Watkins, Calvert. 1976. "Towards Proto-Indo-European Syntax: Problems and pseudo-problems". *CLS Parasession on Diachronic Syntax* ed. by Sanford B. Steever, Carol A. Walker & Salikoko S. Mufwene, 305–326. Chicago: Chicago Linguistic Society.

Watkins, Calvert. 1996. "Proto-Indo-European: Comparison and reconstruction". *The Indo-European Languages* ed. by Anna Giacalone Ramat & Paolo Ramat, 25–73. London: Routledge.

How much syntactic reconstruction is possible?*

Acrisio Pires & Sarah G. Thomason
University of Michigan, Ann Arbor

This paper explores ways to synthesize methods from generative linguistics and historical linguistics to develop explanatory criteria that need to be satisfied by different attempts to carry out syntactic reconstruction. It addresses various questions such as (i) the need to define exactly what it means to reconstruct *a language*; (ii) characterizing the formal entities that count as the basic elements in the analysis of the empirical evidence for syntactic reconstruction, and whether the units of analysis and the elements that need to be reconstructed are formal entities of the same kind; (iii) whether it is possible to establish general principles for reconstructing syntax; (iv) to which extent the methodology adopted for the reconstruction of other properties of a linguistic system can be applied successfully to the reconstruction of syntax; and (v) identifying methodological criteria to assess the success of a syntactic reconstruction, and devise empirical tests for the reconstruction model.

0.   Introduction

An important debate relating formal syntax and historical linguistics concerns the question whether it is possible to reconstruct the syntax of a proto-language by comparing the syntactic properties of two or more of its daughter languages.[1] This

1.   The term 'language' is used in this instance without qualification, an issue we come back to below. We put aside questions regarding the possibility of internal reconstruction, given its more general limitations that extend beyond syntactic properties.

paper addresses this question in detail, by focusing on several important empirical and methodological questions that we take to be crucial in this debate:

(1) a. The term 'language' has been used with many different, often imprecise meanings in the literature, and it is important to define exactly what it means when one refers to the attempt to reconstruct *a language*.

   b. Once one defines in a formal way what counts as a language for purposes of syntactic reconstruction, what are the formal entities that count as the basic elements in the analysis of the empirical evidence for syntactic reconstruction? Are the units of analysis and the elements that need to be reconstructed formal entities of the same kind?

   c. Once we define the basic formal entities that are relevant for any attempt to carry out a syntactic reconstruction, is it in fact possible to reconstruct any properties of syntax?

   d. Is it possible to establish any general principles for reconstructing syntax and, if so, how do these principles relate in general to formal properties of human language syntax?

   e. Can the methodology adopted for the reconstruction of other properties of a linguistic system be extended successfully to the reconstruction of syntax?

   f. Are there methodological criteria that enable us to assess the degree to which a syntactic reconstruction has been successful? If so, can one provide empirical tests for the reconstruction model, given these methodological criteria?

Some of these questions have been discussed in detail in the literature on syntactic reconstruction, but we think that there have not been widely accepted answers to them. For instance, a recent (and in our view unsettled) debate has focused on the general question whether syntactic reconstruction is possible at all, addressing especially questions (1b) and (1e) (Harris & Campbell 1995, henceforth H&C; Lightfoot 2002a,b, Campbell & Harris 2002, henceforth C&H). Whereas Harris and Campbell are very optimistic about the prospects for syntactic reconstruction, Lightfoot has long expressed strong reservations about attempts to reconstruct syntax, pointing out methodological hurdles that must be overcome in any successful model of syntactic reconstruction (see also Lightfoot 1979; Watkins 1976). Lightfoot (1979) presents a detailed criticism of attempts to reconstruct word order in PIE (as does Watkins 1976), and he argues that "caution is in order in undertaking any work in syntactic reconstruction by simply applying to syntax the traditional methods of phonological reconstruction" (1979: 155).

Longobardi (2001, 2003) adopts a methodological perspective similar to Lightfoot's, in that he uses the Principles & Parameters approach to syntactic theory (see e.g. Chomsky 1986; Haegeman 1994; Webelhuth 1995), but he is more optimistic about the prospects for syntactic reconstruction, addressing issues

related to questions (1c) and (1d) above. Regarding question (1d), Longobardi proposes that his approach, which he claims is simply the adoption of the Principles & Parameters approach to syntax, can overcome the well-known deficiencies in typology-based approaches to reconstruction. His approach, however, differs sharply from that of other generativists who have discussed these issues, including Lightfoot; it also diverges radically from standard methods in historical linguistics more generally. This is especially evident when he proposes to use parametric variation – specifically, as determined through counts of shared parametric values among proposed sister languages – as a means of establishing distant genetic relationships among languages (Longobardi 2003). He does not explain how this essentially typological measure can overcome the deficiencies in other typology-based proposals for discovering genetic relationships. Here he does not argue directly for the use of his technique for reconstruction, but he implies that this approach can reconstruct the parameter settings for the parent of a set of related languages.[2] We do not find his specific approach promising, for the same basic reason that typological approaches to genetic linguistic classification have failed in the past: languages whose development is historically closely related (as far as we can tell by standard methods) can and do share numerous typological features, but when a partly shared history is in fact responsible for shared typological features, distinguishing inherited from diffused features is no easy task. In

---

2. Longobardi tells us (p.c. 2005) that his effort to reconstruct 'phylogenetic relations among languages' does not need to involve the reconstruction of any syntactic structures in the parent language of a proposed family. We understand his reservation here, and that he intends his use of the term 'reconstruction of phylogenetic relations' to exclude actual historical comparative reconstruction. But to a historical linguist, any claim that certain languages are genetically related, based on the evidence of shared syntactic properties, does indeed imply a measure of syntactic reconstruction (even aside from the fact that 'reconstruction' is only used in historical linguistics to refer to the construction of hypotheses about features of proto-language structure). One reason is methodological: a convincing demonstration of genetic relationship always includes showing that reconstruction is possible – and the only sure way of showing this is to do some reconstruction. The other reason is empirical: if the shared syntactic structures that Longobardi and his collaborators adduce as evidence for genetic relationship are *not* supposed to be shared inheritances from a common parent language, then what could they possibly be? It's true that drift (linguistic change in response to inherited pattern pressures of various kinds) is taken to produce identical and similar changes in sister languages, but this phenomenon is largely concentrated in the first centuries after their divergence from their parent language. Longobardi's focus is on distant genetic relationships, where drift is no longer very likely to be driving major alterations in a daughter language's structure. Shared inheritance and diffusion are therefore the only likely explanations for shared structures; if shared inheritance accounts for them, then they must have been present in the parent language; therefore we believe that Longobardi's proposed syntactic evidence for genetic relationship implies a claim about syntactic reconstruction.

Gianollo et al. 2004; Longobardi and colleagues elaborate on his proposed method, arguing that trees based on the kind of syntactic evidence he proposes correlate well with trees based on the standard Comparative Method. Among other things, they appeal to a subset of parameters whose settings, they believe, are especially stable, not easily subject to parameter resetting either through internal change or through borrowing. They do not explain their criteria for identifying this subset, but in fact most or all of the parameters settings we can identify on their list can be shown to have been borrowed by one language from another sometime, somewhere.[3] Longobardi and his collaborators do not seem to notice this difficulty.[4] We will not discuss his approach to genetic classification in detail because it is only tangentially relevant to the question of syntactic reconstruction.

In his 2001 article, however, Longobardi addresses issues of syntactic reconstruction more directly, arguing for a 'very restrictive theory of language change' (p. 300). Like other historical linguists, we are dubious about the prospects of a theory of change as restrictive as the one Longobardi aims for (but see fn. 4), and more especially about the prospects for a deterministic theory of change.[5] The set of changes Longobardi (2001) considers includes, for instance, the shift of French *chez* from a noun to a preposition. Category shifts of lexical items are much more common than Longobardi seems to believe: would anyone really want to claim, for instance, that the English word *Google*, which entered the language a few years ago as a proper noun and has very recently also become a verb (as in *I'll google that and get back to you about it*), had to undergo a category shift along the same lines?[6]

For these reasons, we are skeptical about the prospects for reconstructing parameter settings for parent languages in cases where the daughter languages do not agree in their settings. As we argue in detail below, there are important methodological limitations on the enterprise of syntactic reconstruction (along the lines of what is proposed by Lightfoot). However, especially given current developments in syntactic theory, more specifically within the minimalist approach to syntax (see e.g. Chomsky 1995, 2000, 2001; Lasnik 1999a; Epstein & Seely 2002b; Radford 2004 and references therein), we suggest ways in which syntactic reconstruction may be a feasible goal, albeit under strict methodological constraints that stem both from the formal properties of syntactic theory and from the Comparative Method, the methodology of reconstruction that has been developed by historical linguists especially on the basis of phonology and morphology. Where it is necessary, we frame the discussion about formal syntax using general properties of the Principles & Parameters approach to syntax, and where it is crucial, such as in Section 3, we appeal more specifically to properties of Minimalist syntax, also understood as a Principles & Parameters approach to syntax.

1.   The object of inquiry

The term 'language' has been used with many different, often imprecise meanings in linguistics, and it is important to define exactly what linguistic reconstruction stands for from different perspectives. Here we will take into account as much as possible the perspectives of both formal syntax and historical linguistics, to see to what extent their goals overlap regarding linguistic reconstruction. We take it to be the case that both research fields share one common goal, within the enterprise of investigating syntax from a diachronic perspective: that of understanding and explaining the linguistic phenomena corresponding to what historical linguists refer to as language change. However, at least from the perspective of formal syntax, both the term *language* and the term *change* need to be made precise.

---

3.   Neither Gianollo et al. (2004) nor Guardiano & Longobardi (2005) provide much discussion about the individual parameters they discuss, and some of their very abridged names are completely opaque; one cannot tell from the papers what they mean by (for instance) 'contextually inflected As', 'gramm. prox. in D' or 'def spread on APs.'

4.   Other aspects of their proposal will also likely meet with considerable skepticism among historical linguists. For instance, they propose 'the Anti-Babelic principle', which states in part that 'differences [among languages] can only be due to chance (i.e., no one ever confused languages on purpose)' (p. 5). This is false on empirical grounds. A small but growing number of examples from various parts of the world shows that people can and do change their languages deliberately in structurally nontrivial ways, sometimes abruptly, often in order to differentiate their speech more sharply from their neighbors' speech (see Thomason 2006a and sources cited there, and see also fn. 9 below).

From the perspective of syntactic theory, the range of differences among languages is in fact expected to be significantly constrained by the human faculty of language (Universal Grammar); therefore, these differences are not entirely subject to chance variation. It is rather the re-setting of parameters of variation in the course of development of new grammars that is dependent on chance variation in the linguistic input learners are exposed to in order to set the properties of their grammars (see Hale 1998; Lightfoot 1999, 2006; Roberts & Roussou 2003, and references therein for relevant discussion).

5.   However, in section 2 we do argue in a different way that a formal approach to syntax provides independent mechanisms that can impose restrictions on how syntactic change takes place. See also fn. 4 above.

6.   Bowern (this volume) argues for other potential problems in the conclusions Longobardi (2001) draws from the analysis of the development of *chez*.

It is especially from the perspective of generative syntax that the term 'language' as a focus of scientific investigation needs to be restricted in very specific ways. We do not aim at discussing this matter in detail here (for seminal work, see Chomsky 1957, 1965, and for more recent and general discussion, Chomsky 1986, 1988; Jackendoff 1994; Pinker 1994). For generative linguistics, the central properties of human language correspond to the system of specialized rules and principles internalized by humans (their native, internalized grammar) that allows them to combine basic linguistic units (sound units or phonemes) into larger structures (morphemes, words, phrases, and sentences). This corresponds to a discrete system with infinite generative power (that is, discrete infinity), or with the infinite creative potential characteristic of human language.

From this perspective, one can distinguish the internalized grammatical system of principles and rules of human language from the possible linguistic data that this system can generate. These have been taken as two different ways in which the term language can be understood. The first one, internalized grammars, is the main focus of the scientific investigation of human language syntax, but it is at most only indirectly accessible to detailed scrutiny, by means of hypotheses constructed by linguists on the basis of careful analysis of different sources of evidence. The second way in which the term language can be understood is in terms of the linguistic data that one can access to construct hypotheses about human language. Linguistic data, or linguistic outputs, in the form of naturally-occurring data/corpora and grammaticality judgments obtained from native speakers about stimulus data, are all relevant as indirect sources of evidence that can be used to construct hypotheses about the possible properties of internalized grammatical systems. Even though hypothesized internal grammars and linguistic outputs could a priori be taken to be two different aspects of human language, only the former constitutes a goal in the inquiry about the formal properties of human language. That is, hypotheses about the properties of human language are primarily hypotheses about grammars internalized by humans. Of course, successful hypotheses about the properties of internalized grammars also make testable predictions about what one should expect to find as linguistic outputs, or as linguistic evidence.

