Part II
Methods for Studying Language Change

1 The Comparative Method

ROBERT L. RANKIN

The comparative method is a set of techniques, developed over more than a century and a half, that permits us to recover linguistic constructs of earlier, usually unattested, stages in a family of related languages. The recovered ancestral elements may be phonological, morphological, syntactic, lexical, semantic, etc., and may be units in the system (phonemes, morphemes, words, etc.), or they may possibly be rules, constraints, conditions, or the like, depending on the model of grammar adopted. The techniques involve comparison of cognate material from two or more related languages. Systematic comparison yields sets of regularly corresponding forms from which an antecedent form can often be deduced and its place in the proto-linguistic system determined. In practice this has nearly always involved beginning with cognate basic vocabulary, extraction of recurring sound correspondences, and reconstruction of a proto-phonological system and partial lexicon.[1]

1 The Goal of the Comparative Method

Kaufman (1990: 14–15) states: "The central job of comparative-historical linguistics is the identification of groups of genetically related languages . . . [and] the reconstruction of their ancestors." He continues (p. 31): "it should be clear that while archeology, genetics and comparative ethnology will help flesh out and provide some shading in the picture of pre-Columbian . . . Man, it is comparative linguistic study, combined with some of the results of cross-cultural study, that will supply the bones, sinews, muscles, and mind of our reconstructed model of early folk and their ways." Linguistic reconstruction is one of our primary tools for learning about the prehistoric past. In many ways it is our best, and this is especially true at time depths where archeology has trouble identifying the ethnicity of its subject matter. Archeology is our best tool for recovering material culture – settlement patterns, dwelling types, tools,

subsistence, and related information – but it contributes much less to our understanding of what archeologists call *ideoculture* and *socioculture*.[2] These are areas in which linguistic reconstruction is potentially much more productive. The comparative method is our primary tool for arriving at such linguistic reconstructions.

   While the principal goal of most linguists who are also historians has been to learn as much as possible about earlier languages and about past cultures through their languages, other branches of linguistics have benefited a great deal from the by-products of comparative work. Many who are philosophically synchronic linguists have looked to comparativists to inform them about the possible types and trajectories of language change. The study of attested and posited/reconstructed sound changes has played an important role in the formulation of notions of naturalness in phonological theory, and modern theories of markedness and optimality often rely, implicitly if not explicitly, on historical and comparative work. The same can be said for the establishment of the grammaticalization clines that result from much morphosyntactic change.[3] Our understanding of the complexities of the synchronic polysemy often associated with grammaticalization is informed by the study of attested and posited intermediate steps in their histories. To a lesser extent the same may be said of semantics and semantic change. But such essentially typological studies may not be considered by some historical linguists to be one of the goals of the comparative method per se. They are important bonuses that result from a consistent and thorough application of the method to families of languages, but they will not receive much additional coverage in this chapter.

2   Why Does the Method Work?

The comparative method relies on certain characteristics of language and language change in order to work. One important factor is, of course, the arbitrariness of the relationship between phonological form and meaning (non-iconicity). To the extent that the linguistic sign is arbitrary, sound change can operate unhindered and will normally be rule governed. Where iconicity is present (in sound symbolism, nursery terms, onomatopoeia) normal change may be impeded or prevented.[4] Linguists therefore avoid comparison of such items until the basic correspondences among the languages being compared are understood.

   A second factor is the regularity of sound change.[5] To the extent that sound change is regular, we can, with the help of phonetics and an understanding of sound change typology, work backward from more recent to earlier stages. And indeed most phonological change ends up being change of articulatory habit, that is, rule change, and thus ultimately regular. Fairly salient interference is required in order to breach such regularity.

