.195.17kro.

Kuhn, A. 1853. Αἰών, aevum, aivs. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen* 2(3H). 232–235.

Labov, William. 1981. Resolving the neogrammarian controversy. *Language* 57(2). 267–308. DOI: 10.2307/413692.

Le Roux, Pierre. 1957. *Le verbe Breton*. Rennes: J. Plihon.

Ledgeway, Adam. 2012. *From Latin to Romance*. Oxford: Oxford University Press.

Ledgeway, Adam & Ian Roberts (eds.). 2017. *The Cambridge handbook of historical syntax*. Cambridge: Cambridge University Press. DOI: 10.1017/9781107279070.

Ledgeway, Adam & Nigel Vincent. 2025. Latin and Romance: Language contact from a typological perspective. In U. Hoinkes, J. Müller-Lancé & R. Schöntag (eds.), *Manual of historical contact linguistics in romance*. Berlin: de Gruyter.

Lehmann, Winfred P. 1974. *Proto-Indo-European syntax*. Austin: University of Texas Press.

Lehmann, Winfred P. 1993. *Theoretical bases of Indo-European linguistics*. London: Routledge.

Lewis, Henry. 1956. *Llawlyfr Cernyweg Canol*. Cardiff: Cardiff University Press.

Lewis, Henry & Holger Pedersen. 1961. *A concise comparative Celtic grammar*. Göttingen: Vandenhoeck & Ruprecht.

Lewis, Henry & J. R. F. Piette. 1966. *Llawlyfr Llydaweg Canol*. Cardiff: University of Wales Press.

Lightfoot, David W. 1980. On reconstructing a proto-syntax. In P. Ramat (ed.), *Linguistic reconstruction and Indo-European syntax*, 27–45. Amsterdam: John Benjamins.

Lightfoot, David W. 1999. *The development of language: Acquisition, change, and evolution*. Oxford: Blackwell.

Lightfoot, David W. 2002. Myths and the prehistory of grammars. *Journal of Linguistics* 38(1). 113–136. DOI: 10.1017/S0022226701001268.

List, Johann-Mattis, Michael Cysouw & R. Forkel. 2016. Concepticon: A resource for the linking of concept lists. In *Proceedings of the tenth international conference on language resources and evaluation LREC16*, 2393–2400. Marseille: LREC Proceedings.

List, Johann-Mattis, Simon J. Greenhill & Russell D. Gray. 2017. The potential of automatic word comparison for historical linguistics. *PLOS ONE* 12(1). e0170046. DOI: 10.1371/journal.pone.0170046.

Longobardi, Giuseppe. 2017. Principles, parameters and schemata: A radically underspecified UG. *Linguistic Analysis* 41. 517–557.

Longobardi, Giuseppe & Cristina Guardiano. 2009. Evidence for syntax as a signal of historical relatedness. *Lingua* 119(11). 1679–1706. DOI: 10.1016/j.lingua.2008.09.012.

Longobardi, Giuseppe, Cristina Guardiano, Giuseppina Silvestri, Alessio Boattini & Andrea Ceolin. 2013. Toward a syntactic phylogeny of modern Indo-European languages. *Journal of Historical Linguistics* 3(1). 122–152. DOI: 10.1075/jhl.3.1.07lon.

Longobardi, Giuseppe & Alessandro Treves. 2023. Grammatical parameters from a gene-like code to self-organizing attractors. *arXiv preprint arXiv:2307.03152*. DOI: https://doi.org/10.48550/arXiv.2307.03152.

MacCana, Proinsias. 1991. Further notes on constituent order in Welsh. In James Fife & Erich Poppe (eds.), *Studies in Brythonic word order*, 45–80. Amsterdam: John Benjamins.

McCloskey, James. 1990. Resumptive pronouns, A'-binding, and levels of representation in Irish. In Randall Hendrick (ed.), *The syntax of the modern celtic languages*. 199-248. DOI: 10.1163/9789004373228_008.