In sum, the main notion of 'language' that we take to be of potential interest for both formal syntax and historical linguistics is one corresponding to the grammars internalized by humans, and about which we can construct hypothesis on the basis of different linguistic data. In what follows we will often use the terms language and proto-language, but these terms should in fact be understood as references to grammars, as formal systems internalized by humans and which can generate outputs corresponding to the languages in question (e.g., the grammars internalized by speakers of English, French, Chinese and so on). Of course, even though individuals may have internalized grammars that are different in subtle

ways from the grammars of other individuals, these grammars are expected to have many properties in common, as the result of the fact that they instantiate a single human faculty of language. Furthermore, given that different individuals are often exposed to similar primary linguistic data, their internalized grammars can share more specific subsets of properties. The latter kind of similarity is one of the sources of the idealized notions of dialects and languages as shared by certain groups of individuals and not others, which plays an important role in the construction of hypotheses about language change. These are normally hypotheses about properties of grammatical systems as shared by communities, and which do not aim primarily at identifying or investigating subtle aspects of grammatical variation that are manifested in individual grammars.[7]

2.   On the nature of syntactic change and reconstruction

Turning back to the question of syntactic reconstruction, we first need to define how it relates to diachronic syntax as a field of inquiry and to the notion of syntactic change as a phenomenon. From the perspective of formal syntax, one goal of diachronic syntax should be to construct hypotheses about internalized grammars which have the appropriate properties to generate the linguistic outputs one finds in diachronic data. In this sense, the enterprise is not much different from that of synchronic syntax, insofar as one is interested in proposing successful hypotheses about human internalized grammars that would be able to generate certain linguistic outputs and not others. A second goal of diachronic syntax (and diachronic linguistics in general) should be to propose models that explain different phenomena of language change. Here we need to make precise what we mean by *language change* and, more specifically, *syntactic change*.

2.1   Syntactic change

From the perspective of both formal syntax and traditional historical linguistics, statements of change (e.g., 'the Proto-Indo-European labialized velar stop $*/k^w/$ became /k/ in Proto-Slavic', and 'the verb moved to a clause-second position in declarative clauses in Old English, but no longer does so in Modern English') are shortcuts that refer to a sequence of independent events that correspond either to (i) the acquisition of different grammars by two or more generations (the formal syntax perspective; see e.g., Lightfoot 1999, and Pires 2005, for discussion along

7.   The distinction between dialects and languages can be of little relevance in a context in which more fine-grained distinctions can or need to be made among individual grammars.

these lines, and see also Hale 1998, and Roberts & Roussou 2003) or to (ii) the innovation and then the spread of the innovation among adults (the traditional historical linguistics perspective). In this paper we focus on the formal child-centered view, according to which every child with normal language faculty undergoes a process of language acquisition that results in the development of at least one internalized adult grammar. Within generative syntax, this process is taken to be the result of the interaction between the human language faculty and the properties of the primary linguistic data (henceforth PLD) that each child is exposed to.

It is important to note, however, that the formal syntactician's position on language change is compatible with the view that innovations in adult speech form the basis for some, many, most, or even all linguistic changes (see Thomason 2006a and sources cited there), provided these innovations serve as input for the acquisition of new grammars. Because our focus here is on syntactic change from the perspective of formal syntax, we will not address the question of how syntactic changes that may occur in adult grammars, or at least in adult speech, are to be integrated into a view of the individual's grammar as internalized during the process of first-language acquisition. In all cases of systematic change in adult grammar and/or speech, we assume that whatever innovations the adults in a speech community introduce into their language will form part of the primary linguistic data for the next generation of child language-acquirers.

In a generative approach, then, the child, once exposed to primary linguistic data, is able to identify in the data relevant properties that permit progress from an initial state unspecified for the properties of any individual language to a sequence of later states (and eventually a final state) in the course of which the properties of an adult grammar are acquired. Children with similar linguistic experience may converge on different adult grammars. This is often the case, given deficiencies, incompleteness and noise in the data children are exposed to (for detailed discussion see Gleitman & Newport 1995; Pinker 1994, 1995; Hornstein & Lightfoot 1981). More interestingly from the perspective of diachronic syntax, individuals may acquire grammars that are substantially distinct in their formal properties from the grammars of their ancestors, especially given variation in each individual's linguistic experience – that is, variation in the PLD different individuals are exposed to. (See Lightfoot 1991, 1999; Hale 1998; Roberts & Roussou 2003 for different implementations of models along these lines.)

Schematically, the different steps corresponding to the process referred to as language change are given in (2). A child from generation 1 is exposed to the PLD set y, acquiring grammar 1, as in example (2a). Children from a subsequent generation (generation 2) undergo the same development, with the difference that the grammar they acquire is distinct in certain ways from the grammar of generation 1. In order to consider the existence of a process of language change, there should be some plausible link between the two generations. As it turns out, a subset of the linguistic outputs X produced by generation 1 constitutes at least part of the PLD set x' that individuals from generation 2 are exposed to.

(2)   a.   Individual from generation 1:
      $PLD_y \rightarrow$ faculty of language $\rightarrow$ Grammar 1 $\rightarrow$ set of outputs X

      b.   Individual from generation 2:
      $PLD_{x'} \rightarrow$ faculty of language $\rightarrow$ Grammar 2 $\rightarrow$ set of outputs Z

According to example (2), language change corresponds to the innovative properties of Grammar 2 as compared to Grammar 1. These innovations arise in the grammar of an individual from generation 2 as the result of exposure to $PLD_{x'}$, which can include, among others, a subset of the outputs X produced by individuals from generation 1.

Notice that there is no point at which Grammar 1 and Grammar 2 relate directly, which should be expected, since these are part of the internalized mental systems of different individuals. Similarly, there is no direct causal relation between the properties of the outputs X and the properties of outputs Z, even though these outputs are exactly what constitutes the main source of evidence for research on language change.

In other words, the sequence of independent acquisition events by different generations involved in *diachronic change* is connected only by means of the fact that the linguistic output produced by one generation functions as the primary linguistic data used by the next generation to construct hypotheses leading to the acquisition of their own grammars, which may or may not match properties of the grammars of previous generations. Under this view there is *no* such thing as diachronic change as a single, atomic event that links directly the structural source or object of the change (grammar 1), the description of the change, and the resulting changed object (grammar 2). Grammar 1 and Grammar 2 are part of the mental representation of different individuals, and as such do not come into any direct relation.[8] This provides a solution to a problem raised by Longobardi

---

8.   This puts aside the question whether the grammars internalized by bilinguals interact in any direct sense, allowing for the possibility of effects of one grammar on the other. Another situation that is not directly relevant here is language development itself, by which a child's grammar goes from S_o (the initial grammar) to S_i (the grammar attained by adults) by going through a number of intermediate states; these intermediate states count as a set of steps corresponding to the acquisition of a single adult grammar. Therefore, although this development represents different stages in grammar acquisition, in this case by the same individual, it is not relevant from the perspective of diachronic linguistics.

(2003: 283–4) in his interpretation of Lightfoot's position (and see also Hale 1998). According to Longobardi,

> *A priori*, the *ideally* restrictive theory of language change should probably claim that there is no such a thing as diachronic change (also cf. Lightfoot 1999). This is so because, if diachronic change exists, we are faced with a dilemma: either one must assume that at least some primitive change is unmotivated, i.e., largely beyond the scope of scientific inquiry (which is something incompatible with the *ideal* theory) or one loses any understanding of why the previous synchronic state was possible at all.

Given that we characterize (syntactic) change not as a primitive, but as a result of a series of events corresponding to the acquisition of grammars by different generations, it is in principle possible to avoid the complication raised by Longobardi and others regarding the motivation for change, since that motivation is now seen primarily as a set of factors affecting the acquisition of grammars by new generations.[9]

---

9. As indicated above, in the schema of (2), changes in the adult grammar – regardless of whether they are changes in the grammar acquired during childhood or comprise instead a partially parallel set of structures – will contribute to PLDx', namely, the data available to the next generation of language-acquirers. There is no doubt that dramatic changes do occur in adult language. However, it is an open question whether these changes affect the grammar that the adult internalized during child language acquisition. The clearest cases are those in which processes of first-language acquisition can be ruled out as a significant factor in the change process. One type of example is found in instances of what appear to be deliberate, conscious changes, such as the change in gender assignment of nouns in Uisai (a dialect of Buin spoken in Bougainville, Papua New Guinea): all nouns that are masculine in other dialects of Buin are feminine in Uisai, and vice versa (Laycock 1982). This lexical change resulted in reversed gender agreement markers throughout the syntax, and even if comparable reversals were known in other languages, the very shallow time depth that separates Buin dialects rules out any possibility that this reversal proceeded by gradual, incremental steps.

Another set of cases (which may or may not involve deliberate change) is the emergence and further development of pidgin languages, a process that, by definition, does not involve young children. The initial stages of pidgin genesis might not be considered to be language change at all; but for several relatively well-documented pidgins, such as Tok Pisin (Papua New Guinea) and Chinook Jargon (Washington, British Columbia, etc.), syntactic changes can be traced in the later history of the pidgin, well before it was acquired by native speakers (see e.g. DeGraff 1999 and references therein). A third set of cases would be a small number of well-documented bilingual mixed languages, which combine grammatical subsystems of two different languages. Michif (Manitoba, etc.), for instance, consists of French noun phrases (with their phonology, morphology, and syntax) inserted into Cree sentential frames, which include fully inflected Cree verbs. Michif must have been created by fluent bilinguals, because there is little or no distortion in either of its components; it does not resemble anything reported from the numerous studies

---

Longobardi (2003: 284) goes on to say that "Since it seems to be a fact that changes exist, the ideal (or perfectly minimalist) theory cannot be fully pursued." Given the approach to language change we adopt here, we argue that in fact a theoretical approach to syntactic change can be pursued which is fully compatible with the tenets of formal syntax.

2.2   Syntactic Reconstruction: What one would need to reconstruct

In syntactic reconstruction, the task of the analyst is to propose hypotheses about the grammars of previous generations (the grammars of speakers of the ancestral language) without having access to direct evidence in the form of outputs produced by those grammars. We thus seek empirical grounds for arguing that the distinct grammars acquired by subsequent generations (3b) (the grammars 2, 3, and 4 of individual speakers of the daughter languages) all resulted from exposure to a PLD set x that would have included at least part of the outputs X produced by individuals from generation 1 (the proto-language). In formal syntactic terms, Grammar 1 in (3a) corresponds to the set of grammars internalized by speakers from generation 1, and in historical linguistics terms, it corresponds to the proto-language. Grammars 2, 3, and 4 are the grammars of the daughter languages. Grammars 2, 3, and so on here are meant to represent the sets of grammars internalized by different speakers of each daughter language. So, for instance, Grammar 2 is likely to correspond to heavily overlapping grammars of different individuals, resulting from the fact that they constructed their grammars on the basis of very similar primary linguistic data. The analyst has access to outputs produced by individuals who acquired grammars 2, 3, and 4. These grammars need to be hypothesized on the basis of the available evidence in conjunction with what one knows about the human language faculty. This is not different from the study of synchronic syntax, which consists in constructing hypotheses about the grammatical system internalized by humans on the basis of indirect evidence. However, the analyst interested in syntactic reconstruction crucially needs to construct hypotheses about the properties of Grammar 1 (the grammar of the proto-language, underlined in (3a)), without access to actual outputs produced by individuals who acquired that

---

of bilingual first-language acquisition. Moreover, there is some evidence that it arose as a symbol of a new mixed-blood ethnic group, and young children do not have the social skills or motivations needed to create a language for such a purpose. Michif is now spoken almost entirely by people who know English and (in some cases) Chippewa, but not French or Cree, so its current independence from its linguistic roots is undeniable (see Thomason 2006a).

In cases of all these types, we assume that whatever innovations the adults in a speech community introduce into their language, the next generation of children will receive as part of their PLD.

grammar. That is, the analyst will make hypotheses about Grammar 1 solely on the basis of the outputs of Grammars 2, 3, and 4 (and also on the basis of the postulated properties of these grammars), which were by hypothesis acquired as the result of exposure of different individuals to a PLD set X that would have been largely produced (set of outputs X) by individuals with Grammar 1.

(3)   a.   Individual from generation 1:
           $PLD_y \rightarrow$ faculty of language $\rightarrow$ <u>Grammar 1</u> $\rightarrow$ ~~set of outputs X~~

      b.   Individuals from one or more subsequent generations:
           $PLD_x \rightarrow$ faculty of language $\rightarrow$ Grammar 2 $\rightarrow$ set of outputs W
           $PLD_x \rightarrow$ faculty of language $\rightarrow$ Grammar 3 $\rightarrow$ set of outputs Y
           $PLD_x \rightarrow$ faculty of language $\rightarrow$ Grammar 4 $\rightarrow$ set of outputs Z

This is of course a simplified picture of the empirical domain of reconstruction. In most actual cases the links between the outputs of an ancestral grammar and the acquisition of grammars by subsequent generations are much more indirect, mediated by a few or many generations between the ancestral grammar (Grammar 1) and each daughter language (Grammars 2, 3, and 4). In the vast majority of actual cases, we have only modern grammars, so that only two stages are relevant: one corresponding to the (reconstructed) grammar of the ancestral language, and the other one corresponding to the grammars of the daughter languages.

Crucially, even though the only directly accessible evidence is in the sets of outputs W, Y, Z produced by individuals with grammars 2, 3, and 4, the task of the reconstruction involves multiple steps: (i) postulate for the grammars 2, 3, and 4 models compatible with constraints imposed by the human faculty of language and which could possibly yield the sets of outputs W, Y, and Z; (ii) provide empirical grounds for the proposal that grammars 2, 3, and 4 result largely from exposure of different individuals to a PLD set that would have been to a large extent produced by individuals who internalized an unknown grammar 1 (the grammar of the ancestral language); (iii) postulate, solely on the basis of the properties of the postulated grammars 2, 3, and 4, hypotheses about the properties that would have been part of Grammar 1 (without having access to any subset of the outputs produced by Grammar 1 which could be directly informative about the grammatical properties one wishes to reconstruct).

In both syntactic change and syntactic reconstruction, as in formal synchronic syntax, the task of the analyst should then be to construct hypotheses about the internalized grammars of different individuals (the speakers of the proto-language, in the diachronic case) or, as is more often the case, *about the overlapping properties of the grammars of different sets of individuals* or, in more traditional but less precise terms, about the grammars of individual languages or dialects – American English, Quebec French, Classical Latin, Sanskrit, and so on, in more or less

fine-grained ways. The endeavor has been quite substantial from the perspective of synchronic syntax but so far it has been only partially successful, despite the availability of rich sources of linguistic evidence corresponding to different grammars. The task is daunting for syntactic change (given the limited availability of evidence), and even more so for syntactic reconstruction, given the usually complete unavailability of direct evidence from the ancestral language.