Recognition of regularity and of the role it plays in reconstruction has been considered both a strength and a weakness of Neogrammarian linguistics. It has most often been considered a strength because, of course, without ultimate regularity there can be no phonological reconstruction. It has sometimes been considered a weakness of the Neogrammarian position, however. Beginning with Hugo Schuchardt (1885) and continuing until the present, analogical extension of changes and the pervasive role of dialect borrowing with resultant diffusion of forms has occupied many linguists, dialectologists, and creolists.[6] Copious amounts of ink have been spilled in discussions of the extent to which the Neogrammarian "hypothesis" is really "true." But, as most Indo-Europeanists have always known, the exceptionlessness of sound change was not so much a hypothesis for Neogrammarians as it was a definition. Those changes that were sweeping and observed after several centuries to be essentially exceptionless qualified for the term *Lautgesetz* (sound law), while changes that seemed to affect only particular words or groups of words did not so qualify.[7]

Most linguists believe that change in articulation begins as a geographically and/or socially limited but regular, unconscious, and purely phonetic process, which then spreads by several different mechanisms, including dialect borrowing (social and otherwise) and rule formation during the language acquisition period in children, until regularity over a greater area is achieved. A perceived dichotomy in the methods of diffusion has variously been described as *sound change* versus *borrowing and analogy* (the terms traditionally favored by most comparativists), *primary* versus *secondary* sound change (Sturtevant 1917: chs 2 and 3), *actuation* versus *implementation* (Chen and Wang 1975), and others, although the pairs of terms do not always correspond 100 percent. The precise extent to which ultimate regularity results from, or is independent of, dialect borrowing doubtless varies from language family to language family.[8] As a practical matter, comparative linguistics generally involves compilation and analysis of the reflexes of sound changes that occurred, diffused, and regularized long ago. Within comparative Indo-European linguistics the problem of variability within sets of reflexes has not been acute. Whatever the mechanisms that contribute to ultimate regularity in particular instances, its existence, although sometimes obscured by diffusion and analogy, is not seriously disputed and is of primary importance for operation of the comparative method.

3   Family Tree and Wave Diagrams of Language Relationship

The comparative method was developed for the study of the well-defined and quite distinct linguistic subgroups of Indo-European, so comparanda there have tended to be similarly well defined. Obviously such definition is not

always possible (and some might argue that it seldom is). Clearly there are language families (e.g., northern Athabaskan, Muskogean, some Austronesian) in which some unique subgroups are difficult to specify with clarity.[9] This has given rise to another red herring frequently encountered in discussions of the comparative method, namely the assumption that it must be based on some inflexible notion of *Stammbaumtheorie*. And here again much ink has been spilled by amateurs wondering which theory, the family tree (*Stammbaum*) or the supposedly competing wave theory (*Wellentheorie*), is "true."[10] Both are true. But they are oversimplified graphic representations of different and very complex things, and it seems hyperbole to call them theories in the first place. One emphasizes temporal development and arrangement, the other contact and spatial arrangement, and each attempts to summarize on a single page either a stack of comparative grammars or a stack of dialect atlases. Neither is a substitute for a good understanding by the linguist of both the grammars and the historical, social, and geographical interrelationships found among his or her target languages. The comparative study of languages or dialects that are arranged in chains or other adjacent or overlapping continua is certainly a challenge, but it is a challenge to the linguist rather than to the method.[11]

4   Uniformitarianism

Lastly, the method also relies on the more general scientific notion of *uniformitarianism*, here the understanding that basic mechanisms of linguistic change in the past (e.g., phonetic change, reanalysis, extension, etc.) were not substantially different from those observable in the present. Most linguists operate with this as a given and it has not received detailed treatment in most studies of language change, but without the assumption of uniformitarianism, reconstruction would not be possible (Allen 1994: 637–8).[12]

5   Steps in Application of the Comparative Method

The comparative method proceeds in several recognizable stages, which in practice overlap considerably. Internal reconstruction is useful when applied to the daughter languages initially and may also be practiced at various points along the way (see Ringe, this volume). There is relatively little in the way of strict ordering of procedures. A relatively full comparative treatment of a family of languages would include most or all of the following, beginning with the discovery of cognates, both lexical and morphological, and concomitant confirmation of genetic relationship.[13] Most of these topics are discussed below.

