The Regularity of Syntactic Change

Mark Hale & Madelyn Kissock
Concordia University, Montréal

Abstract

Contemporary syntactic theory, with its exclusive focus on functional features of lexical items as the sources of (syntactic) variation and change, requires us to review the questions traditionally posed regarding observed diachronic differences in the syntax of natural languages. This paper discusses 'regular' syntactic change and the different strategies required to account for such change under current theoretical assumptions. A comparison with phonological change suggests that it and syntactic change have more in common than is often realized. Based on the analysis developed, the paper goes on to discuss the appropriate methodology for syntactic reconstruction (SR), particularly in light of recent work such as Walkden (2014), as well as whether or not SR, as is often claimed, is qualitatively different than phonological reconstruction (PR), or, indeed, even possible.

X.1 Introduction

The advent of Minimalism, with its premise that the syntactic computational system is invariant across individuals and output variation is entirely a function of lexical (functional) feature input, saw the loss of parameter-setting as an explanatory mechanism.[1] Earlier parameters were tied to varying synchronic behaviors such as WH movement vs. *in situ* WH, *pro*-drop, headedness, and others. These parameters targeted sets of Lexical Items (LI's) that shared some specific feature(s) (e.g., +WH) and 'set' particular types of syntactic behavior such as movement for that set of LI's. Without parameters, an alternative explanation, based solely on syntactic features, is necessary. For diachronic purposes, it is exactly in cases such as these (WH et al.) that we see a 'regularity' of change parallel to that in the phonological domain. In the phonology, regular sound change affects all and only lexical items that share a segment with a specific bundle of phonological features — a [p], for example, in some particular context. Similarly, where we see a diachronic difference in WH behavior (say from fronted WH to *in situ*, as in some Indic I-E languages), this difference is manifested in multiple WH words (all LIs that have a +WH feature, say). We know of no documented cases which indicate that, for example, 'what' and 'when' go from *in situ* to fronted but 'who' and 'how' remain *in situ*.[2] At first glance, the sound change and syntactic change cases appear to be very parallel — both are based on changes targeting specific features of LIs and both are regular (i.e., all LI's with that feature(s) behave in the same way). However, parameters have never been appealed to to explain sound change, in spite of the fact that they have been critical in capturing 'group' behaviors both synchronically and diachronically in syntax.[3]

[1] We note that there is, in fact, considerable division on this topic with a number of people arguing for the retention of parameters in some form (Roberts 2012; Holmberg and Roberts 2014, inter alia) as well as those against any 'reintroduction' of parameters in any form (Newmeyer 2004, Boeckx 2014).

[2] To the extent we may find differences in the behavior of distinct WH-elements, 'why' vs. 'who', for example (see Ko 2006), the obvious minimalist lesson would be that +WH needs to be replaced by a more fine-grained set of features. 'Change' would still be predicted to be regular for lexemes which bear one of these new features.

[3] We focus narrowly on the changes themselves here rather than the diffusion of such changes throughout the population, the latter obviously driven by a different set of causal factors.

In a framework where parameter-setting is possible, consistent shared diachronic behavior of a set of LI's is accounted for by a change event that sets a parameter to a different value than the source grammar's value. In a framework like the current one, whose sole mechanism is the acquisition of feature sets for individual LI's, capturing similar behaviors across multiple LI's must be accomplished in some other way.[4] For current purposes, we will limit our discussion to ways in which 'regularities' might arise in the initial acquisition of an L1, assuming that this is a primary locus of change in the domains of both phonology and syntax.

The paper is organized as follows. Section 2 presents our general theoretical and working assumptions. Section 3 discusses the similarities and differences between syntactic and phonological systems and whether they have an immediate bearing on change in either case. Section 4 presents three aspects of the lexicon that interact crucially to produce the observed patterns of acquisition/change. Section 5 discusses the implications of our proposal for methods of syntactic reconstruction vis á vis other proposals for that process in the literature. Section 6 addresses the issue of syntactic correspondence in reconstruction. Section 7 offers some conclusions.

X.2 Background assumptions

Our analysis and proposal are made following a specific set of assumptions which are outlined in this section. We also give, as explicitly as possible, both the definition and scope of the 'regular' change that we will be examining.

- All features in the set of UG features (phonological, morpho-syntactic, and semantic) are available to be assigned to LI's by the acquirer at birth;[5]
- The initial featural representations that the acquirer assigns to LI's are based solely upon deduced syntactic or phonological (UG available) features of the input item without regard for aggregate lexical data, typological distributions, 'inventory' symmetry, or other 'meta-analysis' devices being employed by the acquirer.[6]
- The regularity of sound change, in the Neogrammarian sense, often represents observed, shared behavior in some population that differs from the observed shared behavior (of the relevant item) in an earlier generation of that population. A variety of mechanisms can come together to create the observed effect, only some of which have their origin in the acquisition process. Nothing in this population-based regularity entails that, for example, in the origin of some common sound change such as [θ] to [f], one or more acquirers set up the features of [f] for *every* instance of [θ]

[4]We believe that taking this very strict position — that variation and change are solely the result of differing feature sets on LIs, not parameters — as the starting point is in keeping with both minimalist assumptions and scientific principles. Note that if syntactic change can be accounted for solely with features (and without parameters), then synchronic variation falls out from that with no further machinery necessary.

[5]Note that, if we allow features to 'come online' during the course of maturation, then 'relearning' will need to be initiated as these late features come on line (since lexical items would have been stored without the use of such features). This 'relearning' stage would demand the re-initialization of lexical learning (since the acquirer can't know which of the lexical items s/he has posited to that date are 'missing' critical features). In essence, the last 'relearning' stage would represent the true onset of acquisition (earlier learning having been discarded). Thus, even if our assumption that all features are available at birth were to turn out to be incorrect, it is conceptually necessary that all features be available at the onset of adult-state relevant acquisition.