3.   The formal primitives of syntactic reconstruction

Now that we have defined in a formal way what counts as a language for purposes of the reconstruction of syntax, and what the goals of inquiry about syntactic change and syntactic reconstruction are, we will now address other important methodological questions that may be the source of much of the controversy about whether syntactic change is possible: What are the formal entities that count as the basic elements for syntactic reconstruction? Are the units of analysis and the elements that need to be reconstructed formal entities of the same kind? In considering these questions, we will provide important background for a discussion about how feasible syntactic reconstruction may be.

In section 2 we concluded that the task of syntactic reconstruction must involve the construction of hypotheses about the internalized grammars of different individuals, in this case the speakers of the ancestral language (here understood in an idealized way, abstracting away from variation within and across their different internalized grammars). Any specific hypothesis about a reconstructed grammar will then be a hypothesis about the formal entities that constitute that grammar, in the same way that hypotheses about synchronic syntax are hypotheses about the formal entities (features, rules, principles) that constitute internal, mental grammars of speakers and that can produce the kinds of output found in synchronic data. Given that to a large extent the formal properties of an internalized grammar are determined by the human language faculty,[10] they may be universal properties that one must find in all grammars. In this respect, at least in this case, one need not appeal to the limited and extremely indirect evidence available for the task of reconstruction in order to be able to reconstruct properties of the grammar of an ancestral grammar. Assuming that the same universal properties identified on the basis of current linguistic evidence hold for any human language

---

10.   Note, however, that certain general, natural principles of organization in the physical world may impose restrictions on the organization of the human mind, hence determining some of the properties found in human language (as discussed recently in e.g., Chomsky 2004a, b).

(the continuity hypothesis), one should expect that the grammar of any individual, a thousand or five thousand years ago, shared universal properties with human internalized grammars nowadays. For instance, consistently with most current theories of syntax, we can postulate without difficulty that every reconstructed grammar will have precise mechanisms of concatenation of lexical items in the syntax, including feature distinctions that constrain such concatenation (e.g., category, agreement, and Case features), locality constraints that restrict the ways in which lexical items can relate to different syntactic positions, predicate-argument distinctions that map in precise and restricted ways to syntactic positions, and so on. However, none of this should cause much excitement among historical linguists, since the existence of such properties simply predicts that core properties of human language remain the same, and do not change.

In fact, given current models of syntactic theory, at least within the Principles & Parameters framework, even the areas in which one does find variation, and which should be the appropriate locus of language change as formalized above, are areas in which the range of possible settings is dramatically restricted by properties of the human language faculty. Widely discussed examples within earlier P&P approaches included the parameters in (4):

(4) a. the null subject parameter (e.g., lack of null subjects in English vs. their presence in the Romance languages).

  b. the head(-complement order) parameter (distinguishing e.g. V-initial, prepositional languages such as English from V-final, postpositional languages such as Japanese).

  c. a V-raising parameter that determines that the verb starts in a base position and may overtly move to the left – raising in a tree structure – in some languages (distinguishing e.g. English John *often* dances, in which the verb *dances* stays to the right of the adverb, from French Jean danse *souvent*, in which the verb *danse* 'dances' moves to the left of the adverb *souvent*; see Emonds 1978, and also Pollock 1989 for the first analysis of this phenomenon in minimalist syntax). Old English behaved like French in this respect, which indicates that overt verb raising was lost in Modern English grammars.[11]

The locus of parameterization of syntactic properties has in fact been taken, within the Principles & Parameters framework, to be in the specification of features of individual lexical items, more specifically, functional categories (inflection, tense,

complementizers, determiners; see e.g. Borer 1984). Minimalism takes up this idea in detail and proposes an approach to syntactic theory in which operations in the syntax are massively, if not entirely, dependent on the interaction among features of lexical items. For instance, subject-verb agreement and nominative Case on a subject, as in <u>She</u> likes <u>chocolate</u>, results from the interaction, through partial or full identity, between the agreement (person, number) and Case (nominative) features of *She* and the inflectional head corresponding to *likes*. In this approach, the locus of syntactic variation is restricted to the properties of individual lexical items.

If we adopt this feature-based approach, especially as developed recently in Minimalism, the task of syntactic reconstruction can then be restricted to identifying variation in the feature specification of (functional) lexical items that determine syntactic structure and syntactic variation. In the case of the verb raising contrast between English and French/Old English, the prediction is that the variation is determined by variation in the feature specification of the inflectional head which attracts the verb overtly in French but not in English. This inflectional head is the one that carries tense and or agreement (see e.g. Pollock 1989; Chomsky 1995). In this theory, a syntactic functional head – the locus of inflection (tense and/or agreement) – displays the feature specification (strong, in Pollock's/Chomsky's terms) which forces the verb to raise overtly to the inflectional head in French and Old English, but not in Modern English. Of course, this does not mean that it suffices to compare the forms of individual words to identify such feature distinctions. Given that many if not most formal feature distinctions do not map directly to morphophonological distinctions among individual lexical items that carry the feature distinction, one still needs to have access to purely syntactic evidence in order to postulate hypotheses about most of the lexical (formal) feature distinctions that are assumed to have effects on syntactic properties. Again, in the case of the French/English verb-raising example, the relevant evidence that has been used to show the distinction involves the surface position of the verb and inflection with respect to certain manner and frequency adverbs and to negation. Even this evidence is not sufficient, since one also needs to consider the base and surface position of the verb and its arguments as well, to determine, for instance, whether the language is V-final (SOV) or V-medial (SVO). Terms such as SOV and SVO characterize the surface order in the clause, but each one of them may correspond to multiple possibilities in terms of base/underlying order in the clause.[12]

---

11. See e.g., Battye & Roberts 1995 for references regarding diachronic correspondents of the phenomenon, and see Lasnik 1999a, b and references therein for an alternative proposal involving overt verb movement in certain cases in Modern English.

12. The analysis of word order presents additional complications, which we come back to below. For instance, in an SVO sentence, all three elements may have raised or not, once or more than once. The different possibilities need to be considered in detail, if one hopes to make

Longobardi (2001, 2003) has suggested

> the possibility of the ideally restrictive theory relativized to syntactic change, namely the hypothesis that at least syntax, by itself, is diachronically completely inert and, thus, cannot be a primitive source of language change (the Inertial Theory of syntactic change) (2003: 284).

By 'syntactic change' Longobardi means "a change in the formal (non-phonetic, non-semantic) features FF of the items of the lexicon, including all the values of the parameters set for the particular language, taken to be encoded in the various entries, essentially along the lines of proposals by Borer (1984)" (2003: 284, fn. 6; see also Chomsky 1995: 6, 230–1). On this view, the set of formal features FF is the collection of features of lexical items that determine the set of steps going from the selection of lexical items from the lexicon and their assembly in the syntax to the submission of a syntactic object to the phonological interface. These features include, among others, categorical features (V, N, Tense, Determiner), agreement features (number, person, gender), and Case.

On the one hand, we have argued in section 2 that diachronic change does not exist as a primitive, atomic operation, but rather is just a convenient way (a shortcut) to refer to the acquisition of distinct grammars by different generations and to the fact that the output of a set of grammars 1 can function as the input triggering the acquisition of a (maybe different) set of grammars 2. This should apply to the acquisition of grammars in general, and not just to syntactic properties. On the other hand, under this complex, multiple-event view of what constitutes change, nothing in principle prevents distinctions among the grammars of different generations from affecting the specification of formal features of lexical items, e.g., agreement, Case, and categorical features. In fact, most of the literature on morphosyntactic change deals precisely with the occurrence of such distinctions. The characterization of parametric variation is grounded on distinctions in the formal feature specification of different (functional) items, and such specifications are determined in the process of language acquisition, thus introducing this as the main, if not the only, source of crosslinguistic syntactic variation (therefore diachronic variation as well). In sum, one can deal with the phenomena of syntactic variation and change by allowing for the possibility of acquisition of different formal feature specifications for lexical items. By doing so, one is in fact manipulating features that have a direct effect on syntax, and which are normally understood to be syntactic features. Under this view, we can

---

any precise claims about the properties of the grammar that would have given rise to a surface order such as SVO.

assume that in fact syntax is *not* inert in the case of syntactic change, as the latter is understood here.

We reach two conclusions here. First, the investigation of syntax goes significantly beyond the observation of descriptive generalizations about surface patterns that one finds in linguistic data. Syntactic variation is determined by variation in the specification of features of certain lexical items, arguably only functional lexical items that affect the properties of other lexical items that they relate to in the syntax. For instance, variation in the feature specification of the inflectional head tense/agreement affects the syntactic operations that the verb and its arguments undergo. Second, given that variation in certain lexical features may yield crosslinguistic syntactic variation, syntactic properties (as determined by the interaction among features of lexical items) can also be the object of change as the result of the acquisition of grammars by different generations, contrary to what Longobardi appears to argue for.

We also disagree with Harris & Campbell on the issue of the units of analysis for syntactic reconstruction. They argue (p. 344) in favor of the possibility of syntactic reconstruction, although they say there are clear limitations on the effectiveness in syntax of the comparative method (and also of internal reconstruction). In order to evaluate their proposal, we would like first to discuss the way in which they define syntactic correspondences (or equations). They cite various authors who have argued against the possibility of syntactic reconstruction, given the impossibility of establishing syntactic correspondences in a way similar to what is done in phonology and morphology. Two passages cited by H&C (p. 346) are especially relevant here:

> [...]For sentences ... acquisition by learning is most unusual ... Sentences are formed, not learned; morphemes and simple lexemes are learned, not formed ... Syntax deals almost exclusively with entities not learned, but constructed – or generated. (Winter 1984: 622–3)
>
> [...] Or, to use a different terminology, comparative phonology and morphology are interested in tokens and, through them, in types, whereas the domain of comparative syntax as practiced in recent years is exclusively that of types. (Jeffers 1976b: 6)

H&C then argue that:

> at least certain parts of syntax do have multiple "tokens" of a single "type"; although syntactic tokens are not memorized in the context of words, as phonological "tokens" are, they are often embedded in the context of a strictly limited syntactic pattern, which is the "type". (p. 347)

As Jeffers points out, the focus of comparative syntax (and of formal syntax in general) is not on tokens, and this does not seem to be disputed by H&C. However,

it is not clear from H&C's discussion what exactly comprises their types. They say, and we agree, that "it could never be claimed that a particular sentence was replaced by another, except for formulaic utterances" (p. 347). They use the term *syntactic patterns*, but then they go on to make statements such as this:

> in the examples below, there are simple sentences in Mingrelian, the (a) examples, which *correspond to* sentences in Laz, the (b) examples. [...]
>
> In these closely related languages, [fn. deleted], the paired Mingrelian [...] and Laz [...] examples are in an intuitively clear sense (see [their] section 12.2.3.2) cognate sentences, the "same" sentence in the two languages." (pp. 347–8; emphasis in the original)

Like Lightfoot (2002a, 2002b), we don't find H&C's argument on this point convincing. Although the output of the grammar provides the empirical evidence for syntactic reconstruction, the notion of 'cognate sentences' is dubious, especially in the absence of cognate grammatical morphemes.[13] The reason is that we cannot a priori consider that two sentences from different languages have related structures, without identifying precisely the grammatical structures that both sentences represent. The sentences found as part of linguistic data cannot be used directly to establish correspondences among different grammars. Only the internal properties of these grammars, which are identified by analyzing the linguistic data, can be used to establish such correspondences, and identifying such grammatical properties must be both a preliminary step (regarding the daughter languages) and an ultimate goal (regarding the ancestral language) from the perspective of syntactic reconstruction, as we argued above (section 2.2).

This also answers a question we raised in our introduction, by showing that the elements of analysis (the linguistic data) and the elements that need to be reconstructed are not formal entities of the same kind. Under this view, the best chance for achieving viable (if limited) syntactic reconstructions, we believe, is to compare individual grammatical properties, as they are identified from different syntactic structures. This includes structures in which shared grammatical morphemes play a major syntactic role for the reconstruction at hand. (See section 6 below for further discussion of H&C's approach.)

---

13. Harris (this volume) attempts to define in a more precise way what could constitute examples of syntactic correspondence for syntactic reconstruction, appealing especially to cases of morphosyntactic reconstruction. In most respects her approach follows the one developed in Harris & Campbell (1995), although she attempts to clarify that approach in certain respects, for instance in her discussion of the notion of syntactic patterns, instead of sentences, as the object of syntactic reconstruction.

To sum up, in scientific inquiry about diachronic syntax, surface linguistic outputs of the kind usually available for research in historical linguistics must be used as evidence for the development of models for (mental) grammars that could actually have been the source of such outputs (as discussed in section 2). That is, the task of syntactic reconstruction is not simply one of finding descriptive generalizations about surface patterns found in historical data, and how such generalizations connect. The need to construct models of mental grammars – if one wants to be successful in the investigation of formal syntax – imposes three important burdens on the task of syntactic reconstruction, which involves the construction of three types of hypotheses: (i) hypotheses about the properties of the mental grammars that could generate the outputs possible in the daughter languages; (ii) hypotheses about the properties of the mental grammars that could have been internalized by speakers of the ancestral language (the grammar of the proto-language); and (iii) hypotheses about how the different grammars of the daughter languages could have developed from the exposure of earlier generations to the output of the proposed ancestral grammar.