i   *Phonological reconstruction*:
   a   Extraction of phonological correspondence sets.
   b   Classification of sets by articulation (place/manner).
   c   Preliminary reconstruction of proto-phonemes.
   d   Distributional analysis of proto-phonemes; collapse of complementary sets.
   e   Assignment of phonological/phonetic features to proto-phonemes (the reality debate).
   f   Possible adjustment of reconstructions in line with typological considerations (in Indo-European, issues such as laryngeal theory and, more recently, glottalic theory).
ii   *Reconstruction of vocabulary* per se:
   a   Reconstruction of structured lexical and semantic domains within vocabulary such as kinship or numeral systems, in which reconstruction of certain members of the system may enable additional reconstruction of less well-attested or even missing cognate sets within the same system.
   b   Possible semantic reconstruction of cells in a structured matrix even if lexical material is lacking.
iii   *Reconstruction of morphology to the extent that morphological reconstruction is merely an extension of phonological and lexical reconstruction*:
   a   Paradigmaticity may materially aid in reconstruction where cognate morphemes are poorly attested.
iv   *Reconstruction of syntax*.

5.1   Cognate searches

In order to undertake any comparison at all one must have something to compare. The search for cognate vocabulary is, oddly enough, usually the single most challenging task facing the comparativist. If the linguist has already established the existence of a genetic relationship between two or more languages (see Campbell, this volume), she or he has already located a certain number of important cognates. These are normally searched for among the most basic of inflectional forms and among the most basic vocabulary items. A list of 100 or 200 basic words is often used initially in cognate searches, the idea being that basic concepts are the least likely to have been borrowed. We have learned that any such list should be used with care, however, and then only after careful attention to known areal phenomena in the zone where one is working. In English around 10 percent of such basic vocabulary is borrowed, mostly from French. In East and Southeast Asia, though, it is well known that even the most basic numerals are often borrowed from Chinese. In table 1.1, note that the first four languages are related, while the last three are not. Such known vulnerabilities should obviously be considered and avoided, something that was often not possible a century ago but which is often possible today.

Table 1.1 Basic numerals in East Asian languages illustrating both cognates and loanwords

Numeral	Tibetan	Chinese I	Chinese II	Burmese	Japanese	Korean	Thai
'one'	ciq	i	či	tiʔ	iči	il	ʔet
'two'	ñīi	ər	nəŋ	hniʔ	ni	i	(sɔ̀ɔ́ŋ)
'three'	sūm	san	sã	ˈθouŋ	san	sam	sàám

Atypical syllable structures, clusters, and marginal phonemes are obviously suspect also.

Regularly corresponding phonemes in basic vocabulary and in basic grammatical formants (if typology permits, preferably in paradigms) are the goal. The affixal morphology searched should be largely inflectional, as derivational morphology is borrowed relatively easily and can wait until basic regularities have been worked out.

5.2 *Phonological reconstruction: comparanda*

The question of comparanda in phonological reconstruction is important and is one of the most underdiscussed questions in the literature: one obviously must know what to compare at all levels. The degree of abstraction of the comparanda used in phonological reconstruction is significant and can have important implications, both for relative ease of application of the comparative method, and for the accuracy of reconstructions. Technically one could compare transcriptions of virtually any degree of abstractness from a tight phonetic notation that reveals the greatest degree of lectal and individual variability to a highly abstract underlying and underspecified phonological representation in which only the non-predictable features are noted. There are good reasons to choose neither of these extreme alternatives, however.

It is not the primary job of the comparativist to document superficial dialect variation, and subphonemic variability should usually be factored out of transcriptions used for comparison (although it can be very valuable in charting sound change trajectories). Variable dialect data turn out to be much less variable if they are first phonemicized.[14] Thus, even though the comparative method is in principle capable of dealing with any number of variant forms, it is simpler to introduce a degree of abstraction that eliminates as many as possible without compromising necessary distinctions. Degree of phonological abstraction then becomes a question the comparativist must address.

The usual way in which the number of comparanda is reduced is to perform a preliminary internal reconstruction on the data of each of the languages to be compared before attempting to use the comparative method. This reduces

(or eliminates) allomorphy and makes further comparison simpler. Phonemicization is an obvious first step in such reduction.