[6]Such 'meta-analysis' devices would yield useless results if applied continuously, since, over small lexical inventories, they would trigger a wealth of false generalizations, which would have to be discarded later in the learning process.

that they got as input. That is, the 'Neogrammarian hypothesis' is not concerned with regularity *at the level of individual grammars*.[7]

- A more formal, acquisition-based notion of regular change relates only to an individual's grammar/lexicon. Specifically, this notion of regularity refers to an observation that differences between the acquirer's grammar (from the data source's grammar) can manifest themselves in what are usually taken to be *classes* of lexical items and not simply individual LIs. This is the type of regularity that parameters addressed, as noted above, and the primary focus of this paper.[8]

- Lexicon Optimization is economizing.

- Lexicon Optimization, in the requisite sense, targets all features of LIs — phonological and syntactic.[9]

- Lexicon Optimization is a multi-step process, each optimization product serving as the input for the next optimization step.[10]

- Finally, our expectation is that the differences in feature representations that an acquirer sets up for an LI vis-à-vis the features of the corresponding LI in the lexicon of a data provider will be minimal. This follows from the overall 'success' of acquisition in replicating the data. Moreover, evidence from syntactic change supports the assumption that change is minimal (discussed at greater length below).

Most of these assumptions are familiar and many are shared by some subset of researchers in syntax and phonology.

X.3 Syntax vs. phonology

As noted earlier, change in both syntax and phonology is assumed to be a change in featural representation, where the acquirer posits some different (syntactic or phonological) feature for an LI than the source lexicon's representation of that LI. However, since different properties are often attributed

[7]This in spite of the Neogrammarian's own assertion that "die sprache kein ding ist, das ausser und über den menschen steht und ein leben für sich führt, sondern nur in individuum ihre wahre existenz hat, and dass somit alle veränderungen im sprachleben nur von den sprechenden individuen ausgehen können" ["language is not an object which exists outside of and above humans and leads a life for its own purposes, but rather only has its true existence in an individual, and that, therefore, all changes in the life of a language can only arise from speaking individuals,"] (Osthoff & Brugman 1878:xii-xiii).

[8]There is no clear evidence that this type of regularity exists, by the way. We base this notion, too, on observed synchronic grammars without direct knowledge of the state of the input grammar/lexicon. For example, cross-linguistic data suggests that there is considerable variation in at least some of the phenomena for which parameter setting was adduced, such as pro-drop, since there are languages where pro-drop is limited to certain pronouns to the exclusion of others. We return to this point later on.

[9]As well as, presumably, semantic — given the observed resolution of over/under generalization in content features during acquisition.

[10]For computationally-relevant features of syntax and phonology, this process of Lexicon Optimization must have an end point, by virtual conceptual necessity. If it does not, then differences between adult grammars and those of a 18-month old child acquirer would represent 'change' in the historical linguistic sense, and would need to obey constraints on change. No one, to our knowledge, believes that they do represent 'change' in the relevant sense. Since, at *some* point the differences between the adult source and the acquirer's constructed grammar *does* represent 'change', we can identify that point as the 'end point' of grammatical acquisition. (Of course lexical items may be acquired beyond that point. More on this later.)

to each of these modules, theoretically or empirically, it is worth examining these differences at least briefly for potential effects on or sources of (regular) change.[11]

X.3.1 Features

Our present conception of syntactic features is significantly different than that of phonological features. Phonological features all have the same 'status' with respect to the phonological computational system, however we wish to describe that status. There are, for example, no phonological features which are present solely for the purpose of the computation but are uninterpretable/irrelevant to the interface (phonetic representation), at least in our current conception of phonology. Syntactic features, on the other hand, come in two varieties, both of which serve a function in the computation. *Interpretable* features are preserved through the computation and are present in its output to the interface systems. *Uninterpretable* features, on the other hand, are present on an LI solely to facilitate computation, with no effect on (the interpretation of) that LI (e.g., Case, uninterpretable features and the strong/weak distinction, all of which we assume must be featurally represented). Note that both types of features still need to be part of the lexical specification and therefore both can be 'changed' during acquisition.

X.3.2 Computational system properties

Another potential difference is the representational vs. derivational nature of phonological vs. syntactic computation.[12] This is a significant difference (if real) in theoretical apparatus, but that does not necessarily entail that it will have a significant effect on what is essentially a data-(i.e., feature)-driven process. The more critical aspect of both computational systems, for our purposes, is that under OT-type assumptions for phonology both it and the syntactic computational systems are typically assumed to be invariant across individuals. (Data-driven input results in reranking of constraints in an OT phonology, but the constraints themselves are universally present and their content unchanged.)

X.3.3 Lexicon and computation

The 'local' nature of phonological processes has often been contrasted with the 'long-distance' nature of syntactic relationships and operations. It is not clear to us that the notion of 'distance' (of LI's) in phonology is, in its essence, different from that of syntax *mutatis mutandis* regarding the notion 'locality'. In both syntax and phonology, features of distinct LI's may interact with one another in the course of a computation (cf. at morpheme boundaries, in sandhi environments, in syntactic agreement). It is the surface result of such interactions which is of particular interest in acquisition and change. Determining which features (from the UG set) to assign to an LI is a function of observation of properties of the surface string (the *only* data available to an acquirer). Using the sentence-initial position of a WH word as evidence that a WH feature on a specific C is strong is no different than using (by 'undoing') the co-articulation effects between two segments as evidence for the featural properties of each segment — that is, syntactic displacement is a virtual 'syntactic coarticulatory' effect. Both cases represent surface outcomes which arise from the interaction of two distinct elements, such that the acquirer must determine in which of the two elements one should posit the various relevant features. Thus some of the apparent properties of these surface products themselves are, in some sense, an epiphenomenal product

[11] This is not intended to be any kind of complete, detailed comparison, for which space is lacking here.