This clarifies briefly what is at stake, that is, what one is supposed to reconstruct in attempting to reconstruct syntax. We can now consider whether the current methodology used in historical linguistics (especially for phonology) can be successful in attempts to reconstruct syntax. Lightfoot (1999, 2002a, 2006) raises questions as to whether this task can be successful in the different cases. Even though we mostly agree regarding the specific cases of dubious reconstruction he discusses, we will argue below that at least some limited hypotheses can be supported about the properties of reconstructed grammars.

4. Constraints on syntax and their extension to syntactic change

This section addresses the question whether it is possible to establish general principles for reconstructing syntax, and if so how these principles relate to formal properties of human language syntax. Longobardi (2001, 2003) has argued that the task of syntactic reconstruction is a realistic goal and that the Principles & Parameters approach provides the necessary tools one needs to develop an appropriate discovery procedure. Nevertheless, he states that "current theories place very mild constraints on possible diachronic changes, something at odds with the trivial observation that actual 'language change' represents a tiny fraction of the variation made a priori available by Universal Grammar" (Longobardi 2003: 275).

Even though the cases of language change that have been attested and investigated in detail may instantiate only a small part of the crosslinguistic variation that has been identified so far, it is not at all clear that *possible* diachronic change

should in any sense be more restricted than the space of possible crosslinguistic variation made available by properties of the human linguistic system. If certain grammatical properties are possible, as verified by means of attested crosslinguistic variation, then it is a priori possible to identify a process of grammatical change that led to attested variants, even though there may be no direct historical data that can represent a source for these variants. The absence of historical correspondents for certain crosslinguistic phenomena is simply due to the absence of historical records, and cannot be assigned to restrictions imposed by the linguistic system, since such restrictions would not have a synchronic correlate.[14]

For this reason, we argue against the view that it is necessary to impose restrictions applying specifically to syntactic change, beyond the constraints that are imposed by the human faculty of language on properties that arise in the linguistic system from a synchronic perspective. Current theories of syntax do have the means to constrain what counts as possible diachronic changes, given that they narrow down the space of grammatical innovation to grammatical properties that can be learned by humans (more specifically, acquired by children) as part of the grammar of a natural language, on the basis of exposure to a minimal amount of linguistic input.[15] The linguistic input provides the relevant triggers for the acquisition by the child of the properties that will eventually converge into her adult grammar, according to precise and restrictive boundaries imposed by the specialized linguistic system that only humans have. Longobardi (2001) also argues that:

> syntactic change not provably due to interference should *not* occur at all as a primitive – that is, unless forced by changes in the phonology, the

---

14.   An anonymous reviewer does point out that "there is reason to think there are restrictions on possible changes. [...] for instance, English could not simply become a free word order language without first developing some case (or adpositional) marking." We assume the reviewer is referring to the possibility of restrictions on change that would not be restrictions on linguistic variation. First, although it may be widely attested that a freer word order co-occurs with overt morphological marking (e.g., case or adpositions), it is not clear that dependence between the two is mandatory. Second, as we proposed, if it is indeed the case that a freer word order is dependent upon the existence of certain kinds of overt morphology, this dependence should follow from a restriction on the space of linguistic *variation* (e.g., freer word order would be restricted to grammars with certain kinds of overt morphology) and not be taken only as the result of a restriction on *possible* diachronic change.

15.   Of course, this does not warrant the conclusion that it is only children acquiring their native grammar who can manipulate the linguistic system in ways that will lead to grammatical innovation. Research on language contact, in particular, provides examples of linguistic innovation that may not have been triggered directly in the process of language acquisition, although they may have been primarily the result of language contact among adults (see e.g. Thomason & Kaufman 1988 and references therein).

semantics, or the lexicon, perhaps ultimately by interface or grammar-external pressures, [...] (p. 275).

It is unclear to us what Longobardi intends by claiming that syntactic change that is not the result of interference needs to be treated in a different way from syntactic change that is due to interference. The standard meaning of the term 'interference' in historical linguistics is the transfer of structural features from the grammar of one language to another. However, from a formal perspective, it is not clear why syntactic change (i.e., innovation) resulting from interference should in principle have significantly different structural properties from syntactic change that does not result from interference.

Here we would like to discuss Longobardi's claim that syntactic change (not due to interference) has to be forced by changes in the phonology, the semantics or the lexicon. We focus on this point in the remainder of this section. One possible source of confusion has to do with defining what exactly should count as syntactic change. Within the Principles & Parameters/Minimalist approaches the locus of parametric variation has been taken to be restricted to properties of lexical items. More specifically, it has been argued that the locus of variation is restricted to features of functional lexical categories (such as inflection, complementizer and determiner, to cite a few functional categories that have been widely adopted). In a somewhat related way, within Minimalism, it has been argued that all the operations that take place in the syntax emerge from the manipulation of features of lexical items, including features that are relevant only for syntactic operations, but also features that are relevant for the phonological and semantic interfaces. If we adopt this approach, it is reasonable to assume that the properties of the syntactic system depend directly on the specified features of different lexical items, and in this perspective the source of grammatical innovation is expected to be dependent on changes in this feature specification. Furthermore, such innovations can arise in the specification of features that are relevant for syntax, for (morpho)phonology and for semantics. On this view, something else needs to be said about Longobardi's argument that syntactic change "should not occur at all as a primitive – that is, unless forced by changes in the phonology, the semantics or the lexicon". We argue that syntactic change is simply the result of changes in the features of lexical items, more specifically the features that directly affect the output of syntactic operations. These features may be (formal) features that are not manipulated directly either by the semantic or the phonological interface, provided they are manipulated as part of syntactic operations. One example may be illustrated by the contrast in (5), showing the loss of V-movement in English to a position to the left of negation. This change has been attributed to the weakening of formal features of V and of the inflectional head (Infl) in English. In an earlier stage, in the absence of auxiliary verb, the V had to move obligatorily to Infl to check a feature that is argued not to be interpretable

by the phonological interface in Old English (thus the need for it to be checked or eliminated before any manipulation by the phonological interface). But this feature specification later changed so that the relevant feature was specified differently, and no longer needed to be eliminated or checked before the phonological interface, leading to the loss of overt V-movement to Infl in English, as in (5b).[16]

   (5) a. *Mary *speaks* not Greek.
        b. Mary does not *speak* Greek.

In this case, the relevant features that led to the change in question are arguably *not* phonological or semantic features, even though their manipulation as part of syntactic operations affects the output of the phonological interface.

5.   Can the Comparative Method be applied to syntactic reconstruction?

Now that we have discussed briefly how the investigation of syntactic change relates to formal syntax, and what would be at stake in an attempt to reconstruct properties of syntax, we turn in this section to the evaluation of cases in which the standard methodology of historical linguistics is applied to (morpho)syntactic changes that have been described in some detail.

    We note first that those who propose methods for syntactic reconstruction, in particular Harris & Campbell (1995), have primarily used examples whose history is not independently known. There is therefore no way to evaluate the methodology directly; the fruitfulness test is of course available, but typically a good deal of time must elapse before the results of a substantial reconstruction have been widely applied in the daughter languages. For instance, in the case of Saussure's 19th-century Laryngeal Theory – the most successful example to date of internal reconstruction – it took decades before enough results accumulated to convince skeptics. (It must be confessed, however, that the theory might still be controversial if the hypothesized consonants had not been discovered in newly-deciphered Hittite documents.) The reasons for adducing examples with unknown history have to do with the paucity of usable ancient syntactic attestations. Nevertheless, there are many texts in Ancient Greek and Latin, and in a number of other ancient languages in several language families, so the tests could be applied; but as far as we know, they haven't been. Current proposals, both Harris & Campbell's and Longobardi's, therefore seem to us to be inconclusive.

---

16.   This example omits many details that are not directly relevant for the present discussion. See e.g. Kroch & Taylor 1997; Pintzuk 1999; Pintzuk et al. 2000 for different analyses and other references.

    We illustrate this potential pitfall in our evaluation of three different cases below. In one case, the problem of word order, we will point out flaws in different attempts to reconstruct the corresponding syntax of the earlier grammar. In two other cases we take phenomena that have been described with respect both to the daughter languages and to the ancestral language. We will begin by using subsets of the daughter languages as the only initial source of evidence, attempting to reconstruct the syntax of the earlier grammar. Then we compare the results we obtain with what is actually known about the syntax of the ancestral language. First, we examine a set of changes in the morphosyntax of noun declensions in Slavic, and conclude that no firm hypothesis can be supported about the ancestral grammar. For the second example, the development of synthetic forms of the future in Romance languages, we suggest that the reconstruction hypothesis we are led to in fact matches what is known about the ancestral grammar, but it is still only tentative, given that it ultimately appeals to notions of naturalness or directionality of change that have not been independently motivated. As we will argue, these results show that no a priori conclusion can be drawn about the prospects for syntactic reconstruction. We suggest that in some cases it is possible to construct interesting and potentially useful hypotheses about the properties of earlier grammars, especially when the relevant source of syntactic variation is grounded in morphophonological distinctions. However, contrary to what different proposals might lead one to believe, the complexities associated with the investigation of syntax make the prospect for successful syntactic reconstruction much more limited than we might wish, despite the success of the comparative method in other domains of the grammar (see Fox 1995 for a general discussion about reconstruction).

    One source of what we take to be a misunderstanding about the success of the enterprise of syntactic reconstruction lies in the fact that in the literature on syntactic reconstruction, it's not always clear whether the authors are talking about the same thing. There appears to be some confusion about the nature of the enterprise. Lightfoot observes, for instance, that 'if there are no principles of history, there would seem to be no basis for reconstructing proto-languages where the most archaic languages do not agree' (2002a: 134); this view would rule out phonological, lexical, and morphological reconstruction as well as syntactic reconstruction. Longobardi (2001, 2003) apparently shares this view, at least with respect to syntactic reconstruction, since he argues for what appears to be a deterministic prediction of syntactic changes. But this approach does not correspond closely to what historical linguists have actually done in carrying out reconstructions over many decades; and, regardless of the general lack of historical principles that govern language change, especially outside the phonology, the achievements in this area have been very substantial.

    In historical linguistics, as elsewhere in the sciences, one methodological principle stands out: all things being equal, the simplest hypothesis wins, in efforts to

account for sets of data. Observers can and certainly do disagree often about what the simplest hypothesis is and about whether all things are truly equal. Nevertheless, this principle guides much in reconstruction, often overriding considerations of 'naturalness' (which is a slippery concept, even in the area of sound change). The principle should apply to syntactic reconstruction as well as to other areas. Let's turn now to the evaluation of different attempts to do syntactic reconstruction.

5.1   Why surface word order is not enough

Reconstructing basic sentential word order is notoriously controversial, for a variety of reasons: in many languages word order is sufficiently free that the concept of one basic word order has much less relevance than in languages with more rigid word order patterns; more importantly, the same surface word order pattern in two different languages may result from very different grammatical structures; and the languages of a family may vary so much in this area of the grammar that there is no clear candidate for the simplest hypothesis. Proto-Indo-European is a case in point. Different problems for attempts to reconstruct word order in PIE were discussed in detail as early as Watkins 1976 and Lightfoot 1979. Some Indo-European (IE) languages have quite free word order; some IE languages are SOV, some are SVO, and others are VSO; and most of the ancient SOV languages were arguably in close contact with SOV languages, so that this easily-transferred syntactic feature could be the result of foreign interference (see also Thomason 2006b for brief review). No matter which word order one might select as the basic PIE word order pattern, numerous changes will be required in many daughter languages. The majority view (given the SOV pattern of most of the oldest attested IE branches) is SOV, but there is no general acceptance that this is the best historical hypothesis. In spite of the recent boom in research on historical IE syntax, the crucial test of the validity of a historical hypothesis has not been met here: no generally-accepted research results have shown any hypothesis about PIE word order to be scientifically fruitful, enabling historical linguists to explain things that are otherwise mysterious.

But not all language families are like IE in this respect. Uralic, to take one example, is not. Of the three major branches of the Uralic, two are solidly SOV and the third, Finno-Permic, is mostly SOV (Harris & Campbell 1995, p. 364). But one sub-branch, Balto-Finnic, together with Lapp, has SVO word order. There is an explanation for this exception: the SVO languages, among them Finnish, have all been heavily influenced by Slavic and Germanic languages during their speakers' westward movement, and their SVO word order is generally agreed to be the result of IE influence. Moreover, Finnish, at least, shows apparent traces of a former SOV word-order pattern (Harris & Campbell 1995: 365).

In contrast, there is no contact or other explanation available to explain the SOV word order of the other Uralic languages. (Some of them may well have come under the influence of Altaic and other SOV languages in the past, but no contact-related hypotheses have been proposed to account for the Uralic SOV pattern.) It is therefore relatively uncontroversial to reconstruct SOV word order for Proto-Uralic. Does this mean that Proto-Uralic certainly had SOV word order? No, of course not. But it does mean that the simplest hypothesis promotes reconstructing SOV for the family as a whole: this would require only one word-order change, in the immediate ancestors of the Balto-Finnic and Lapp languages. Positing SVO or any other word order pattern would require at least three independent word-order changes, one in Ugric, one in Samoyedic, and one or more in non-Balto-Finnic sub-branches of the Finnic branch of the family. Furthermore, it is unclear whether the proposed SOV identified in the different branches and proposed for the ancestral languages corresponds to the same underlying grammatical properties across the different languages, showing that the task of reconstruction is both imprecise and incomplete here, along the lines of the discussion in section 2.2, since the observations about surface word order alone are restricted to the description of linguistic outputs, requiring additional investigation about the grammatical properties that may have yielded the surface orders in the different languages.