Changes in synchronic phonological theory since about 1960 have clouded the picture somewhat. Only two levels of notation have been significant in most generative phonologies, the underlying phonological and the surface phonetic. We have already eliminated the phonetic as excessively detailed, but the underlying turns out to be unsuitable for comparisons also.[15] This is because the procedures generally used for arriving at synchronic underlying notation, although they often do lead to results that look superficially like reconstructions, can sometimes lead the analyst in an ahistorical direction. The resultant abstract phoneme may look like the results of an internal reconstruction, but internally reconstructed and merely abstract phonemes can differ.

Numerous authors have noted the similarity between the procedures of internal reconstruction and those used for abstracting underlying segments. It is often claimed that the procedures are really the same (e.g., Fox 1995: 210). Both procedures do involve treating allomorphs as cognates (which, internally, they are), but synchronic phonological theory places a high value on *productivity*, which may in turn be the result of analogical change, whereas internal reconstruction stresses the importance of irregularities, often so rare that synchronic phonologies would merely assign them an exception feature of some kind. The least productive and most irregular alternations are often the most revealing for the comparative linguist, but the most productive and least irregular alternations are the ones that best serve the synchronist. So the two methodologies may lead in different directions and should be kept distinct.

So it would seem that the comparativist must begin with something not far removed from the conservative notion of surface phonemes, and that abstraction beyond cover symbols for the most automatic of alternations must be treated as an avowedly historical procedure and justified by a careful and explicit application of internal reconstruction.[16] The use of some variety of surface phonemes as comparanda at once eliminates the most superficial levels of lectal variation while preventing a confusion of internally reconstructed with merely underlying forms.

5.3 Correspondence sets and phonological reconstruction

Phonological and lexical reconstruction proceeds according to the procedures outlined above. Take, for example, the cognate sets from several Siouan languages shown in table 1.2.[17] The sets of stop correspondences that can be extracted from these are shown in table 1.3. Major subgroups here are separated by a solid line and minor subgroups within the central Mississippi Valley subgroup by a broken line.

The comparative method requires that these sets recur regularly in a great many other basic Siouan words. With that requirement fulfilled, we see a

Table 1.2 Cognate sets from Siouan languages

Language	'fire'	'four'	'blue/green'	'throw'	'mark'	'bison'
Crow		šo:pá	šu:-	kuss-	-ka:xi	bišé:
Hidatsa		to:pá	tóʔo-		-ka:xE	wité:
Mandan	pte	top	toho-	-kų:te	-kax-	ptį:
Dakotan	phéta	tópa	tho	khuté	káɣA	pte
Winnebago	pe:č	jo:p	čo:		ga:x	če:
Ioway-Otoe	phé:je	do:we	tho	khú:je	gá:xe	čhe:
Dhegihan:						
Omaha	ppé:de	dú:ba	ttúhu	kkí:de	gá:ɣe	tte
Kansa	ppé:je	dó:ba	ttóho	kküje	gá:ɣe	čče
Osage	hpé:ce	tó:pa	htóho	hküce	ká:ɣe	hce
Quapaw	ppétte	tó:pa	ttóho	kkítte	ká:ɣe	tte
Biloxi	peʔti	topa	tóhi	kité		
Ofo	aphéti	tópa	ithóhi			
Tutelo	pé:ti	to:pa	oto:	kité:		

Table 1.3 Sets of stop correspondences from table 1.2

Language	I	II	III	IV	V	VI	VII	VIII
Crow		š	k	p	š	k	š	
Hidatsa		t		p	t	k	t	
Mandan	p	t	k	p	t	k	t	t
Dakotan	ph	th	kh	p	t	k	th	t
Winnebago	p	č		p	ǰ	g	č	č
Ioway-Otoe	ph	th	kh	w	d	g	čh	ǰ
Dhegihan:								
Omaha	pp	tt	kk	b	d	g	tt	d
Kansa	pp	tt	kk	b	d	g	čč	ǰ
Osage	hp	ht	hk	p	t	k	hc	c
Quapaw	pp	tt	kk	p	t	k	tt	tt
Biloxi	p	t	k	p	t			t
Ofo	ph	th		p	t			t
Tutelo	p	t	k	p	t			t