[12] This varies in OT-based theories, with some versions such as Stratal OT or Harmonic Serialism approaching a derivational framework at least to some extent. Similarly, opinions on this are not uniform in syntactic theory (OT syntax, *et al.*).

of the interface levels' processing.

As far as we can see, none of the above-described differences between phonology and syntax, in and of itself, is likely to have a direct effect on the presence or absence of regular change, nor necessarily on the mechanisms underlying such change. This leave properties particular to the lexicon and its development to examine more closely. In the next section, we look in some detail at three domains within the lexicon that seem most likely to be involved in change.

X.4 Three Aspects of the Lexicon

We turn now to three aspects of the lexicon that interact crucially in acquisition and which we therefore believe are instrumental in acquisition-based syntactic and phonological change. Each of these is independently necessary and has been posited elsewhere for specific reasons unrelated to their role in change events.

X.4.1 Lexicon Optimization

The process of lexicon optimization has been appealed to to motivate a variety of synchronic phonological phenomena (e.g., constraint ranking in OT). We outline here the specifics of what lexicon optimization must entail in order to account for both synchronic states, as well as for the possibility of 'failure' (change).

Optimization, in our model, is a complex process containing multiple, ordered sub-processes. The initial lexicon upon which optimization operates is the set of stored forms (LI's) with full feature representations (phonological, syntactic, semantic) drawn from the entire UG available set of such features. Crucially, this initial lexicon will show an 'overgeneration' of LI's due to slight variations in features (phonological, syntactic, semantic), for what is, in the adult input grammar, only a single LI. (This seems inescapable given diachronic differences in grammars where the acquirer must have set up at least some representations of a given LI that are close to, but different from, the representation of the data source.) The (mis)assignment in some specific individual instance(s), because of some acoustic irregularity, of an [a] for an [ə] or vice versa; of a WH feature for a Relative Pronoun feature or vice versa (because of any number of obscuring factors in the input string); or any of a multitude of cases that will cause a 'match' between an input and a stored LI to fail, will result in the storage of a new, but almost identical LI.

A necessary subsequent step, then, and the first subprocess of lexicon optimization, is to optimize across the set of stored LIs, to recognize and then reduce what are essentially duplicates. For phonological representations, this might entail considering a set of stored forms that would ultimately be reduced to a single form 'pet', all of which had slightly different semantic features assigned to them (pet dog; pet cat; pet bird) perhaps corresponding to slightly different phonetic realizations (aspiration on [p] or lack thereof due to some sandhi environment; released vs. unreleased [t], and so on.). The optimization process will reduce these to a single stored form, arguably with aspirated *p* and unreleased *t*.

The parallel for (morpho-)syntax would be multiple stored instances of a first person pronoun; one stored as 'inclusive' another as 'exclusive' another as 'plural' and another as 'dual,' possibly with corresponding slight differences in phonological features and reduced to a single LI (e.g., exclusive plural). We refer to this initial subprocess as 'single-item optimization' because it reduces multiple LI's to a single LI. Subsequent to that, a subprocess that we refer to as 'cross-cutting optimization' scans across LI's for shared, particular properties, e.g., feature bundles for /p^h/ /p/ /$p^=$/; all LI's with a [+WH] feature, all V-category LIs, and so on. This subprocess results in positing LIs and a constraint reranking which gives rise to allophony and allomorphy, among other things, and correspondingly results in differences

between underlying representations and phonetic output representations. Note that without regularizing single-item lexical items in the earlier subprocess, the cross-cutting process would not produce the correct (attested) results. A final sub-process identifies lexical exceptions and stores them appropriately.

Stages in Lexicon Building and Optimization

1. Assign features to LIs and store them (URs identical to SRs)
2. Single-item optimization for LIs that show identical features in at least one dimension and 'too close for chance' similarity on other dimensions (homophones survive this step). This step gives a single result for the initial segment of 'thanks', either [θ] or [f] — both present in at least one stored version of 'thanks' — and a single result for the vowel, either [ɛ] or [æ] — both occurring in some stored forms.
3. Cross-cutting optimization. Optimization across all LI's: all LI's with /p/ in their rep; all pronouns with phi features of 2sg in their rep; (and presumably all LI's with 'dog' features in their rep). This step produces URs that are distinct from computed SRs (allophony, morphological alternations, etc.). Identification of LIs that need to be marked with exceptional, irregular features (postposed 'galore' in English, for example) is done as a final step here.

The data for Step (3) is, of course, the output of Step (2). Change may result at either step when there is a shift in data distribution such that the optimization algorithms (which we assume are invariant) produce a different outcome for the acquirer than for the adult source. However, while the data distribution that feeds Step (2) is dependent upon a wide variety of (external) factors, the data for Step (3) is dependent only on the output of Step (2).

We take Step (2) — a first optimization procedure — to be the point at which most diachronic differences arise in both syntax and phonology. Crucially, though, this step targets only those sets of stored LI's that are too similar to one another to be explained by chance and can therefore be collapsed into a single LI. However, this scenario only illustrates how a featural difference in an individual LI might come about between one grammar and the next (by an acquirer setting up an LI that differed in some featural way from the adult stored form). Nothing in this step is predicted to target LI's that are significantly different from one another featurally (where reduction to a single LI is not possible) such as all WH words (for WH-movement) or all pronouns (for 'pro-drop'), etc. However, logic dictates that, if an acquirer's data is distributed in such a way that the outcome of Step (2) for a single LI is different than that of the source grammar, then it is highly likely that a significant number of LI's with the relevant properties will also share that difference (the source of data being *grosso modo* the same for all LI's). Based on this, the output of Step (2) will already be a preponderance of 'changed' forms, such as LI's with [f] in place of source [θ], for example, or null subject pronominals for several persons (earlier subject pronominals having been reanalyzed as agreement morphology).