McCloskey, James. 1996. On the scope of verb movement in Irish. *Natural Language and Linguistic Theory* 14(1). 47–104.

McMahon, April & Robert McMahon. 2005. *Language Classification by Numbers*. Oxford: Oxford University Press.

Meelen, Marieke. 2016. *Why Jesus and Job spoke bad Welsh: The origin and distribution of V2 orders in Middle Welsh*. Utrecht: LOT.

Meelen, Marieke. 2017. Object-initial word order in Middle Welsh narrative prose. In Erich Poppe, Karin Stüber & Paul Widmer (eds.), *Referential properties and their impact on the syntax of Insular Celtic languages*, 145–178. Münster: Nodus Publikationen.

Meelen, Marieke, Nathan W. Hill & Hannes Fellner. 2022. What are cognates? *Papers in Historical Phonology* 7. 44–80. DOI: 10.2218/pihph.7.2022.7405.

Meelen, Marieke & Silva Nurmio. 2020. Adjectival agreement in Middle and early Modern Welsh native and translated prose. *Journal of Celtic Linguistics* 21(1). 1–28. DOI: 10.16922/jcl.21.2.

Meid, Wolfgang. 1968. *Indogermanisch und keltisch*. Innsbruck: Sprachwissenschaftliches Institut der Universität Innsbruck.

Meillet, Antoine. 1954. *La méthode comparative en linguistique historique*. 6th edn. Paris: Champion.

Miller, D. Gary. 1975. Indo-European: VSO, SOV, SVO, or all three? *Lingua* 37. 31–52. DOI: 10.1016/0024-3841(75)90003-0.

Morrison, David A. 2016. Genealogies: Pedigrees and phylogenies are reticulating networks not just divergent trees. *Evolutionary Biology* 43. 456–473. DOI: 10.1007/s11692-016-9376-5.

Neureiter, Nico, Peter Ranacher, Nour Efrat-Kowalsky, Gereon Kaiping, Robert Weibel, Paul Widmer & Remco Bouckaert. 2022. Detecting contact in language trees: A Bayesian phylogenetic model with horizontal transfer. *Humanities and Social Sciences Communications* 9(1). 1–14. DOI: 10.1057/s41599-022-01211-7.

Newmeyer, Frederick J. 2005. Against a parameter-setting approach to typological variation. *Linguistic variation yearbook* 4(1). 181–234. DOI: 10.1075/livy.4.06new.

Newton, Glenda Elizabeth. 2006. *The development and loss of the Old Irish double system of verbal inflection*. University of Cambridge. (Doctoral dissertation).

Nichols, Johanna. 1992. *Linguistic diversity in space and Time*. Chicago: University of Chicago Press.

Niyogi, Partha & Robert C. Berwick. 2009. The proper treatment of language acquisition and change in a population setting. *Proceedings of the National Academy of Sciences* 106(25). 10124–10129. DOI: 10.1073/pnas.0903993106.

Nolan, Brian. 2018. The encoding of negation in Modern Irish: Negation at the layered structure of the clause and noun phrase. In Rolf Kailuweit, Lisann Künkel & Eva Staudiger (eds.), *Applying and expanding role and reference grammar*, 45–70. Freiburg: Freiburg Institute for Advanced Studies.

Ohala, John J. 1981. The listener as a source of sound change. In Carrie S. Masek, Roberta A. Hendrick & Mary Frances Miller (eds.), *Papers from the parasession on language and behavior*, 178–203. Chicago: Chicago Linguistics Society.

Osthoff, Hermann & Karl Brugmann. 1878. *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*, vol. 1. Leipzig: S. Hirzel.

Ouhalla, Jamal. 1993. Subject-extraction, negation and the anti-agreement effect. *Natural Language and Linguistic Theory* 11(3). 477–518. DOI: 10.1007/bf00993167.