5.2   Reconstruction in the absence of precise criteria?

Other examples, even without methodological discussions, are inconclusive. Kroeber's impressive achievement, in his 1999 book on syntactic reconstruction of Proto-Salishan, is a compilation, comparison, and analysis of large amounts of syntactic data from most of the 23 Salishan languages. He does make proposals for reconstruction, but many of these are tentative. In discussing the pan-Salish morpheme -m, for example, he observes that

> All Salish languages have ... [a] construction ... used to mark reduced topicality of the agent of a transitive clause, and/or increased topicality of the patient. ... [D]iscourse functions are analogous to...passives in European languages, but their morphological form in most Salish languages is rather different... (1999: 25).

Kroeber goes on to observe that the two most common forms of this A[gent] D[emotion] marker are -m and -t. In the Interior Salish branch of the family, -m is used with a 1SG or 3 patient, and -t is used with all other patients; in the Coast Salish languages, however, -m is used in main clauses and -it (or a similar suffix) in both nominalized and non-nominalized subordinate clauses (pp. 26–27). He then says that probably the Interior Salish distribution of -m and -t reflects the merger of two distinct AD paradigms, perhaps similar to Coast

Salish: "at any rate, it is difficult to see how the Coast Salish distribution of *-m* and *-(i)t* could have evolved from the Interior Salish one" (1999: 27). But he also points out that Bella Coola, a one-language branch of the family generally considered coordinate with Coast Salish and Interior Salish, has a distribution of these two suffixes that closely resembles the Interior Salish pattern; this has been suggested as one piece of evidence for a closer relationship between Bella Coola and Interior Salish (p. 27). An alternative possibility comes to mind: perhaps Bella Coola and Interior Salish are archaic in this respect, and Coast Salish has innovated the main clause/subordinate clause pattern. This seems the simpler hypothesis, in the absence of considerable other evidence in favor of a closer Bella Coola/Interior Salish link. In sum, the difficulty of imagining how it could have occurred cannot be criterial, especially given how ignorant we are about syntactic change generally.

5.3   When morphosyntactic reconstruction fails

Harris & Campbell (1995) apparently wish to appeal to a syntactic analogue of the regularity hypothesis in sound change.[17] They address this issue in detail, arguing that "syntactic change is indeed regular, in the sense that it is rule-governed, non-random" (p. 347; see also p. 326). But this is emphatically not the sense in which sound change is regular. According to the regularity hypothesis of sound change, if *x* changes to *y* in environment *z* in one word in language A, then *x* will change to *y* in that environment in every word of language A. Lightfoot (1979: 164, quoted by Harris & Campbell, p. 367), by contrast, argues that "[s]yntactic change is in large measure analogical"; Harris & Campbell counter by saying that the effects of analogy can be overcome when doing phonological reconstruction and therefore also in doing syntactic reconstruction. But this misses the point. Sound change does have analogic components. But where analogy affects sound change directly, it always depends on phonetic categories; a rather common example is the extension of a palatalization rule from the environment before a high front vowel to the environment before any front vowel, simplifying the palatalization rule. This kind of phonetic analogy presents no difficulties for comparative reconstruction. Morphosyntactic analogy affects only the results of sound change within paradigms. These distorting effects are overwhelmingly less prominent than the effects of regular sound change, so (as Harris & Campbell say) the problems they cause for comparative reconstruction can easily be

---

17.   Longobardi (p.c. 2005) tells us that he has never argued for regularity of syntactic change, considering, for instance, that change of syntactic features may spread regularly and incompletely through similar lexical items.

overcome. This simply is not true of morphological and syntactic change, where – as the Neogrammarians already knew in the 1870s – the primary mechanisms of change are to a large extent analogic. There is no foundation of non-analogic morphosyntactic change that can be used to sort out the effects of analogy; the effects of analogy are currently the whole picture, or close to it, for morphosyntactic changes that are not brought about by phonological change alone.

The sense in which analogic change is regular (in Harris & Campbell's terms) is apparently the same in both the morphology and the syntax. Here's a typical example. In late Proto-Slavic, the domain of the grammatical category of 'animacy' was restricted to just one place in the elaborate noun declensional system, the accusative singular of masculine nouns in a single noun class, the *o*-stems. Specifically, in animate masculine *o*-stem nouns, the accusative singular suffix was identical to the genitive singular; in inanimate masculine *o*-stem nouns, the accusative singular suffix was identical to the nominative singular. The domain of grammatical animacy has expanded through analogic changes in most modern Slavic languages, though not in the same ways in each daughter language of Proto-Slavic. In Novi Serbo-Croatian, for instance, animacy is still restricted to the *o*-stems, but now it holds in the plural as well as in the singular – that is, the rule no longer specifies 'singular', as in (6b):

(6)   a.   Late Proto-Slavic: *o*-stem MASC. ANIM. ACC. SG. = *o*-stem GEN. SG.
      b.   Novi Serbo-Croatian: *o*-stem MASC. ANIM. ACC. = *o*-stem GEN.

In Modern Standard Russian, by a set of analogic processes that presumably began with a spread from the *o*-stem singular to the *o*-stem plural and from there to the plurals of other noun classes, the situation is more complicated, as in (7) and (8):

(7)   a.   *o*-stem MASC. ANIM. ACC. = GEN.
      b.   *o*-stem MASC. INAN. ACC. = NOM.

(8)   a.   ANIM. ACC. PL. = GEN.
      b.   INAN. ACC. PL. = NOM.

The *o*-stem MASC.ANIM. plural is thus doubly specified; there's no way to collapse the two rules into one, because the singular forms of the other noun classes do not show any animacy distinctions. Now, the rules for marking grammatically animate accusative forms are quite regular in Late Proto-Slavic, modern Novi Serbo-Croatian, and Modern Standard Russian, and the analogic changes that led to the current states of the two modern languages were regular in that they affected all nouns in the relevant class, case, and number categories. But they were not regular in the sense of regular sound change. The fact that regular sound change is blind to all considerations of morphological, syntactic, semantic, and

discourse categories is precisely the source of its tremendous value for the comparative reconstruction of prehistoric linguistic states. Duality of patterning is crucial for tracking sound changes; as Harris & Campbell say (p. 349), there is nothing 'quite like' it in syntax. We would make it stronger: there may be nothing *at all* like it in syntax.

In any case, a decision on the validity of Harris & Campbell's attempts to use some concept of regularity in carrying out syntactic reconstruction will of course have to wait until their specific analyses either have been tested or have produced results that show them to be fruitful scientific hypotheses. The concept of testing a hypothesis about comparative reconstruction is rather straightforward: examine material from sister languages whose parent language is attested, carry out a comparative reconstruction, and then compare the results to the actually attested structures of the parent language. Admittedly, there are very few language families for which such tests can be carried out; but several branches of Indo-European are available for the purpose, together with a handful of other (sub-)families, among them Semitic. The most famous effort to test the Comparative Method is the partial reconstruction of Proto-Romance by Robert A. Hall, Jr. (1950). Hall found that, although some features of Vulgar Latin phonology could not be reconstructed because they were lost in all the daughter languages, overall the reconstruction was very successful.

To illustrate the practice of morphosyntactic reconstruction, we will return to the history of the grammatical category of animacy in Slavic, a sub-branch of the Balto-Slavic branch of Indo-European, which has been attested for about a thousand years. The oldest extensive Slavic textual material is Old Church Slavic, comprising Bible translations and other religious writing. Old Church Slavic (OCS), which is attested from the 11th century CE, is recognizably South Slavic rather than East or West Slavic, but it is considered to have been very close phonologically and morphologically to late Proto-Slavic. Slavic languages are very closely related, with a time depth of perhaps 1500 years. There were certainly dialect differences among the three branches by the time OCS was written down, but they are easily identified and not very extensive. Interpreting Proto-Slavic syntax is much more difficult, because of the problem of translation: the extent to which the oldest texts – the Bible translations – were influenced by the Greek original from which they were translated is unknown, since we have no truly independent documentary evidence of early Slavic syntactic patterns besides these translations. We therefore focus on inflectional morphology here. We will not extend our analysis to interconnected syntactic patterns and changes, but of course inflectional changes have syntactic as well as strictly morphological consequences. Our goal is to see whether the known late Proto-Slavic source structures turn out to be recoverable through comparative reconstruction based on the daughter languages. To keep the example within manageable bounds, we will restrict our comparison to a small but representative subset of Slavic languages.

Modern Slavic languages have a morphosyntactic category of animacy, which is expressed in noun declension and in adjectival agreement patterns. There is semantic variation in the modern languages with respect to the range of creatures that receive the lexical designation 'animate', but in most or all of the languages grammatically animate beings include at least humans, other mammals, birds, and other vertebrates, in at least some grammatical contexts. The animate/ inanimate system of noun classification co-exists with two other classificatory systems: a division into biologically-based categories of masculine, feminine, and neuter, and a partly cross-cutting set of declension types with no semantic content but with separate sets of case/number suffixes. This last set of declension types is traditionally designated by the last segment in the noun stem: *o*-stem, *a*-stem, *i*-stem and so forth. Most masculine nouns and almost all neuter nouns belong to the old *o*-stem declension; most feminine nouns belong to the old *a*-stem declension. In some of the modern languages the old *i*-stem declension contains both masculine and feminine nouns, and in others there are only feminine *i*-stems. In addition, there are a few masculine *a*-stems and a handful of irregular nouns that partially preserve other old declension types. These are found in all Slavic languages that have preserved substantial noun declension from Proto-Slavic. Only Bulgarian, Macedonian, and some southeastern dialects of Serbo-Croatian have lost declension almost entirely. Although the designations *o*-, *a*-, and *i*-stems are now anachronistic, we will use them below for convenience, as ways of referring to the different declensions.

In most Slavic languages, only nouns in certain declensions are identified lexically as animate. In some languages only masculine nouns can be grammatically animate, while in others both masculine and feminine nouns show animacy distinctions. Neuter nouns are never grammatically animate in Slavic. In the examples given in Tables 1–4 below, accent marks are omitted – not because they're unimportant, but because they don't affect any of the animacy patterns. For the same reason, we give no examples from neuter nouns.

Balto-Slavic languages preserve Proto-Indo-European (PIE) noun declension better than any other branch of IE. In addition to the two older noun classification systems, Slavic has preserved up to seven of the eight PIE cases: Nominative, Accusative, Genitive, Instrumental, Dative, Locative, and – in a few languages only – Vocative. We omit Vocative from our partial paradigms below, since it has no direct connection with the animate/inanimate distinction. In tables 1–4, then, are four partial declensions from modern Slavic languages. A comparison of animate with inanimate forms shows where the two categories differ (see Thomason 1976 for a more detailed discussion).

Table 1. Standard Serbo-Croatian noun declension (partial paradigms)

Case/Number	*o*-MASC. ANIM	*o*-MASC. INAN	*i*-FEM. ANIM/INAN	*a*-FEM. ANIM/INAN
	'priest'	'table'	'thing'	'fish'
NOM.SG	pop-0	sto-0	stvar-0	rib-a
ACC.SG	**pop-a**	**sto-0**	stvar-0	rib-u
GEN.SG	pop-a	stol-a	stvar-i	rib-e:
INST.SG	pop-om	stol-om	stvar-ju	rib-o:m
DAT.SG	pop-u	stol-u	stvar-i	rib-i
LOC.SG	pop-u	stol-u	stvar-i	rib-i
NOM.PL	pop-ov-i	stol-ov-i	stvar-i	rib-e
ACC.PL	pop-ov-e	stol-ov-e	stvar-i	rib-e
GEN.PL	pop-ov-a:	stol-ov-a:	stvar-i:	rib-a:
INST.PL	pop-ov-ima	stol-ov-ima	stvar-ima	rib-ama
DAT.PL	pop-ov-ima	stol-ov-ima	stvar-ima	rib-ama
LOC.PL	pop-ov-ima	stol-ov-ima	stvar-ima	rib-ama

Table 2. Novi (Čakavian) Serbo-Croatian noun declension (partial paradigms)

Case/Number	*o*-MASC. ANIM	*o*-MASC. INAN	*i*-FEM. ANIM/INAN	*a*-FEM. ANIM/INAN
	'priest'	'roof'	'bone'	'cow'
NOM.SG	pop-0	krov-0	kost-0	krav-a
ACC.SG	**pop-a**	**krov-0**	kost-0	krav-u
GEN.SG	pop-a	krov-a	kost-i	krav-e:
INST.SG	pop-on	krov-on	kost-u:n	krav-u:n
DAT.SG	pop-u	krov-u	kost-i	krav-i
LOC.SG	pop-i	krov-i	kost-i	krav-i
NOM.PL	pop-i	krov-i	kost-i	krav-e
ACC.PL	**pop-i:h**	**krov-i**	kost-i	krav-e
GEN.PL	pop-i:h	krov-i:h	kost-i:h	krav-0
INST.PL	pop-i	krov-i	kost-ami	krav-ami
DAT.PL	pop-o:n	krov-o:n	kost-a:n	krav-a:n
LOC.PL	pop-i:h	krov-i:h	kost-i:h	krav-ah

Table 3. Standard Russian noun declension (partial paradigms)

Case/Number	*o*-MASC. ANIM	*o*-MASC. INAN	*i*-FEM. ANIM	*i*-FEM. INAN	*a*-FEM. ANIM	*a*-FEM. INAN
	'wolf'	'table'	'horse'	'bone'	'fish'	'hill'
NOM.SG	volk-0	stol-0	lošad'-0	kost'-0	rib-a	gor-a
ACC.SG	**volk-a**	**stol-0**	lošad'-0	kost'-0	rib-u	gor-u
GEN.SG	volk-a	stol-a	lošad'-i	kost'-i	rib-i	gor-i
INST.SG	volk-om	stol-om	lošad'-ju	kost'-ju	rib-oj	gor-oj
DAT.SG	volk-u	stol-u	lošad'-i	kost'-i	rib-e	gor-e
LOC.SG	volk-e	stol'-e	lošad'-i	kost'-i	rib-e	gor-e
NOM.PL	volk-i	stol-i	lošad'-i	kost'-i	rib-i	gor-i
ACC.PL	**volk-ov**	**stol-i**	**lošad'-ej**	**kost'-i**	**rib-0**	**gor-i**
GEN.PL	volk-ov	stol-ov	lošad'-ej	kost'-ej	rib-0	gor-0
INST.PL	volk-ami	stol-ami	lošad'-ami	kost'-ami	rib-ami	gor-ami
DAT.PL	volk-am	stol-am	lošad'-am	kost'-am	rib-am	gor-a:m
LOC.PL	volk-ax	stol-ax	lošad'-ax	kost'-ax	rib-ax	gor-ax