Step (3) of the process is the point at which the same general algorithm is extended across the entire lexicon, reviewing sets of LI's with shared properties on a variety of dimensions. This step will reveal predictable distributions in phonological form, for example, resulting in a smaller inventory of feature bundles present in UR representations than is present in surface forms. Similarly, productive morphemes will be adduced as separate LIs with the effect of increasing the number of lexical entries. Morphological alternations (e.g., *a/an* in English) will be recognized, collapsing what were earlier separate entries.

Finally, to complete the lexical building process and approximate an adult lexicon, the acquirer must identify irregular forms as exceptional. It is not clear that this last process will apply to phonology as it does to morphology and (potentially) syntax.

X.4.2 Linearity in the lexicon

The traditional idea that linear order needs to be captured in the syntax has been replaced with theories that separate linearity from structure – that linearity is not a matter for the Narrow Syntax (NS) (Chomsky 1995) and that linearity should be accounted for by independent algorithms (e.g., the Linear Correspondence Algorith, LCA of Kayne 1994). At the same time, since the advent of minimalism, it has been proposed that all variation is the result of differences in lexicons (the Borer-Chomsky Conjecture, Baker 2008), the syntactic computational system being itself invariant across individuals. If the lexicon is the locus of surface variation that we observe in syntactic output, and linear order is part of that variation, then it follows that lexicon should be the first place to look as a source of variation in linear order. Strong support for this comes from the fact that indicators of adjacency and precedence (presumably via specific features) *already* are a necessary part of lexical representations. Phonological feature bundles in an LI are necessarily ordered, at least as intrinsic properties of the LI /tin/$\neq$/nit/, and phonological operations make reference to those orders. Lexical marking of linear order for larger components than phonological feature bundles is also already firmly instantiated in the storage of affixes and other bound morphemes. Furthermore, the parallels between what is necessary to mark such affixal morphology and what is necessary for linear syntactic order are striking. For example, the suffix *-ly* not only needs to be stored as left-edge deficient, but with specific features on that edge to accept only a merger with an Adj category (edge), and crucially not an N category edge. Like morphological concatenation, syntactic concatenation is both sensitive to specific edges (head-final or head-initial) *and* specific categories — *N merge with Adv. We do not claim here that the linearity-related features needed for morphology are necessarily identical to those needed for syntax (although that would be ideal) but the parallels are sriking. A complete exploration of how features are to be instantiated in the lexicon is beyond the scope of this paper.

X.4.3 Feature Deactivation

We appeal here to the now well-known phenomenon described in the experiments of Werker *et al.* 1981, Werker and Tees (1984) and many subsequent similar experiments by these and other researchers. The phenomenon in question is the apparent loss of discrimination ability of certain non-native sound contrasts at an early age (10-12 months). To summarize, at the earliest ages (4-6 mos.) children performed equally well in perception experiments whether the (sound) contrasts tested were present in the environment language or not (suggesting that all UG features were available initially). However, their performance declined dramatically with age, specifically on those contrasts not available in their language environment (typical loss age (10-12 mos)). Hale and Kissock (1997) discussed these and similar results along with the question of bilingual second language acquisition (which seemed precluded by the Werker and Tees findings). For our purposes here, the basic experimental results, and our reinterpretation of these in terms of phonological features, are the relevant issue. To explain, from a theoretical standpoint, what was at the root of the experimental results, we proposed that phonological features that were not 'active' (used in a feature bundle for an LI) by an acquirer within a certain time frame (most probably based on lexicon optmization) were actually 'de-activated' for the grammar that was under construction (crucially, not removed from the innate set of features, simply from that particular instan-

tiation of a grammar acquisition algorithm). We propose, in what follows, that a similar de-activation, for similar reasons, occurs with syntactic features.

X.4.4 Sources of observed, regular syntactic change

Before turning to the details of how these aspects of the lexicon interact to produce change (and variation), it is important, first, to set aside cases of only 'apparent' regularity, where the change in behavior of a group of LI's is not actually due to a change in any features of those LI's. We find such apparent regularity specifically where the strength of a single feature on a head determines whether or not there is overt movement. For example, movement (or lack thereof) of LI's with an inherent WH feature is, we assume, determined by the strength of a feature on a single LI (a specific interrogative C, in this case). While superficially it appears that there is a 'group behavior' phenomenon here (and parameter-setting treated it as such), in fact this is not the case. The diachronic difference, should one arise, will have been a difference solely in the strength of the feature on C between the acquirer-posited C-element lexeme and that of the earlier generation. Since many of the diachronic syntactic differences that we see involve movement and ordering, we believe that some amount of 'group behavior' can be explained by change in a syntactic feature on a single, (possibly null) head, as in the WH case. Crucially, in cases such as these, there was *no change* in the WH LIs themselves – they simply continue to have a +WH feature. Since cases such as these are simply changes to a feature on a single LI, they do not count as 'regular change' in the sense we are interested in here. A considerable number of so-called 'regular' syntactic changes are actually of this type.

There are several types of syntactic change that do appear to at least affect more than one LI in a particular category, unlike the case above. An example of one such type is pro-drop. The traditional pro-drop parameter focussed on null subjects of the type seen in Spanish or Italian, where all pronominal, non-emphatic subjects were null. More complete cross-linguistic investigation revealed, however, that there was considerable variation in where null pronominals occurred. They could be in different syntactic positions (objects as well as subjects), they could alternate with overt pronominals for a variety of different featural reasons (discourse antecedents *inter alia*), their occurrence was not determined by presence or absence of overt (verbal) agreement, and most critically, some languages limited null pronominals to certain persons (3rd, for example), so that *within a specific grammar* there was no uniformity across pronouns. The point here is that there is little evidence that null pronominals should be treated as a coherent group with respect to syntactic change. The cost-free analysis of such pronominal behavior is that null pronominals are acquired in exactly the same way all other null LI's are acquired (cp. the many other LI's with heavy functional load which are often null — C's, T's, etc.). For all null LI's, the acquirer must get indirect evidence for the features of the LI and assign those features — if the presence of those features in the representation does not correspond to any overt phonological material, the appropriate phonological entry will be empty. The varying behavior of null pronominals within and across languages can be attributed to the varying features assigned to each LI and the possible contexts in which such features can occur.