Pacchiarotti, Sara. 2020. On the origins of the ergative marker *wã* in the Viceitic language of the Chibchan family. In Jóhanna Barðdal, Spike Gildea & Eugenio Lujan (eds.), *Reconstructing syntax*, 241–288. Leiden: Brill. DOI: 10.1163/9789004392007_007.

Padel, Oliver J. 2021. English influence on the Middle Cornish verbal system. *Zeitschrift für celtische Philologie* 86(1). 255–296. DOI: 10.1515/zcph-2021-0011.

Panagiotidis, Phoevos. 2015. *Categorial features*. Cambridge: Cambridge University Press.

Pedersen, Holger. 1907. Neues und nachträgliches. *Zeitschrift für vergleichende Sprachforschung* 40. 129–217.

Pellard, Thomas, Robin Ryder & Guillaume Jacques. Forthcoming. The family tree model. In Adam Ledgeway, Edith Aldridge, Anne Breitbarth, Katalin Kiss, Joseph Salmons & Alexandra Simonenko (eds.), *Wiley Blackwell companion to diachronic linguistics*. Oxford: Wiley Blackwell.

Pires, Acrisio & Sarah G. Thomason. 2008. How much syntactic reconstruction is possible? In Gisella Ferraresi & Maria Goldbach (eds.), *Principles of syntactic reconstruction*. Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.04pir.

Poppe, Erich. 1995. Negation in Welsh and 'Jespersen's Cycle'. *Journal of Celtic Linguistics* 4. 99–107.

Press, Ian. 1986. *A grammar of modern Breton*. Berlin & New York: Mouton de Gruyter. DOI: 10.1515/84975.

Press, Ian & Herve Ar Bihan. 2004. *Colloquial Breton: The complete course for beginners*. London: Routledge.

Ram-Prasad, Krishnan & Marieke Meelen. 2022. *Syntactic parameters and the linguistic geography of Indo-European*. Paper presented at the Linguistics Association of Great Britain Annual Meeting.

Rama, Taraka. 2018. Three tree priors and five datasets: A study of Indo-European phylogenetics. *Language Dynamics and Change* 8(2). 182–218. DOI: 10.1163/22105832-00802005.

Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100(1). 59–129. DOI: 10.1111/1467-968X.00091.

Roberts, Ian. 1997. Restructuring, head movement, and locality. *Linguistic Inquiry* 28. 423–460.

Roberts, Ian. 1998. Review of historical syntax in cross-linguistic perspective. *Romance Philology* 51(1995). 363–370. DOI: 10.1017/s0022226798267119.

Roberts, Ian. 2007. *Diachronic syntax*. Oxford: Oxford University Press.

Roberts, Ian. 2010. *Agreement and head movement: Clitics, incorporation, and defective goals*. Cambridge: MIT Press.

Roberts, Ian. 2019. *Parameter hierarchies and Universal Grammar*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198804635.001.0001.

Roberts, Ian, James Baker, Marieke Meelen, Krishanan Ram-Prasad & Elena Isolani. 2022. Extending parametric comparison: Some preliminary results in older Indo-European languages. In *Diachronic generative syntax*.

Roberts, Ian & Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.

Ross, Malcolm. 2015. The argument indexing of early Austronesian verbs. In *Historical linguistics 2013: Selected papers from the 21st International Conference on Historical Linguistics, Oslo, 5-9 august 2013*, vol. 334, 257–79. Amsterdam: John Benjamins.

Rumsey, Alan. 1987. The chimera of Proto-Indo-European ergativity: Lessons for historical syntax. *Lingua* 71. 297–318.

Schleicher, August. 1853. 'Die ersten Spaltungen des indogermanischen Urvolkes. *Allgemeine Monatsschrift für Wissenschaft und Literature* 1853. 786–7.

Schrijver, Peter. 1995. *Studies in British Celtic historical phonology*, vol. 5. Rodopi.

Schrijver, Peter. 1997. *Studies in the history of Celtic pronouns and particles*. Dublin: Department of Old Irish, National University of Ireland.