pattern emerging among the correspondence sets (in spite of the fact that some of the sets here are incomplete because cognates have not been found in some subgroups). There are two sets of labial stops, two sets of dentals (we shall return to sets VII and VIII momentarily), and two sets of velars. And where they differ, they seem to differ by a feature of aspiration or gemination. If we assume that the gemination is secondary and comes from total assimilation of the *h* portion of the stop to what it is adjacent to (i.e., *hC > CC* in the Dhegihan subgroup), then it appears probable that we should reconstruct an aspirated and a plain (non-aspirated) set of stops for each of the three places of articulation. To do this, however, we must answer several questions. Were the Proto-Siouan aspirates pre-aspirated, *hC*, or post-aspirated, *Ch*? Were the plain stops voiced or voiceless? What kind(s) of general evidence should we look for and consult in answering these questions?

5.4   Geographic distribution and reconstruction

Meillet (1964: 381, 403) required that cognates be present in at least three distinct subgroups in order to qualify for reconstruction within Indo-European. Obviously the applicability of such a requirement will vary with the size of the language family. Within Siouan, post-aspirated stops are found in Dakotan, Ioway-Otoe-Winnebago, and Ofo. Pre-aspirated or geminated stops are found only in the Dhegiha subgroup (Omaha, Ponca, Kansa, Osage, and Quapaw) of Mississippi Valley Siouan. So the type of aspiration found in Siouan cross-cuts well-established subgroup boundaries. Ordinarily, distribution of post-aspiration in two or more major subgroups would be a pressure toward reconstruction of that feature. Not only are pre-aspirates in the minority but they are found only in one small subgroup of central Siouan. In this instance, however, it is instructive to note that additional factors intervene and cause Siouanists to reconstruct the minority preaspirates.

There are synchronic rules in Dakotan, Ioway-Otoe-Winnebago, and Ofo which reverse *h-C* sequences when they occur in clusters at a morpheme boundary. So Dakotan *mąh-* 'earth' + *-ka* 'nominalizer' gives [màkha]. The clinching argument is that there are additional, conflicting cognate sets which contain real post-aspirated stops. A few of these may represent borrowings, but if they are borrowings they are very old as they are represented in virtually all Siouan subgroups. They include 'cow elk, grizzly, mosquito', and numerous other terms. These problems are discussed in Rankin (1994) and in Rankin et al. (1998). Lastly, there are post-aspirates that arise morphophonemically, and they behave differently from our pre-aspirated sets. So it is the minority pattern, *hC*, that is reconstructed, and, as often happens in comparative linguistics, the qualitative evidence outweighs the quantitative. These cases also serve to illustrate the importance of the comparativist's knowing the synchronic grammars and phonologies of his or her target languages.

The second group of stop correspondence sets shows generally similar articulations but lacks the aspiration. Several languages voice the simplex stops,

but voicing is inconsistent even within the smallest subgroups, and philological evidence of variation in the transcription of voicing in the eighteenth and nineteenth centuries strongly suggests that it is recent.

So the comparative method leads us to reconstruct three places and two manners of articulation for Proto-Siouan stop consonants. Given the above discussion, these are fairly transparently *hp, *ht, *hk and *p, *t, *k. Nothing that could be called guesswork was involved.

5.5   *Complementarity and reconstruction*

Returning to sets VII and VIII, we see that these groups overlap III and IV, the *ht and *t sets, somewhat. Examining all such cognate sets it emerges that sets III and IV nearly always precede non-front vowels, while VII and VIII nearly always precede *i* or *e*. Thus III and VII are complementary, so are IV and VIII, and we are entitled to collapse them into two sets and reconstruct a single stop for each, thereby deriving one set as a positionally determined "alloset"of the other. Such distributional analysis and amalgamation of sound correspondence sets is what Hoenigswald (1950) called the "principal step in comparative grammar."