Another type of observed, regular change is a change of headedness for some particular syntactic category, such as V. Within a single grammar, values for headedness are observably not necessarily uniform *across* categories. The general, theoretical assumptions have been that for any one category in an individual's grammar, headedness is set as either initial or final. It should be noted, however, that observed surface strings in many languages do not correspond well with this assumption, showing considerable surface variation for any particular category and its complement (e.g., Sanskrit has both prepositions and

postpositions — the category 'P' is thus not head-uniform). It is, of course, exactly in such variable cases that we predict and find changes in headedness features. Not surprisingly, very strictly verb-final languages such as Telugu do not seem subject to mis-analysis by the acquirer insofar as V-headedness goes. For the following reasons, it is somewhat unclear whether the Telugu-type stability is system-driven or epiphenomenal. If every piece of evidence for every V is obviously head-final, we cannot distinguish between an acquirer giving each V a head-final feature on a V by V basis or whether the acquirer had some system which assigned head-finality to all V's. This will turn out to be a distinction without a difference. We support this view with evidence that (1) featural changes are minimal and (2) their effects are often difficult to detect in the surface string. We expect an LI that has 'changed' (been assigned a different feature by the acquirer) to behave in most ways like its 'original' source counterpart because most of its features will have stayed the same. Expletive negation in English, for example, illustrates this — all the features of NEG (which we assume was the source model) remain except the actual negation semantics. Syntactically and phonologically expletive NEG behaves like 'negative NEG' as shown below.

(1) And so didn't John (go and) walk down those stairs anyway?

From an interpretational standpoint, there is no negation in the above string (nor interrogative, as it happens). It is infelicitous to answer 'Yes' or 'No' and the correct interpretation is affirmative — John walked down the stairs. The structure and phonological form of the expletive negation however is surface identical to negation. We note that, under the empirically-supported assumption that featural changes are minimal, Lightfoot's (1979) 'radical reanalysis' problem appears to be no impediment to working on syntactic change (and reconstruction).

Difficulty in determining the effects of a change — critical since the acquirer has only the surface string, apart from properties of UG, to base their featural assignment on — is illustrated by the emergence of a quotative marker (QM) in Telugu. The marker, *ani* is homophonous with the absolutive participle form of the verb 'to say' (*anu*) and was apparently a reanalysis of that form (which is still perfectly productive as an absolutive). Like the English expletive negation case, QM *ani* appears in the surface string in the same position as an absolutive form of *anu* would. However, absolutive *ani* is productively formed from a root and suffix whereas QM *ani* is almost certainly stored as an unanalyzed whole.[13] We point out the underlying difference in morphological structure here not so much as a confound for an acquirer (though it certainly is not helpful) but as one of many confounds that occur when we observe data.

X4.5 Lexicon Building and Change

We return now to headedness as the strongest example of uniform change of LI's within a single category (all V's, all N's, and so on). Our proposal for this and similar types of 'global' change is straightforward. One way to model linearity features in the lexicon is to associate them with categories (just as affix edges must be associated with categories, as noted earlier). In this way, UG would provide two distinct N category features to the acquirer, one left and one right, just simplistically. The result of feature assignment (based on interpretation of the data) and Lexicon Optimization, will be that unless there is evidence in the data for both types of N (both left and right) or the result of Lexicon Optimization produces LI's with both types of N (via the first subprocess), the grammar will not maintain both as active. Once a UG feature is deactivated in this way, the acquirer no longer has it available as a choice in setting

[13]There are, of course, more significant syntactic differences between these two forms — the QM lacks any argument structure or predication ability, among a number of other differences.

up a new LI. We assume that this is a very early process in lexicon building.[14] The result of such a process for surface linearity facts is that, assuming most categories have only one active category-direction feature available, the acquirer must look to other mechanisms (other features on LI's) to produce surface orders that run counter to their only active headedness choice for that category. One can imagine that this type of cascade could produce the implicational results noted in Rizzi (1982) for 'that-trace' effects and null pronominals.

X.5 Reconstructing syntactic features

We next turn to a consideration of some implications for this conception of syntactic change for the widely discussed matter of the possibility of, and appropriate methodology for, syntactic reconstruction. We focus on the so-called 'Correspondence Problem' — i.e., what elements are to be included in the correspondence sets that are the basis for reconstruction (Lightfoot 2002a,b; Campbell and Harris 2002; Walkden 2014; among many others). The stumbling block for creating syntactic correspondence sets for many scholars seems to be a notion that the proper cognates would be 'sentences'. Two problematic issues arise for this conception. First, for many scholars it seems to be problematic that sentences, which are *built*, rather than *stored*, can form a sound foundation for establishing 'corresondences' from which reconstruction can be undertaken. Secondly, no coherent concept of 'cognate sentence' has been proposed.

Under the assumptions above, syntactic correspondence can be reduced to a relationship between the syntactic feature bundles of 'corresponding' lexical items, parallel to the phonological featural representations for 'corresponding' phonological forms. In short, lexical items are sets of feature bundles: morphosyntactic and phonological.[15] In tracing the history of a lexical item — something we have been successfully doing for some centuries now — we are placing the relevant feature bundles into correspondence.