Seržant, Ilja A. 2015. An approach to syntactic reconstruction. In Carlotta Viti (ed.), *Perspectives on historical syntax*, 117–154. Amsterdam: John Benjamins. DOI: 10.1075/slcs.169.05ser.

Sheehan, Michelle. 2014. Portuguese, Russian and the theory of control. In Hsin-Lun Huang, Ethan Poole & Amanda Rysling (eds.), *Proceedings of the 43rd annual meeting of the North East Linguistic Society*, vol. 43(2), 115–26.

Shlonsky, Ur. 1992. Resumptive pronouns as a last resort. *Linguistic Inquiry* 23. 443–68.

Sims-Williams, Patrick. 2003. *The Celtic inscriptions of Britain: Phonology and chronology, c. 400-1200*. Boston: Blackwell.

Smith, Jennifer. 2000. *Synchrony and diachrony in the evolution of English: Evidence from Scotland*. University of York. (Doctoral dissertation).

Smitherman, Thomas. 2015. Reconstructing non-canonical argument structure for Proto-Indo-European: Methodological questions and progress. In Carlotta Viti (ed.), *Studies in language companion series*, vol. 169, 93–116. Amsterdam: John Benjamins. DOI: 10.1075/slcs.169.04smi.

Stump, Gregory T. 1984. Agreement vs incorporation in Breton. *Natural Language and Linguistic Theory*. 289–348. DOI: 10.1007/BF00133790.

Sutrop, Urmas. 2012. Estonian traces in the tree of life concept and in the language family tree theory. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 3(1). 297–326. DOI: 10.12697/jeful.2012.3.1.15.

Szemerényi, Oswald. 1996. *Introduction to Indo-European linguistics*. Oxford: Clarendon Press.

Thomason, Sarah G. 2007. Language contact and deliberate change. *Journal of language contact* 1(1). 41–62. DOI: 10.1163/000000007792548387.

Thurneysen, Rudolf. 1885. Oskische einschubsvocale. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 27(2. H). 181–182.

Thurneysen, Rudolf. 1927. Altirische grammatik. *Indogermanische Forschungen* 44(s1). 370–372.

Toman, Jindrich. 1998. A discussion of resumptives in colloquial Czech. In Z. Boskovic (ed.), *Formal approaches to slavic linguistics*, vol. 6, 303–318. Ann Arbor: Michigan Slavic Publications.

Toorians, Lauran. 2014. *Towards a grammar of Middle Cornish*. http : / / laurantoorians.com/?page_id=128 (5 June, 2025).

Trask, R. L. 1996. *Historical linguistics*. London: Arnold.

Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. First edition (Oxford linguistics 6). Oxford: Oxford University Press.

Uhlenbeck, C. Cornelis. 1901. Agens und Patiens im Kasussystem der indogermanischen Sprachen. *Indogermanische Forschungen* 12. 170–171.

Vaillant, André. 1936. L'ergatif indo-européen. *Bulletin de la Société de Linguistique de Paris* 37. 93–108.

Van der Wal, Jenneke. 2020. From microparameters to macroparameters: A Bantu case study. In Theresa Biberauer, Sten Vikner, Andras Bárány & Jamie Douglas (eds.), *Syntactic architecture and its consequences I: Syntax inside the grammar*. Language Science Press. DOI: 10.5281/zenodo.4041229.

Van Gelderen, Elly. 2011. *The linguistic cycle: Language change and the language faculty*. Oxford: Oxford University Press.

Van Gelderen, Elly. 2022. *Third factors in language variation and change*. Cambridge: Cambridge University Press.

Villar, Francisco. 1983. *Ergatividad, acusatividad y género en la familia lingüística indoeuropea*. Salamanca: Universidad de Salamanca.

Viti, Carlotta. 2015. Historical syntax: Problems, materials, methods, hypotheses. In Carlotta Viti (ed.), *Perspectives on historical syntax*, 1–32. Amsterdam: John Benjamins. DOI: 10.1075/slcs.169.01vit.