5.6   *Naturalness and typology in reconstruction*

Linguists often appeal implicitly or explicitly to sound change typologies and the notion of naturalness when deciding among several possibilities for reconstruction. In the complementary Siouan sets, we are dealing with a relatively shallow time depth and a common and relatively transparent palatalization of dentals preceding front vowels. It is important to note, though, that our reconstruction, however easy, is actually being informed by an understanding of phonetic naturalness that, in turn, is derived historically from the combined knowledge of the sound changes that have occurred in hundreds of languages worldwide.[18] It was largely the study of such changes that indicated to early phoneticians such as Eduard Sievers, Paul Passy, and Maurice Grammont just where they would need to search for the kinds of articulatory and acoustic explanations to which we appeal today. One must know what requires explanation before one may explain it. The study of sound change has consistently provided the raw material for phonological typologies and phonetic explanation. And comparativists, in turn, use these constructs in their hypotheses about sound change trajectories and in their reconstructions.[19]

5.7   *Reconstruction of lexicon*

Working from these and other sets (which account for the remaining vowels and consonants in the cognates), we are able to reconstruct entire lexemes for

most of the cognate sets. In a few instances independent derivation within particular subgroups or languages prevents us from reconstructing more than the root morpheme. The reconstructions thus far are Proto-Siouan: *ahpé:te* 'fire', *tó:pa* 'four', *ihtó:-* 'blue/green', *hkú:te* 'throw', *ká:xe* 'make marks', *wihté:* 'bison cow'.

Caution is in order, of course. The examples above were chosen carefully in order to represent fairly what is usually encountered in Siouan languages. These languages abound in simple lexemes of the sort reconstructed here. Even though Siouan is not polysynthetic in structure, there are both nominal and verbal compounds. One of these is a term for distilled spirits: 'fire-water':

Winnebago	pé:ǰ-nį́:
Ioway-Otoe	phéh-ñi
Omaha	ppé:de-ni
Ponca	ppé:de-ni
Kansa	ppé:ǰe-ni
Osage	hpé:te-ni
Quapaw	ppétte-ni

These examples illustrate the danger of reconstructing other than simple lexemes. Each is a compound of native reflexes of *ahpé:te* 'fire' and *wirį* 'water.' But of course the Siouan-speaking peoples did not have distilled liquor until post-contact times, and the compound came about either through parallel innovation, based on the properties of the liquid, or through contact with Algonquian-speaking peoples to the east who had a similar compound (equally non-reconstructible) from which the Siouan could easily have been loan-translated. It could even represent a back-translation by whites of the Algonquian pattern.

5.8   Residual problems in reconstruction

There are certain trends that are not visible from the few examples of reconstruction given above. Let us examine a couple of additional phenomena within Siouan that challenge the comparative method in different ways. The method can be defeated by mergers or loss of phonemes in the proto-language. Often, though, linguists must deal with a certain amount of suggestive residual evidence of phonological split that has been left behind. In Siouan linguistics just such a case is often called the "funny-R problem." There are two, somewhat overlapping, sets of liquids. One is reconstructible as a simple *r*.[20] In the other set we find a number of strengthened sonorants and this set is reconstructed provisionally as *R* (table 1.4).

'Wash' and the many words like it are reconstructed with *r*. But 'Indian potato' and 'beg' show the other resonant set. *R* often seems to occur in a cluster following the reflex of Proto-Siouan *w*, as in 'Indian potato.' If this

Table 1.4   The "funny-R problem" in Siouan linguistics

Language	'wash'	'Indian potato'	'beg'
Proto-Siouan	*ruša	*wi-Ro	*Ra
Mandan	rusaʔ-		
Lakota	yužáža	blo	la
Dakota	yužaža	bdo, mdo	da
Ioway-Otoe	ruya	do:	da
Winnebago	ruža	do:	da
Omaha	ðiža	nu	na
Kansa	yüža	do	da
Osage	ðüža	to	ta
Quapaw	diža	to	ta

Table 1.5   Deictic particles in Siouan languages

Language	'this, here, now I'	'this, here, now II'
Proto-Siouan	*re(ʔe)	*Re(ʔe)
Crow	-le:-	-né:
Mandan	re	
Lakota		le
Dakota		de
Ioway-Otoe		je-
Winnebago		de: ~ deʔe
Omaha	ðé	
Kansa	ye	
Osage	ðe	
Quapaw	de	
Biloxi	de	né-
Ofo		le-
Tutelo	lé:	né:

were true everywhere, we could collapse the sets, but in numerous other cases there is no trace of *w, which is from an old nominal prefix, or evidence of any other cluster. Yet it seems that *R is somehow related to *r because of their partial complementarity and because of the sets of deictic particles shown in table 1.5, in which the semantic necessity of some sort of historical relationship is clearer. Note that in some languages doublets for these deictics are common.