Table 4. Literary Czech noun declension (partial paradigms)

Case/Number	*o*-MASC. ANIM	*o*-MASC. INAN	*i*-FEM. ANIM/INAN	a-FEM. ANIM/INAN
	'son'	'castle'	'bone'	'woman'
NOM.SG	sin-0	hrad-0	kost-0	žen-a
ACC.SG	**sin-a**	**hrad-0**	kost-0	žen-u
GEN.SG	**sin-a**	**hrad-u**	kost'-i	žen-i
INST.SG	sin-em	hrad-em	kost-i:	žen-u:
DAT.SG	**sin-ovi**	**hrad-u**	kost-i	žen-e:
LOC.SG	**sin-ovi**	**hrad-u**	kost-i	žen-e:
NOM.PL	**sin-ove:**	**hrad-i**	kost'-i	žen-i
ACC.PL	sin-i	hrad-i	kost'-i	žen-i
GEN.PL	sin-o:	hrad-o:	kost'-i:	žen-0
INST.PL	sin-i	hrad-i	kost-mi	žen-ami
DAT.PL	sin-o:m	hrad-o:m	kost-em	žen-am
LOC.PL	sin-ex	hrad-ex	kost-ex	žen-ax

The tables are ordered according to the scope of animacy within the declensional system, from least to greatest. Inspection of the paradigms shows wide variation in the realization of the animacy category. Standard Serbo-Croatian distinguishes animate from inanimate nouns in just one place: the accusative singular of animate masculine *o*-stem nouns is identical to the genitive singular, while the accusative singular of inanimate masculine *o*-stem nouns is identical to the nominative singular (Table 1).

The same contrast is found in Novi Serbo-Croatian, but here the *o*-stem plurals show the same pattern, although the plural endings themselves differ from the Standard Serbo-Croatian endings (Table 2).

Like the two Serbo-Croatian dialects, Russian contrasts singular animate and inanimate nouns only in the *o*-stem masculine accusative. But unlike Standard Serbo-Croatian and Novi Serbo-Croatian, Russian carries the distinction through the plurals of all declensions, including both *i*-stem and *a*-stem feminine nouns (and also *a*-stem masculine nouns, not illustrated in Table 3).

Finally, Literary Czech has a very different pattern, contrasting animate and inanimate masculine nouns – but never feminine nouns – in a number of cases. In the *o*-stems all the singular cases except the nominative and the instrumental show the distinction (Table 4); in the masculine *a*-stems (not shown in Table 4)

the distinction appears in the singular only in the dative and locative, but in all the plural cases except the accusative.

Now, if these were the only four Slavic languages, what would we reconstruct for Proto-Slavic animacy? There are two main options. One possibility is that animate and inanimate nouns were originally widely distinguished in the declensional system – perhaps as in Literary Czech masculine nouns – and then the distinction lost ground in the other three languages, most strikingly in Standard Serbo-Croatian, where it survived only in the masculine o-stem accusative singular. The other obvious possibility is that animate and inanimate nouns were originally distinguished only in the masculine o-stem accusative singular, perhaps as in Standard Serbo-Croatian, and then spread to the plural and/or to other cases and/or other declensional types in the other three languages. Of course there are numerous other possibilities as well: perhaps Novi or Russian is closest to the original system, with contraction in Standard Serbo-Croatian and expansion in Czech.

The question is, do we have any grounds for choosing among the various possibilities? Even if we restrict our options to just the two most extreme possibilities, it's not necessarily an easy choice. If we start with the Czech pattern for masculine nouns, we can consider the notion of *drift*, according to which closely-related languages undergo numerous similar or even identical changes as a result of inheriting the same set of pattern pressures from their common parent, in this case Proto-Slavic.[18] The contractions in Standard Serbo-Croatian, Novi Serbo-Croatian, and Russian would be seen as variations on the same theme: loss of inflectional range of the animacy category. All the changes involved would be instances of analogic leveling. We could account for the fact that the only part of the pattern shared by all four languages – the animate accusative = genitive singular in the masculine o-stems – survives longest, thus retaining the distinction between nominative singular and accusative singular in this noun class: pragmatically, it is useful to distinguish accusative from nominative in languages which (like Slavic languages) have elaborate case systems and quite free word order, because an animate NP would be a plausible agent: that is, it would be hard to tell who did what to whom in a sentence like 'Hit Bill John.'

It's also possible to devise a scenario whereby Standard Serbo-Croatian is closest to Proto-Slavic in its morphological realization of animacy. Pragmatically, it makes good sense that an innovative distinction between animate and inanimate

---

18.   By referring to drift (Sapir 1921), we do not mean to imply that we have a theoretical account of such processes. We don't. The notion is well-established in historical linguistics as a label for a large range of parallel developments (e.g., Crowley 1987: 189, and see also Trask 2000: 'drift' p. 98, and 'independent parallel innovation', pp. 160–1). There is no widely accepted theory of drift, though it is generally assumed to be connected with pattern pressures (or structural imbalances), and ultimately with ease of learning.

nouns would arise in the first place to distinguish animate from inanimate masculine o-stem nouns: the reason is the same as the reason for (by hypothesis) retaining the distinction only in this part of this paradigm, as we just discussed. The other major declension class, the a-stems, already distinguishes nominative from accusative; the i-stems don't, but there are relatively few of those, and even fewer animate i-stems. Analogic spread of the distinction from the singular o-stems to the plural and/or other cases and declension types would be a typical set of changes, all of them analogic extensions.

Semantically, at least, one aspect of our reconstruction task seems easy: all four languages agree rather closely on the range of nouns to be identified as animate – roughly, all vertebrates – so we will confidently reconstruct that range for Proto-Slavic.

Now, what did Proto-Slavic in fact have in this domain? First, the category is an innovation in Slavic; it does not exist in other branches of Indo-European. In fact, it is quite clear that it is a late Proto-Slavic innovation, because its occurrence is not quite consistent in the earliest Old Church Slavic texts. Where it does occur, as we noted earlier, it is identical to the Standard Serbo-Croatian pattern: in the masculine o-stems, animate accusative singular = genitive singular, while inanimate accusative singular = nominative singular. Most later Slavic languages – not just the ones in Tables 1–4 – expanded the category to other parts of the overall declensional system.

Ironically, the one point on which we could feel confident about our reconstruction is dead wrong. In Old Church Slavic, only lexical items for free, adult, human males were grammatically animate; lexical items for slaves, children and animals (along with women, who were not designated by o-stem nouns in any case) were all grammatically inanimate.

5.4   Reconstruction of the Romance future?

This section presents one brief case study suggesting that in certain restricted conditions and under certain restrictive assumptions about language change, it is possible to construct hypotheses about syntactic reconstruction that make testable predictions about earlier grammars. As in previous cases, the methodology we adopt here consists in selecting a phenomenon for which we have a reasonable amount of information corresponding to the grammar of the ancestral language. We then use only the syntactic properties of the daughter languages to evaluate whether different hypotheses about the relevant syntactic properties of the proto-language grammar make the correct predictions. The case in question involves the Romance languages.

The synthetic forms of the simple future in most Romance languages are argued to have originated from a periphrastic structure in Latin in which an infinitive was followed by an inflected form of the verb *habere* 'to have', and which originally did not have a future interpretation (see Benveniste 1968; Bourciez 1946;

Fleischman 1982). The change yielded the forms of the simple future tense in many Romance languages, as indicated in (9) for French, Italian, Portuguese and Spanish:[19],[20]

(9)   Simple future forms derived from Latin: 'sing': stem+fut+person/number:

person/number	French	Italian	Portuguese	Spanish
1SG	chante-r-ai	cante-r-ò	canta-r-ei	canta-r-é
2SG	chante-r-as	cante-r-ai	canta-r-ás	canta-r-ás
3SG	chante-r-a	cante-r-à	canta-r-á	canta-r-á
1PL	chante-r-ons	cante-r-emo	canta-r-emos	canta-r-emos
2PL	chante-r-ez	cante-r-ete	canta-r-eis	canta-r-éis
3P	chante-r-ont	cante-r-anno	canta-r-ão	canta-r-án

In these languages, especially French and Spanish, the future person/number morphemes are still almost entirely identical to the inflected forms of the verb 'have'. The overlap between the two forms is indicated in boldface in 10):

(10)   Present tense of verb 'have'[21]

person/number	French	Italian	Portuguese	Spanish
1SG	**ai**	**ho**	**hei**	**he**
2SG	**as**	**hai**	**hás**	**has**
3SG	**a**	**ha**	**há**	**ha**
1PL	**avons**	**abbiamo**	**havemos**	**hemos**
2PL	**avez**	**avete**	**heis**	**habéis**
3P	**ont**	**hanno**	**hão**	**hán**

Roberts & Roussou (2002: 36–45, henceforth R&R) present a proposal to explain how the periphrastic infinitive+*habere* Latin form gave rise to the synthetic forms of the Romance simple future. Following a standard view regarding this development (as presented, for instance, by Fleischman 1982) they claim that this change involved three steps:

(11)   a.   *habere*, which did not have a future interpretation in Latin, was reanalyzed as a future auxiliary comparable to *will* [or shall] in Modern English;[22]

b.   "the auxiliary *habere*, an autonomous word, was reanalyzed as a syntactic affix" (p. 37);

c.   the morpheme corresponding to *habere* was reanalyzed as a lexical affix attached to the main verb.

According to R&R, the last change (11c) took place almost immediately after the earlier ones in French (Languedoïl); slightly later in Occitan, Catalan and Northern Italian; and considerably later in Ibero-Romance (except Catalan).

We consider here the question whether it is possible to make appropriate hypotheses about these different stages in the development of the Romance future forms from Latin by using only the modern Romance language data as evidence. We know a fair amount of detail about the original forms in Latin, so we should be able to evaluate different hypotheses about the reconstructed syntax of Proto-Romance.

We construct two different scenarios. In the first scenario, we consider only the properties of the Romance simple future in French, Italian and Spanish (9), in which the forms of 'have' were completely reanalyzed as morphological affixes in the future tense forms. From only these languages, it would be impossible to reconstruct with certainty a periphrastic construction in earlier grammars, and we would end up with a synthetic form of the future in Latin, corresponding directly to the modern Romance simple future. The morphosyntactic reconstruction would

---

19.   We're using the spelling of the different forms. Although there are significant changes in pronunciation, especially in French, they are irrelevant for our analysis.

20.   A similar development seems to have taken place in the case of the synthetic conditional forms, which display similar properties to the simple future.

21.   The *h* in the spelling of the 'have' forms in Italian, Spanish and Portuguese is no longer pronounced. The Portuguese and Spanish forms of 'have' are no longer used as possessive verbs. The Portuguese forms are currently more restricted in their use than the Spanish counterparts.

---

22.   According to Brinton (1991: 13), "Fleischman (1982: 58–59) [...] recognizes three stages in the development from a meaning of possession through a "future-oriented modality" to a pure future. In the first stage, *have* has its full meaning of possession, but with a "nuance of obligation", as in *I have a letter to mail*. At this stage, the construction is quite limited in its distribution (Benveniste 1968: 89–90). In the second stage, the possessive meaning of *have* is bleached and the obligative meaning comes to predominate, as in *I have a letter to write*. She points out, correctly, that one cannot actually possess an unwritten letter. There is gradual loosening of the selectional restrictions of *have* and wider distribution of the construction (Benveniste 1968: 90; Pinkster 1987: 207)." However, Brinton (1991: 13, fn. 18) further indicates that "Benveniste (1968: 90–91) argues that the Latin construction does not have a meaning of obligation; rather, it has a meaning of predestination, out of which develops a future meaning: 'what is to happen' becomes 'what will happen'." The specific interpretations of the original complex forms, as suggested by Benveniste and Fleischman, may have played an important role in the ensuing change to a future periphrastic form, although there are other factors discussed by these and other authors that contributed to the change, including the structural properties of these forms and the availability of related grammatical forms in the grammars in question.

therefore not be successful. However, the morphosyntactic similarity between the Romance simple future and the structure ⟨infinitive V+PRES 'have'⟩ would allow us at least to raise the hypothesis of a development relating both, between Proto-Romance and the modern Romance languages, even if we knew nothing about previous stages in the history of the modern Romance languages. A related source of evidence is the fact that the forms of the verb 'have' are used in the formation of many compound tenses in the languages in (9), such as the compound past tense in French (*passé composé*):

(12)   J'ai   /tu   as   /il a   /nous avons/vous avez/ ils   ont chanté
       I have/you have/he has/we   have /you   have/ they have sung
       'I/you/he/we/you/they sang.'