Von Mengden, Ferdinand. 2008. Reconstructing complex structures: A typological perspective. In Gisella Ferraresi & Maria Goldbach (eds.), *Principles of syntactic reconstruction*. Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.06men.

Wackernagel, Jacob. 1892. Über ein Gesetz der Indogermanischen Wortstellung. *Indogermanische Forschungen* 1. 333–436.
Walkden, George. 2009. *The comparative method in syntactic reconstruction*. University of Cambridge. (MA thesis).
Walkden, George. 2011. Abduction or inertia? The logic of syntactic change. In Chris Cummins, Chi-Hé Elder, Thomas Godard, Morgan Macleod, Elaine Schmidt & George Walkden (eds.), *Proceedings of the Sixth Cambridge Postgraduate Conference in Language Research*, 230–239. Cambridge Institute of Language Research.
Walkden, George. 2014. *Syntactic reconstruction and Proto-Germanic*. Oxford: Oxford University Press.
Warner, Anthony R. 1993. *English auxiliaries: Structure and history*. Cambridge: Cambridge University Press.
Watkins, Calvert. 1964. Preliminaries to the reconstruction of Indo-European sentence structure. In Horace G. Lunt (ed.), *Proceedings of the 9th International Congress of Linguists*, 1035–1045. The Hague: Mouton.
Watkins, Calvert. 1966. The origin of the f-future. *Ériu* 20. 67–81.
Watkins, Calvert. 1976. Towards Proto-Indo-European syntax: Problems and pseudo-problems. In Stanford Steever, Carol Walker & Salikoko Mufwene (eds.), *Papers from the parasession on diachronic syntax*, 306–326. Chicago: Chicago Linguistic Society.
Watkins, Calvert. 1995. *How to kill a dragon: Aspects of Indo-European poetics*. Oxford: Oxford University Press.
Weiss, Michael. 2022. Italo-Celtic. In Thomas Olander (ed.), *The Indo-European language family*, 102–113. Cambridge: Cambridge University Press. DOI: 10.1017/9781108758666.007.
Weiss, Michael. Forthcoming. Italo-Celtic preverbs and adpositions. In Joseph F. Eska, Silva Nurmio, Peadar Ó Muircheartaigh & Paul Russell (eds.), *The Palgrave handbook of Celtic languages and linguistics*. London: Palgrave Macmillan.
Williams, Nicholas J. A. 2011. Middle and late Cornish. In Elmar Ternes (ed.), *Brythonic Celtic - Britannisches Keltisch: From Medieval British to Modern Breton*, 237–357. Bremen: Hempen.
Willis, David. 1998. *Syntactic change in Welsh: A study of the loss of the verb-second*. Oxford: Oxford University Press.
Willis, David. 2000. On the distribution of resumptive pronouns and *wh*-trace in Welsh. *Journal of Linguistics* 36(3). 531–573.
Willis, David. 2006. Negation in Middle Welsh. *Studia Celtica* 40(1). 63–88.

Willis, David. 2011. Reconstructing last week's weather: Syntactic reconstruction and Brythonic free relatives. *Journal of Linguistics* 47(2). 407–446. DOI: 10.1017/S0022226710000381.

Willis, David. 2013. The history of negation in the Brythonic Celtic languages. In David Willis, Christopher Lucas & Anne Breitbarth (eds.), *The history of negation in the languages of Europe and the Mediterranean, volume 1: Case studies*, 239–298. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199602537.003.0007.

Willis, David. Forthcoming. Comparative Brittonic syntax. In Joseph F. Eska, Silva Nurmio, Peadar Ó Muircheartaigh & Paul Russell (eds.), *The Palgrave handbook of Celtic languages and linguistics*. London: Palgrave Macmillan.

Wmffre, Iwan. 1998. *Central Breton*. Munich: Lincom.