Finally, with this focus on the features on individual lexical items, we address the question of whether there is any role for the 'sentence' in diachronic syntax, beyond whatever evidence it can provide for the features of LI's. We argue that 'sentences' may still play an indirect role, specifically one of *explaining* change, as context *is* likely to be a critical factor in such cases (as it is in phonological diachrony). These will be cases where syntactic cognates in a correspondence set do not all show the same set of features as each other, yet as cognates are assumed to descend from one proto-form. In such cases, we must seek an explanation for the difference (and decide which cognates are actually deviating, of course). Since the acquirer is exposed to the input forms *in their sentential context*, the first place to look for cues as to why the acquirer posited a different featural analysis will be here. Although we expect much of this to be a language-internal, we see no real block to comparing syntactic output forms/sentences across languages. A brief example is presented below.

Debates about the viability of syntactic reconstruction have revolved around whether or not one can employ directionality predictions, whether the nature of acquisition and change allow for a historical 'connection' (the Radical Reanalysis Problem of Lightfoot 1980), and what elements are to be included in the correspondence sets that are the basis for reconstruction. We concentrate here on the correspondence/cognate issue, as that seems foundational in a way the other topics are not. As Walkden (2014)

[14]Moreover, such a process is likely to be at the root of decision-making on the part of the acquirer as to whether they need to start the construction of a new grammar vs. incorporate LI's into their existing grammar (a distinction necessary if bilingual second language acquisition is to remain a possibility).

[15]We will not further consider 'semantic' features in what follows.

notes, historical phonologists do not compare just *any* [p]'s when determining what should go into correspondence sets. They examine sounds in *cognates* (already established as daughters of a common, etymological ancestor) and linearity (i.e., syntagmatic relationships into which a given [p] enters) is crucial. We note two things here: 1) that cognates do not, in fact, always have corresponding sounds – different and/or missing/additional sounds are extremely common and are the basis for hypothesizing historical changes; and 2) that correspondent sets are made up of *phonetic forms*, that is non-stored, output forms (see Hale 2007 for an explicit discussion).

As Walkden (2014) states, cognates in syntax (under contemporary assumptions) can be reduced to lexical items with similar syntactic feature bundles (still related etymologically), parallel to the featural representations for phonetic/phonological forms. What he fails to recognize is that the choice of phonetic forms for PR, with their very important contextual information, is driven by a need to reconstruct not only the URs of the proto-language but *also the rules/rankings for phonological computation*. Crucially, this same requirement is not there for the proto-language syntax, given that the syntactic computational component is assumed to be identical across individuals — we are not obliged to reconstruct syntactic rules/processes. But in phonology, we cannot simply compare purely UR/phonemic forms of the cognates in a correspondence set without missing aspects of the phonological computational system that derive *variation* in outputs from identical inputs. This is what makes the environment ('context' in Walkden's terms, to include sentential context) so important in PR and, essentially, much less relevant in SR, where the LIs with their syntactic features suffice. Note that this is not a difference in the *methodology* of reconstruction for phonology vs. syntax– it is not about *how* but about *how much*.

Let us be clear about what we think a 'sentence' is. We will assume that the syntax has no access to 'encyclopedic' content of lexical items (e.g., whatever makes 'duck' different from 'jacket' or 'chair') nor, of course, to their phonological content. Thus, assuming identity of all syntactically-relevant features (grammatical gender, grammatical animacy, etc.), 'the cats ate this bread' is *the same sentence* as 'those maruts drank the soma'. The two utterances generated from these sentences differ in their phonology (because of the phonological features of their lexical content, not visible to the syntax, and thus not part of the syntactic computation or output) and in some aspects of their interpretation (because of the encyclopedic semantics of the lexical items, fed into the conceptual interpretive system), but not in their syntax (nor in their syntactic representation).[16] Some such conception of what syntax is is required in any event, if we are to understand why there are processes such as the movement of relative pronouns (*wh*-movement) and auxiliaries (as in I>C movement for question formation), but no such thing as "*duck* movement" or "trisyllable fronting" in human languages. The easiest explanation for this cross-linguistic fact is that the syntax can identify *wh*-elements and auxiliaries (and thus can manipulate them), but can't identify 'ducks' or 'trisyllables' (and thus *cannot* manipulate them).

We conclude that, if our goal is to reconstruct the grammar and the lexicon of some proto-language, i.e., only information that is stored (including elements present by virtue of UG), then the reconstruction methodology for phonology is perfectly suitable for syntax, *mutatis mutandis*. The most salient difference between syntax and phonology in this realm (of reconstruction) is that, based on the assumptions above, PR requires reconstructing both the phonological features of lexical items *and* the rules (or rankings of constraints) whereas SR requires only the reconstruction of formal, syntactic features of lexical items, the computational system being identical across individuals. Finally, we warn against confusing

[16]The 'direct object' status of 'the bread' and 'the soma' is represented in the syntax, and thus, to the extent it feeds the interpretation, that aspect of the interpretation of the sentence is 'syntactic'. But the breadiness and soma-properties of these objects is not visible to the syntax.

our current (lack of) understanding of the (synchronic) features of syntax with a (successful) methodology that uses such features as its foundation.

X.6 'Syntactic' correspondence

The question that arises next is whether there is a limit to the kinds of linguistic object that can be brought into correspondence (in the relevant sense). We note here that in this sense is it *not correct* to say something like "Samoan *laumei* 'turtle' is cognate with Hawai'ian *honu* 'turtle'." Cognancy, in the relevant sense, is intended to cover cases of *phonological* relationship between semantically-related lexemes and is an important mechanism for establishing genetic relatedness — and *laumei* is not the *phonological* correspondent of **fonu*, but the lexical replacement thereof. 'Lexical replacement' is a diachronic change — changes, therefore, can be inferred not only from *cognancy*, but also from *correspondence* in this 'replacement' sense. This is the 'relevant' sense referred to above.