However, this morphological similarity between the conjugated forms of simple present 'have' (10) and the simple future inflectional affixes (9) in modern French, Italian and Spanish alone would be insufficient to support the hypothesis of a periphrastic source for the future in the structure ⟨infinitive V+PRES 'have'⟩. This is due mainly to the fact that none of these three daughter languages displays traces of the earlier stages (11a,b) in the path of innovation that led to the appearance of the modern Romance simple future. First, although a tense inflectional head may correspond in the lexicon either to a verbal inflectional affix or to a free morpheme (e.g., an auxiliary verb),[23] we would still need to find some identity between the semantic interpretation of the future inflectional morpheme and specific forms of the verb *have*, across the grammars of the different daughter languages. Consider the possibility that we take this lexical form-meaning correspondence as a relevant criterion to support a reconstruction also in the syntax (but in a more precise way than the correspondences proposed by H&C, given that here we are not proposing correspondences between sentences or among patterns, as H&C later suggest, but specifically between individual lexical items).

This form-meaning connection between the simple future inflectional morpheme and the auxiliary 'have', however, is not found in the modern Romance languages considered here, except for Romanian, which we discuss later. Second, one remaining, but important, question that also could not be answered on the basis of what is known about the current stage of these languages is why the present tense of 'have' would have become an inflectional affix in the simple future but, crucially, would not have yielded a similar development in other compound tenses, such

---

as the compound past (12). Clearly, one would need to find evidence in one or more of the daughter languages suggesting that the lexical verb (in its infinitival form) immediately preceded the simple present of 'have' in a periphrastic structure ⟨infinitive V+ PRES 'have'⟩ in the grammar of the proto-language. Structures in which ⟨infinitive V + PRES 'have'⟩ forms had future meaning could then have been input for the acquisition of synthetic simple future forms by new generations. However, this kind of evidence is also not present in modern French, Italian and Spanish, thus weakening further the hypothesis that the development in (11) would have taken place, if one considers only the data from these languages.

Consider now a second scenario, in which we add data regarding future forms from two other modern Romance languages: Portuguese and Galician. In at least some dialects of Portuguese and Galician the future affix displays a hybrid behavior in which a pronominal clitic can be inserted or must be inserted between the verb stem and the person/number future suffix, in what is referred to as mesoclisis, a property that is found only in the simple future and simple conditional verb forms, as seen in (13b), as opposed to (13a), in which mesoclisis does not occur, due to the presence of negation in the clause.[24] This has led to the proposal that the person/number future (and conditional) suffix has a special behavior in these grammars. R&R refer to it as a 'syntactic affix', as opposed to the 'lexical affix' status of the corresponding French, Spanish, and Italian forms.

(13)   a.   Eu não lhe   darei   /daria   esse livro. (Port.)
            I   not him.CL give-fut.1SG /give-cond.1SG this book
            'I would not give him this book.'

       b.   Dar-lhe-ei   /Dar-lhe-ia   esse livro. (Port.)
            Give-him.CL-FUT.1SG /Give-him.CL-COND.1SG this book.
            'I will give him this book.'/'I would give him this book.'

Portuguese and Galician provide support for the relative independence of the person/number future suffix from the verb stem. Therefore, there is synchronic evidence, for this set of five Romance languages – French, Italian, Spanish, Portuguese, Galician – for the status of the person/number future inflection as both a lexical affix and a clitic (syntactic affix, in R&R's terms). What can the new data from Portuguese and Galician contribute to hypotheses about the development of the Romance simple future? If we consider simply the lexical affix vs. clitic/syntactic affix status of

---

23.   This would be possible under a weak lexicalist approach to the representation of functional heads, in which inflectional morphemes are concatenated with their hosts in the syntax (e.g., Chomsky 1970; Lasnik 1999b).

---

24.   As mentioned before, the conditional forms apparently underwent a development similar to that of the simple future in most of the Romance languages under consideration here. Colloquial Brazilian Portuguese has lost the possibility of mesoclisis entirely. According to Gerardo Fernández-Salgueiro (p.c. 2004), colloquial Galician also no longer shows this possibility.

the future person/number inflection in the two groups of languages, two directions of change are possible for the future inflection:

(14) a. Hypothesis 1 (French/Italian/Spanish innovate):   clitic > affix
      b. Hypothesis 2 (Portuguese/Galician innovate):   affix > clitic

According to hypothesis 1, French/Spanish/Italian would have innovated, whereas Portuguese/Galician would have at least partially kept the clitic status of the future inflection. Under hypothesis 2, Portuguese/Galician would have innovated, and an affixal form would have turned into a clitic. Neither development is unheard of, and although (14a) has been observed to be much more common, (14b) has been taken to arise in the case of the English genitive marker -s, analyzed as a clitic that developed from a morphological affix (Lightfoot 1999, and see also Allen 1995, 2002 and references cited there). Considering only the number of languages that would have innovated under either hypothesis, hypothesis 2 would be the simplest one, especially if Portuguese and Galician were taken to form a single branch in the family tree. This would be the wrong choice, given what is known to date about the development of the Romance simple future.

What about hypothesis 1 (14a)? Here we would have to argue that French, Italian and Spanish innovated, possibly independently, but undergoing a common change, in the process known as drift (see section 5.3 and fn. 18). This hypothesis is supported by data from previous stages of Romance, as the summary in (11) indicated. However, the purpose of the case studies we consider here is to construct hypotheses on the basis of restricted subsets of the available data, in order to be able to evaluate them against the additional, larger data set. If we consider again the morphological similarity between the future inflection and the present tense of 'have', now in combination with the added facts about Portuguese/Galician, we find limited, fragmentary support for a development of the future along the lines of (11), and which would fit Hypothesis 1 (14a):

(15)   (i) Vinf +'have' > (ii) Vinf+clitic > (iii) Vinf+affix (cf. (11))

However, neither of the changes in (15) is overwhelmingly supported by what has been discussed so far regarding the Romance future. The languages we considered do not provide direct evidence for stage (i), aside from the formal similarity with the present tense of 'have'. Furthermore, all the languages but Portuguese/Galician would have to have innovated in going from stage (ii) to stage (iii), a process that would have to be motivated from the perspective of syntactic change.

In order to make a choice between the two hypothesis in (14), one might wish to invoke considerations of naturalness or directionality of change that would favor the change from clitics to affixes (hypothesis 1, the correct one), and not

the other way around. In section 6 below we point out some problems with this approach.

Before we close this discussion, let us consider whether the addition of data from another modern Romance language affects our conclusion. Modern Romanian, unlike the languages discussed so far, does not have a simple future form. Instead, it has only periphrastic forms for the future, and at least one of these makes use of a form of 'have', in combination with an infinitive (16a) or a subjunctive form (16b) (examples from Joseph 1983: 163):[25]

(16) a. am   a   mă   duce   la oraş.
          have.1SG PRT 1SG.REFL lead.INF to town
          'I will go to town.'

      b. am   să   fac.
          have.1SG PRT do.1SG.SUBJ
          'I will do.'[26]

Interestingly, even though the Romanian data would not provide direct evidence for stages (i) and (ii) of the development in (11), it does provide relevant information that cannot be found directly in any of the other languages we considered – namely, for a periphrastic form of the future that uses the verb 'have' in combination with an infinitive or subjunctive form of the main verb.

Even though this does not constitute direct evidence for the structure ⟨Vinf+'have.PRES.'⟩, which would have led to the stages of the development in (11),

---

25. Modern Romanian usually displays subjunctive forms in the place of infinitives, even though it has been taken to still display infinitival forms in certain contexts (see Joseph 1983: ch. 6 and references cited there).

26. An anonymous reviewer cites Maria Jordan (p.c. to reviewer) as saying that (16a) is not in fact a future, but instead means 'I have to go to the city/I am supposed to go to the city'; the reviewer goes on to say that "even deontic *avea* 'have' has fallen out of use, except perhaps regionally." Other speakers of Romanian, however, interpret sentences like (16a) as a periphrastic future. Larisa Avram (p.c. 2004), for instance, indicates that at least one dialect of Modern Romanian has a periphrastic form of the future with the auxiliary *avea* 'have', and provides example (i) (For similar examples, also interpreted as futures, see Avram 1986, 1999; Dobrovie-Sorin 1994; and references therein).

(i) am   sa   merg
     Have.1st.SG SA.SUBJ go-1st.SG
     'I will go/I am going.'

The same reviewer also reports that Maria Jordan finds example (16b) ungrammatical as it stands, because it needs an overt object. Since this example comes from Joseph 1983, we have retained it; it is possible that there is interspeaker variation on sentences of this type.

it provides supporting evidence for a hypothesis that would reconstruct those stages. Furthermore, if the Romanian evidence, combined with the morphological similarity between the future inflection and 'have' in the languages that have a synthetic future, does constitute evidence for the existence of the first stage in (11a), it also supports the hypothesis that the development in the other Romance languages considered here would have been as in (14a)–(15).[27]

These results do not eliminate the need to address important methodological questions, in particular the ones that arise when competing hypotheses appeal to naturalness (or markedness). However, they do suggest that interesting and reasonable hypotheses about the syntactic properties of proto-language grammars can be established solely on the basis of properties of the grammars of the daughter languages.

6.   Conclusion

As noted just above, there are still methodological questions that must be considered in order to approach the task of doing syntactic reconstruction. Some of these considerations are common to research on language change in general. First, certain problems and questions arise if we use considerations of naturalness in choosing among different hypotheses.

More directly relevant to the discussion about syntactic reconstruction, if considerations of naturalness (or, more specifically, directionality of change) may be invoked to choose among different hypotheses, it still remains to be explained from the perspective of the linguistic system (and its acquisition) what determines what kinds of processes do occur more naturally (e.g., the ones leading from a change from clitics to affixes, as discussed in section 5.4). Specifically in the grammars in which the change does happen, what would explain the change, that is, what would be the structural source in the linguistic system for the acquisition of the innovative grammar?[28]

---

27.   One might point out that the Romanian situation is in fact innovative, even with the 'have+infinitive' future cases in (16a). However, in the present case study, the fact that Romanian might be innovative is not known a priori, since we are introducing only the modern Romanian data into the discussion, in the attempt to further test the reconstruction hypotheses in (11) with access only to a limited set of data.

28.   Roberts & Roussou (2003) attempt to develop a model that explains why this change is favored given complexity measures that come into play in the acquisition of new grammars.

We still need to explain *why* the change happens, and why it happens *only* in some grammars, and not in others, as was the case in Section 5.4 for the last stage that gave rise to the Future inflectional affixes as lexical affixes in French/Italian/Spanish, but not in Portuguese/Galician, at least until recently. Historical linguists are not optimistic about the prospects for making major advances in this area, however: the usual argument is that there are too many social and linguistic variables to be considered.

If certain properties of human language (e.g., clitics) are expected to change naturally into something else (e.g., affixes), and thus disappear, why should such properties arise and remain stable in the grammars of many generations? (See also, regarding this question, Hale 1998).

If the acquisition of some innovative properties in new grammars happens because they are more natural, how can the opposite development take place, as seems to have been the case in the change of the possessive marker *-s* from an affix to a clitic in English?[29]

An important question has to do with the tendency to connect what seems to be more common or probable with what is more natural. This introduces a methodological complication. The fact that certain phenomena are more common does not necessarily result from the fact that they are the most natural phenomena, especially when one has a limited or fragmentary amount of empirical data to appeal to. In any case, the best and possibly only certain approach to invoking notions of directionality might be to provide explanations based on the properties of the linguistic system for why certain phenomena tend to be recurrent, affecting more often the course of language change.

There are further problems in extending to syntax the notion of naturalness. First, even in the case of phonological change, which is better understood than syntactic change, the notion of 'naturalness' is vexed. Sometimes the (apparently) most natural changes fail to happen; sometimes a change that might be considered highly unnatural – or at least very rare – does happen. The latter may have to do with notions of (un)markedness, taken to be the basis of 'naturalness'. Marked things may be taken to be harder to learn. But pinning it down is another matter (see Thomason & Kaufman's discussion of this issue, 1988: ch. 2). Second, our store of knowledge about syntactic change is so puny compared to what is known about sound change that most hypotheses about syntactic reconstruction are guesswork.

---

29.   As an anonymous reviewer points out, there are other examples of change from affix to clitic. The reviewer gives the Polish person markers as an example, citing Andersen 1987 and Witkoś 1998.

In morphophonological change there are many controversial claims, especially in the grammaticalization literature (see H&C for discussion).

In this paper we have focused primarily on foundational and methodological questions that (in our opinion) must be addressed first in order to carry out any syntactic or morphosyntactic reconstruction successfully. (Below we will use the term morphosyntactic to cover both of these closely-related categories.) Here we will sum up our main conclusions about the feasibility of reconstruction in this domain. First, and most importantly, the greater difficulties inherent in morphosyntactic reconstruction by comparison to phonological reconstruction do not justify abandoning the attempt to reconstruct morphosyntax; they do, however, mean that the level of success will almost certainly remain lower than the level of success in phonological reconstruction. Similarly, given the state of current research, our confidence in the accuracy of the results of morphosyntactic reconstruction will probably always be less. A major reason for this pessimism lies in the fact that morphosyntactic changes are not regular in the same sense that most phonological change is regular: regular phonological change is blind to all considerations of meaning, while in morphosyntactic change, by contrast, analogy restricted by certain structural and meaning properties is a major non-phonological driving force. In other words, structural and meaning restrictions play a crucial role in morphosyntactic change but not in regular phonological change.

Second, uniformity of surface forms in all the daughter languages at a particular structure point (e.g., sentential word order or *yes/no* question formation) permits the reconstruction of that surface form in the parent language, but always with the caution that more than one grammar might produce the same surface configuration.

Third, where morphosyntactic patterns share cognate morphemes, as in the Slavic animacy and Romance future examples above, functional identity or overlap in these morphemes leads to useful hypotheses about the proto-language constructions that gave rise to the daughter-language patterns. In some cases, like the Romance future – where our analysis built on the fact that a clitic-to-affix change is much more widely instantiated than an affix-to-clitic change – directional probabilities contribute to the formulation of these hypotheses; in other instances, like the Slavic animacy category, no such directional probabilities can be adduced as supporting evidence for a reconstruction. (In that example, although morphological simplification is the norm in Indo-European languages, morphological elaboration is well known from other language families, typically families that start out with less elaborate inflectional systems than Proto-Indo-European had.) In all these cases the possibility of drift must be taken into account: because daughter languages inherit the same structures from the proto-language, they undergo many similar or identical changes as a result of pattern pressures (structural imbalances, hard-to-learn bits of grammar) in the inherited material.