At the moment there are enough cases of *r and *R in apparent contrast that Siouanists feel constrained to reconstruct both. Yet there is a strong suspicion that *R was secondary and that it developed from *r in a cluster with a preceding resonant or glide. Mandan shows a Ør cluster in one or two such cases, but in many cognate sets (such as "beg", above) there is simply no trace of the hoped-for cluster, and if we follow the comparative method strictly we are left unsatisfied. New data or internal reconstruction may help resolve the question.

5.9 The question of phonetic realism in reconstruction

Since the principle of distinctiveness became dominant in phonology, the goals of comparativists have revolved around reconstructing those segments or features deemed to be distinctive in the proto-language. We often end up having to reconstruct feature by feature. The product is admittedly an abstraction and thus not "pronounceable," and most modern practitioners eschew delving into allophony even where it might be possible. In practice most linguists seem to have quite a bit of faith in their constructs and would be willing to vouch, at least informally, for their phonetic manifestation(s). Obviously this cannot always be true, though, and the Proto-Siouan *r/*R distinction is a case in point. The phonetic feature by which these phonemes differed is unknown, so in this instance, even among linguists who "hug the phonetic ground," *R can only be a cover symbol for a divergent correspondence set. It is reconstructed the same as *r except for one feature, but that one feature (possibly assimilated from an adjacent consonant or glide, since disappeared) remains phonetically elusive.

5.10 Distributional statistics and problems in reconstruction

Part of tying up loose ends in comparative reconstruction involves looking closely at the language one has reconstructed for hints about older changes and deeper alternations. We have seen that we must reconstruct an aspirated and a plain series of stops in Proto-Siouan. After reconstructing about a thousand lexemes an unexpected pattern emerges, however. Virtually all of the pre-aspirated stops reconstructed fall in accented syllables in the proto-language. Pre-aspiration apparently did not occur in unaccented syllables. Plain stops, on the other hand, do appear in Proto-Siouan accented syllables but only a small percentage of the time, perhaps in only about 10 or 15 percent of such stop consonant reconstructions. Words with plain stops in accented syllables include some very basic items, however: "four" and "make marks" in our small sample alone.

What should comparativists make of such distributional skewing? Most Siouanists believe it suggests that in pre-Proto-Siouan there was most likely an aspiration rule: $CV' > hCV'$ (where C was any stop). This cannot be proved conclusively, however, because it is not supported by alternations. Siouan languages utilize prefixes in inflection, and since affixation generally causes accent to move to the left as prefixes are added, we would expect aspirated and unaspirated stops in root morphemes to alternate in paradigms. But they do not.[21] It seems likely that the putative pre-Proto-Siouan aspiration rule operated at one time, but then ceased to function actively in the language, leaving numerous roots with (pre-)aspirates frozen in place. This would have to have involved the analogical extension of the aspirated allomorphs (of verbs especially) to all contexts. The distantly related Catawba language offers no help. Catawba lacked any trace of aspiration. The comparative method is at an impasse here, as is internal reconstruction (because alternations are wanting). Only the distributional pattern of Proto-Siouan aspirates tells us that something is amiss. So in this case also, strict application of the comparative method leaves an unsatisfying residue.