The syntactic question, therefore, is not necessarily about *cognate sentences*, whatever exactly that might refer to, because that is not the notion of 'correspondence' we necessarily need to develop a meaningful diachronic hypothesis, contra e.g. Lightfoot (2002a,b). It is quite normal in work on the syntactic history of languages during their attested stages to bring elements from distinct grammars into 'correspondence' in the relevant sense—indeed, it is impossible to do that kind of diachronic syntax without this notion (just as it is in morphology). So, if we wanted to talk about the diachronic syntax (into later stages of 'English') of an element such as Old English *sceal* 'shall', we would need to draw structures from our two (or more) stages of 'English' which stand in a meaningful correspondence relationship.

One can only imagine that this would include looking at structures which show the descendant of the object under study (*shall*-sentences, for example) as well as sentences which express meanings earlier expressed by *sceal* sentences, but now no longer involving the descendant of that morpheme (i.e., cases of lexical replacement, which are changes we must account for as well). To the degree the sentences culled by this procedure show the same syntax and a reflex of *sceal*, the syntax of *sceal* has not changed. To the extent they show the same syntax but no reflex of *sceal*, there has been some lexical change event. To the extent they show different syntax, and a reflex of *sceal*, the syntax of *sceal* has changed, and we need to explain the change event(s) involved. To the extent they show a different syntax and no reflex of *sceal*, you probably were confused in calling them 'correspondent'!

This seems to be the working procedure of everyone who does this kind of thing, so we apologize for spending so much time on it. The question we now face is this: does it make sense to call forms in sister languages (i.e., in grammars which descend from a common, unattested ancestor) 'corresponding' as we have done for grammars which we assume to stand in a descent relationship with one another. Here we are at a loss for why we would not. To do the work on the history of *sceal* within English we needed a procedure which licensed treating distinct structures drawn from distinct grammars as being 'in correspondence'. We face no challenge identifying other, let's say, West Germanic cognates of *sceal*. It is, as far as we can see, inconceivable that Proto-West Germanic had some proto-form of *sceal* (let's call it **skal-*) that had no syntax at all: therefore, descendant grammar uses of the daughter reflexes of that Proto-West-Germanic form represent developments of that proto-syntax. The syntax of these daughter forms must show a diachronic dependence on the syntax of the relevant item in Proto-West Germanic (or they are not daughter forms); therefore, we find **skal-* correspondences within such sentences.

Our reconstruction of the Proto-West Germanic syntax of **skal-* will be that reconstruction which provides the most plausible and simplest account of the historical syntax of the West Germanic daughter languages. Again, there is nothing unfamiliar in this requirement: our account of the history of the

'modals' in English takes a sampling of data from grammars at a variety of points in time and posits the most plausible and simplest account for the historical developments which link the data points. If we have no capacity for determining what is the 'most plausible and simplest' account for diachronic syntactic data, we can't do diachronic syntax at all. If we do have such a capacity, it can be applied just as coherently to competing theories of the trajectory from Proto-West Germanic to its daughters as it can to competing theories of Old English to its daughters. One hypothesizes a starting point (this is called 'reconstruction' in one case, 'syntactic analysis' of the anterior state in the other) and deduces from that starting point and the attested data point(s) a history.

Now the question reduces to this: can we recover the morphosyntactic properties of lexical items? To the extent we can, and to the extent those features determine the properties of the syntactic output generated by the (invariant) syntactic computational device, we will have reconstructed a system that *generates* proto-language sentences. That some morphosyntactic features are recoverable (i.e., can be reconstructed) is quite clear: that a given reconstructed object is an N, or a D, C, P, A, or V-element is part of any reconstruction. IEists reconstruct grammatical gender, person and number, as well as case (structural and lexical) on nominals, tense, mood, and subject agreement on verbs (including some peculiar aspects of that agreement, such as the failure for neuter plural nouns to trigger plural agreement), participial marking, and the like. Each of these elements was reconstructed from the belief that the structural configurations in which these morphological elements appeared in the various daughter languages stood in the type of *correspondence* relationship that made them appropriate comparanda. Without controlling for such correspondences, how are we to know that the thing we call an ablative in one language is to be compared to the thing we call an ablative in another for reconstruction purposes?

Support for the idea that morphosyntactic features can be reconstructed comes from the simple fact that existing reconstructions include, almost universally to our knowledge, detailed reconstruction of such properties. For example, examine the following PIE reconstructions.

Vedic Sanskrit	Greek (dor.)	PIE 'phonetic'	PIE 'phonemic'	
dyāw+s	zew+s	[*dyēws]	/*dyēw/ + /s/	'sky' (nom.sg.m.)
dyā+m	zē+n	[*dyēm]	/*dyew/ + /m/	'sky' (acc.sg.m.)
gāw+s	bow+s	[*g^{w}ōws]	/*g^{w}ōw/ + /s/	'cow' (nom.sg.f.)
gā+m	bō+n	[*g^{w}ōm]	/*g^{w}ow/ + /m/	'cow' (acc.sg.f.)

PIE Phonological Rules:

Consonant-Stem Root Nouns have 'lengthened grade' nominatives, full-grade accusatives

/w/ deletes with compensatory lengthening before word-final /m/

The outcome of the reconstruction sketched above, in PIE, is a set of morphemes. These morphemes have phonological properties and morphological properties, and via their reconstruction we have been able to recover some morpho-phonological and simple phonological processes of the proto-language. But they also have *morpho-syntactic features*, directly relevant to the generation of syntactic output representations of strings into which they entered. To say that **dyēws* was the nominative singular of 'sky' is to say that it entered into syntactic computation in such a way that it could be inserted into a tree where nominatives are inserted (not accusatives, or datives, or finite verbs), and that it could only be placed in a structure which was going to show, at the end of the relevant computation, singular subject agreement. To say that **dyēm*, by contrast, was the accusative singular of the word for 'sky' is to claim that it could enter into an output tree only in such a way that it surfaces in a position and with an interpretation appropriate for accusatives (and singulars). Thus the latter form, but not the former, could be inserted into the direct object position of a sentence built around a transitive verb (e.g., **speḱyeti* 'he, she it sees') and that the result would be a VP which meant 'sees the sky'.