Fourth, efforts to reconstruct morphosyntax in the absence both of shared morphemes and of identity in surface forms in the daughter languages are much less likely to yield solid results. It is quite possible that further research will support directional probabilities, and these in turn may facilitate progress even here.

Current approaches to reconstruction that rely on clusters of parameter settings are difficult to evaluate because they have not yet been systematically developed, or tested on data where the relevant proto-language constructions are known (such as our Slavic and Romance examples), or presented in conjunction with well-worked-out examples. Our initial reaction to these proposals, however, is that they are essentially typological, not historical: clusters of features do not, in themselves, carry direct historical implications. There is currently no proven metric that would show the extent to which chance might be responsible for any particular clustering; even if a historical connection can be established for such a clustering, no method has been worked out to distinguish inheritance from diffusion; and although a number of case studies have been published in this area, no general paths of change have been established for parameter settings.

References

Allen, Cynthia. 1995. *Case Marking and Reanalysis: Grammatical relations from old to modern English*. Oxford: Oxford University Press.

Allen, Cynthia. 2002. "Case and Middle English Genitive Noun Phrases". Lightfoot, ed. 2002c, 57–80.

Andersen, Henning. 1987. "From Auxiliary to Desinence". *Historical Development of Auxiliaries* ed. by Martin Harris & Paolo Ramat, 21–52. Berlin: Mouton de Gruyter.

Avram, Larisa. 1986. "A few Remarks on Futurity in Romanian". *Revue Roumaine de Linguistique* 31: 3. 203–7.

Avram, Larisa. 1999. *Auxiliaries and the Structure of Language*. Bucharest: Editura Universitatii din Bucuresti.

Battye, Adrian & Ian Roberts, eds. 1995. *Clause Structure and Language Change*. Oxford: Oxford University Press.

Benveniste, Emile. 1968. "Mutations of Linguistic Categories". *Directions for Historical Linguistics* ed. by Yakov Malkiel and Winfred P. Lehmann, 83–94. Austin: University of Texas Press.

Borer, Hagit. 1984. *Parametric Syntax*. Dordrecht: Foris.

Bourciez, Edouard. 1946. *Eléments de linguistique romane*. Paris: Klincksieck.

Brinton, Laurel. 1991. "The Origin and Development of quasimodal *have to* in English". Paper presented at the *Workshop on the Origin and Development of verbal Periphrases, 10th International Conference on Historical Linguistics* (ICHL 10), Amsterdam, August 16, 1991. http://faculty.arts.ubc.ca/lbrinton/

Campbell, Lyle & Alice C. Harris. 2002. "Syntactic Reconstruction and Demythologizing 'Myths and the Prehistory of Grammars'". *Journal of Linguistics* 38: 599–618.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton de Gruyter.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1970. "Remarks on Nominalization". *Readings in English transformational Grammar* ed. by Roderick Jacobs and Peter Rosenbaum, 184–221. Waltham, MA: Ginn Press.

Chomsky, Noam. 1986. *Knowledge of Language*. New York: Praeger.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 2000. "Minimalist Inquiries: The Framework". *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* ed. by Roger Martin, David Michaels, and Juan Uriagereka. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. "Derivation by Phase". *Ken Hale: A life in language* ed. by Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2004a. "Beyond Explanatory Adequacy". *Structures and beyond: Current issues in the theory of language* ed. by Adriana Belletti, 104–31. Cambridge: Cambridge University Press.

Chomsky, Noam. 2004b. "Three Factors in Language Design". Paper presented at the *Annual Meeting of Linguistic Society of America*. Boston, MA, Jan 2004.

Crowley, Terry. 1987. *An Introduction to Historical Linguistics*. Suva, Fiji: Univ. of Papua New Guinea Press.

DeGraff, Michel, ed. 1999. *Language Creation and Language Change: Creolization, diachrony, and development*. Cambridge, MA: MIT Press.

Dobrovie-Sorin, Carmen. 1994. *The Syntax of Romanian: Comparative studies in Romance*. Berlin; New York: Mouton de Gruyter.

Emonds, Joseph. 1978. "The Verbal Complex V'-V in French". *Linguistic Inquiry* 9. 49–77.

Epstein, Samuel, D. & T. Daniel Seely, eds. 2002a. *Derivation and Explanation in the Minimalist Program*. Oxford: Blackwell.

Epstein, Samuel D. & T. Daniel Seely. 2002b. "Introduction: on the Quest for Explanation". Epstein & Seely, ed. 2002a. 1–18.

Fleischman, Suzanne. 1982. *The Future in Thought and Language: Diachronic evidence from Romance*. Cambridge: Cambridge University Press.

Fox, Anthony. 1995. *Linguistic Reconstruction: An introduction to theory and method*. Oxford: Oxford University Press.

Gianollo, Chiara, Cristina Guardiano & Giuseppe Longobardi. 2004. "Historical Implications of a Formal Theory of Syntactic Variation". Paper presented at Diachronic Generative Syntax Conference (DIGS) VIII, Yale, June 2004.

Gleitman Lila, R. & Mark Liberman, eds. 1995. *Language: An invitation to cognitive science*. 2nd edn. Vol. 1. Cambridge, MA: MIT Press.

Gleitman Lila, R. & Elisa L. Newport. 1995. "The Invention of Language by Children: Environmental and Biological Influences on the Acquisition of Language". Gleitman & Liberman, eds.

Guardiano, Cristina & Giuseppe Longobardi. 2005. "Parametric Comparison and Language Taxonomy". *Grammaticalization and Parametric Change* ed. by Montserrat Batllori, Maria L. Hernanz, Carme Picallo & Francesce Roca. Oxford: Oxford University Press.

Haegeman, Liliane. 1994. *Introduction to Government & Binding Theory*. 2nd edn. Oxford: Blackwell.

Hale, Mark. 1998. "Diachronic Syntax". *Syntax* 1: 1–18.

Hall, Robert A., Jr. 1950. "The Reconstruction of Proto-Romance". *Language* 26: 6–27.

Harris, Alice C. & Lyle Campbell. 1995. *Historical Syntax in Cross-linguistic Perspective*. Cambridge: Cambridge University Press.

Hornstein, Norbert & David W. Lightfoot, eds. 1981. *Explanation in Linguistics: The logical problem of language acquisition*. New York: Longman.

Jackendoff, Ray. 1994. *Patterns in the Mind: Language and human nature*. New York: Basic Books.

Jeffers, Robert J. 1976. "Syntactic Change and Syntactic Reconstruction". *Current Progress in Historical Linguistics: Proceedings of the Second International Conference on Historical Linguistics* ed. by William M. Christie, Jr., 1–15. Amsterdam: North Holland.

Joseph, Brian D. 1983. *The Synchrony and Diachrony of the Balkan Infinitive*. Cambridge: Cambridge University Press.

Kemenade, Ans van & Nigel Vincent, eds. 1997. *Parameters of Morphosyntactic Change*. Cambridge: Cambridge University Press.

Kroch, Anthony & Ann Taylor. 1997. "Verb Movement in Old and Middle English: Dialect variation and language contact". Kemenade & Vincent, ed. 1997. 297–325.

Kroeber, Paul D. 1999. *The Salish Language Family: Reconstructing syntax*. Lincoln: University of Nebraska Press.

Lasnik, Howard. 1999a. *Minimalist Analysis*. Malden, MA: Blackwell.

Lasnik, Howard. 1999b. *Syntactic Structures Revisited: Contemporary lectures on classic transformational grammar*. Cambridge MA: MIT Press.

Laycock, Donald C. 1982. "Melanesian Linguistic Diversity: A Melanesian choice?" *Melanesia: beyond Diversity* ed. by Ron J. May & Hank Nelson, 33–38. Canberra: Australian National University Press.

Lightfoot, David W. 1979. *Principles of Diachronic Syntax*. Cambridge: Cambridge University Press.

Lightfoot, David W. 1991. *How to Set Parameters: Arguments from language change*. Cambridge, MA: MIT Press.

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

Lightfoot, David W. 2002a. "Myths and the Prehistory of Grammar". *Journal of Linguistics* 38: 113–136.

Lightfoot, David W. 2002b. "More Myths". *Journal of Linguistics* 38: 619–626.

Lightfoot, David W., ed. 2002c. *Syntactic Effects of Morphological Change*. Oxford: Oxford University Press.

Lightfoot, David W. 2006. *How New Languages Emerge*. Cambridge: Cambridge University Press.

Longobardi, Giuseppe. 2001. "Formal Syntax, Diachronic Minimalism, and Etymology: The history of French *chez*". *Linguistic Inquiry* 32: 2. 275–302.

Longobardi, Giuseppe. 2003. "On Parameters and Parameter Theory". *Syntaxtheorien: Modelle, Methoden, Motive* ed. by Elisabeth Stark & Ulrich Wandruszka 273–290. Tübingen: Gunter Narr.

Pinker, Steven. 1994. *The Language Instinct: How the mind creates language*. New York: W. Morrow.

Pinker, Steven. 1995. "Language Acquisition". Gleitman & Liberman, eds. 135–182.

Pinkster, Harm. 1987. "The Strategy and Chronology of the Development of Future and Perfect Tense Auxiliaries in Latin". *Historical Development of Auxiliaries* ed. by Martin Harris & Paolo Ramat 193–223. Berlin: Mouton de Gruyter.

Pintzuk, Susan. 1999. *Phrase Structures in Competition: Variation and change in Old English word order*. New York: Garland.

Pintzuk, Susan, George Tsoulas & Anthony Warner, eds. 2000. *Diachronic Syntax: Models and mechanisms*. Oxford: Oxford University Press.

Pires, Acrisio. 2005. "Verb Movement and Clitics: Variation and change in Portuguese". *Grammaticalization and Parametric Change* ed. by Montse Batllori, Maria-Luísa Hernanz, Carmen Picallo & Francesc Roca. Oxford: Oxford University Press.

Pollock, Jean-Yves. 1989. "Verb Movement, Universal Grammar and the Structure of IP". *Linguistic Inquiry* 20: 365–424.

Radford, Andrew. 2004. *Minimalist Syntax: Exploring the structure of English*. Cambridge: Cambridge University Press.

Roberts, Ian and Anna Roussou. 2002. "The History of the Future". Lightfoot, ed 2002c. 23–56.

Roberts, Ian and Anna Roussou. 2003. *Syntactic Change: A Minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.

Sapir, Edward. 1921. *Language*. New York: Harcourt, Brace & World.

Thomason, Sarah Grey. 1976. "Analogic Change as Grammar Complication". *Current Progress in Historical Linguistics: Proceedings of the Second International Conference on Historical Linguistics* ed. by William M. Christie, Jr. 401–9. Amsterdam: North Holland.

Thomason, Sarah Grey 2006a. "Language Change, intentional". *Encyclopedia of Language and Linguistics*, 2nd ed., vol. 6, ed. by Keith Brown, 346–349. Oxford: Elsevier.

Thomason, Sarah Grey 2006b. "Syntactic Reconstruction". *Encyclopedia of Language and Linguistics*, 2nd edn., Vol. 10. ed. by Keith Brown, 397–401. Oxford: Elsevier.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.

Trask, Robert L. 2000. *The Dictionary of Historical and Comparative Linguistics*. Edinburgh: Edinburgh University Press.

Watkins, Calvert. 1976. "Towards Proto-Indo-European Syntax: Problems and pseudo-problems". *Papers from the Parasession on Diachronic Syntax* ed. by Sanford B. Steever, Carol A. Walker & Salikoko S. Mufwene, 305–26. Chicago: Chicago Linguistic Society.

Webelhuth, Gert, ed. 1995. *Government and Binding Theory and the Minimalist Program: Principles and parameters in syntactic theory*. Oxford: Blackwell.

Winter, Werner. 1984. "Reconstructional Comparative Linguistics and the Reconstruction of the Syntax of Undocumented Stages in the Development of Languages and Language Families". *Historical Syntax* ed. by Jacek Fisiak, 613–26. Berlin: Mouton de Gruyter.

Witkoś, Jacek. 1998. *The Syntax of Clitics: Steps towards a minimalist account*. Poznań: Motivex.

Reconstruction in syntax

Reconstruction of patterns*

Alice C. Harris
SUNY, Stony Brook

In syntax, reconstruction is limited to patterns, that is, repeated surface forms paired with a consistent function or distribution. Comparative reconstruction is illustrated with the example of the pattern of yes/no questions in Kartvelian languages and the core case marking pattern in the languages of the Nakh-Daghestanian family. The method involves setting up correspondences among patterns in languages known to be related. Determination of the ancestral pattern may make use of syntactic relics, dialect data, phonological correspondences, the principle of economy, and identification of internal or external sources of innovation, among other resources. Co-occurrence of all parts of a pattern provides the required safeguards in reconstruction of syntax, just as co-occurrence of all parts of a word does in reconstruction of phonology. Reconstruction of syntax is more likely to be successful in language families with less time depth and in ones with complex morphology reflecting the syntax, than in families lacking these characteristics.

1. Introduction

A recent book on linguistic reconstruction summarizes the position of its author regarding syntactic reconstruction as follows:

PRINCIPLES OF SYNTACTIC RECONSTRUCTION

Edited by

GISELLA FERRARESI
Johann Wolfgang Goethe Universität, Frankfurt am Main

MARIA GOLDBACH
University of Oxford

2008