Crucially, the reconstruction sketched above is *not* the reconstruction of the accusative singular of the word for 'sky': it is the reconstruction of a grammar and a lexicon which can *generate* the accusative singular of the word for 'sky'. The lexicon must contain the relevant *morphemes* (not output 'words') (e.g., **dyew-* 'sky', **-m* accusative singular, **-s* nominative singular), and a phonological computation system with the relevant rules or constraints so as to generate the appropriate surface outputs. The reconstruction of these lexical features was undertaken by bringing the relevant elements *into correspondence*. In precisely the same way, the morphosyntactic features posited for the reconstructed lexical items above require that we bring the relevant syntactic structures of the daughter languages *into correspondence*. The various reflexes of what we reconstruct as an accusative singular morpheme must find an explanation of their syntactic distribution from a proto-language function which was expressed by this morpheme — e.g., direct object of transitive verb, or, in the case of PIE, goal of motion. But if any Indo-Europeanist doubts that the accusative singular of the word for 'sky' could function as the direct object of a transitive verb, they have kept that view a secret from the field. As far as we know, an (often implicit) theory of the syntactic history of each and every closed-class morpheme posited by IEists (the nominative dual, the optative, the subjunctive, the 1st singular middle, the ablative plural, etc., etc., etc.) was a prerequisite for the reconstruction of these items: without a theory of appropriate syntactic correspondence, one could not know what should be compared between languages to allow for a reconstruction.

What this reveals is that the idea that one of the problems confronting syntactic reconstruction is that sentences are not *stored* (and thus cannot be brought into correspondence) is mistaken: 'words' are not *stored*, either, and yet historical linguists have been bringing these grammatically-generated objects into correspondence for over a hundred and fifty years.

X.7 Conclusion

The view that syntax operates over a UG-provided, well-defined set of morphosyntactic features, and that the operations performed over these features are universal and invariant, has direct implications for our understanding of the mechanisms of syntactic change and, therefore, of syntactic reconstruction. In particular, we have argued that the most salient traditionally-cited challenges to syntactic reconstruction — the 'radical reanalysis' problem and the 'correspondence' problem — appear to dissolve under an acquisition-based theory of change built around such a conception of syntax.

Baker, Mark. 2008. 'The Macroparameter in a Microparametric world, in Theresa Biberauer (ed.) *The Limits of Syntactic Variation* Amsterdam: Benjamins, pp. 351-374.

Boeckx, Cedric. 2014. 'What Principles and Parameters Got Wrong,' in Picallo, M. Carme (ed.) *Linguistic Variation in the Minimalist Framework* Oxford: Oxford University Press, pp 155-178.

Campbell, Lyle and Alice C. Harris. 2002. Syntactic reconstruction and demythologizing 'Myths and the prehistory of grammars'. *Journal of Linguistics* 38:599-618.

Cheng, Lisa Lai-Shen Cheng and Norbert Corver. 2006. *WH-Movement: Moving On*. Cambridge, MA: MIT Press.

Hale, Mark. 2012. Foundational concepts in the scientific study of sound change. Pp. 235-246 in *The Initiation of Sound Change*, Maria-Josep Solé and Daniel Recasens (eds.). Amsterdam: Benjamins.

_____. 2007. *Historical Linguistics: Theory and Method*, Oxford: Wiley-Blackwell.

———. 1998. Diachronic Syntax. *Syntax* 1.1:1-18.

———. 1987. *The Comparative Syntax of the Archaic Indo-Iranian Languages*. Ph.D. dissertation, Harvard University, Department of Linguistics.

Hale, Mark and Madelyn Kissock. 1997. 'Nonphonological triggers for renewed access to phonetic perception' in *GALA '97, Language Acquisition: Knowledge Representation and Processing,* Richard Shillcock, Antonella Sorace & Caroline Heycock (eds.). Edinburgh, UK: The University of Edinburgh.

Holmberg, Andreas and Ian Roberts. 2014. 'Parameters and the Three Factors of Language Design,' in Picallo, M.. (ed.) *Linguistic Variation in the Minimalist Framework* Oxford: Oxford University Press, pp 61-81.

Kayne, Richard. 1994. *The Antisymmetry of Syntax* Cambridge, MA: MIT Press.

Ko, Heejeong. 2006. 'On the structural height of reason, *Wh*-adverbials: acquisition and consequences' pp. 319-349 in Lisa Lai-Shen Cheng and Norbert Corver (2006).

Lightfoot, David. 2002b. More myths. *Journal of Linguistics* 38:619-626.

———. 2002a. Myths and the prehistory of grammars. *Journal of Linguistics* 38:113-136.

Newmeyer, Frederick. 2004. 'Against a parameter-setting approach to langauge variation,' in *Language Variation Yearbook* Vol 4. Amsterdam: Benjamins, pp. 181-234.

Osthoff, Hermann and Karl Brugman. 1878. *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*. Leipzig: Hirzel.

Rizzi, Luigi. 1982. *Issues in Italian Syntax.* Dordrecht: Foris.

Roberts, Ian. 2012. 'Macroparameters and Minimalism: A Programme for Comaprative REsearch,' pp. 320-335 in C. Galves. S. Cyrino, R. Lopes, F. Sandala, and J. Avelar (eds.) Parameter Theory and Linguistic Change. Oxford: Oxford University Press.

Walkden, George. 2014. *Syntactic Reconstruction and Proto-Germanic.* ()Oxford Studies in Diachronic and Historical Linguistics 12.) Oxford: Oxford University Press.

Werker, Janet F., John H. V. Gilbert, Keith Humphrey, and Richard C. Tees. 1981. Developmental aspects of cross-language speech perception. *Child Development* 52:349-355.

Werker, Janet F. and Richard C. Tees. 1984a. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7:49-63.