6   Semantic Reconstruction

Lexical reconstruction of course involves more than just phonology; it must also involve semantics. And if the reflexes of a proto-morpheme or lexeme are semantically diverse, reconstruction can be quite difficult. In some instances the only solution is to reconstruct a meaning vague enough to encompass all the descendant forms or to reconstruct polysemy. In other cases it is sometimes possible to appeal to other links in a greater lexical system or semantic domain. Kinship systems (like systems of inflectional affixes: see below) often lend themselves to a kind of semantic componential analysis which may produce "pigeonholes" that aid semantic reconstruction. In other cases, known or inferable history may aid reconstruction. In the Siouan cognate set labeled 'throw' (table 1.2), the semantics of the descendant forms is more complex than my label suggested. The actually attested meanings of the reflexes in the individual languages are as follows: Crow and Mandan 'throw'; Dakotan, Ioway-Otoe, Omaha, Kansa, Osage, Quapaw 'shoot'; Biloxi 'hit, shoot at'; Tutelo 'shoot.'

In modern times, in the (vast majority of the) languages in which this term is translated 'shoot,' this verb has normally meant 'shoot with a firearm,' but in earlier times, of course, it meant 'shoot with an arrow.' Here, archeology becomes the handmaiden of linguistics. We know, thanks to a great deal of work by North American archeologists, that the bow and arrow appear in sites in the Illinois Country and adjacent areas west of the Mississippi River only in about the sixth century AD, long after Proto-Siouan had split into its major subgroups. Before that there were no bows in Siouan-speaking areas

and people hunted using atlatl darts propelled by throwing sticks. Knowing this, it is a simple matter to reconstruct the semantic progression: earlier 'throw,' originally applied to atlatls, became later 'shoot,' applied to bows and finally to guns. 'Throw,' attested only at the northwest corner of Siouan-speaking territory, virtually has to be the older meaning. Semantic reconstruction most often must be done on a word-by-word basis.

7 Morphological Reconstruction

In morphology, internal reconstruction deals with the comparison of allomorphs, and the comparative method should ordinarily not have to deal with allomorphy. Comparative reconstruction must then rely pretty strictly on the comparison of cognate morphemes. The requirement that comparative reconstruction of common affixal morphology be based on established sound correspondences is pretty much taken for granted, although there have been attempts to reconstruct grammatical categories from the comparison of analogs rather than cognates. This would never be considered in lexical reconstruction, however, where comparison of French *maison* with Spanish, Portuguese, Italian *casa* would be unthinkable. Some have found such comparisons more tempting in morphology where morphotactics (fixed common position in templatic inflectional morphology) may offer limited support for such reconstruction. For example, in the Mississippi Valley Siouan subgroup there is a pluralizing morpheme, *-api*, that occurs as the first suffix with verbs (aspect and mood morphology follows this affix). In the related Ohio Valley Siouan subgroup (Biloxi, Ofo, and Tutelo) the analog (not cognate) of *-api* is *-tu* 'pl.,' and it fills exactly the same post-verbal slot in the template. Is morphological pluralization reconstructible for Proto-Siouan verbs? Most would say not, because the morphemes in the recognized subgroups are not cognate, but it brings up the question of whether or not morphotactics alone may contribute at all to the notion of cognacy or of category reconstructibility.

To generalize these observations, comparison and reconstruction of empty templates are not generally accepted as legitimate. If the morphemic contents of the templates are properly cognate, then reconstruction of the morphology along with its positional restrictions becomes possible. Otherwise a much better understanding of the reasons for lack of morpheme cognacy is necessary before positional reconstruction can proceed.

The comparative method per se does not really provide for morphological reconstruction as distinct from phonological reconstruction. As Lass (1997: 248) puts it, "When 'standard' comparative reconstruction is carried out in morphological domains, it is (if done strictly) only projecting paradigmatic segmental correspondences to the syntagmatic plane." However, "morphs expound categories . . . and genuinely morphological change takes place at the category level." Comparison of morphological categories and paradigms can

create a matrix with cells (pigeonholes) for reconstructed members. This often provides help to the linguist, who then knows roughly what to expect in the way of inventories. If the material in expected/established cells in an inflectional matrix fails to correspond phonologically, however, recovery of the proto-morpheme can be problematic.

Loss of entire grammatical categories can lead to inability to reconstruct large parts of the system. In the morphology of the Romance languages, for example, less than half of Classical Latin inflectional endings are reconstructible. Much of the problem is due to early loss of the Latin passive subsystem, nothing of which is really preserved in the modern languages, and the loss of most (not all) nominal case marking. Almost all of the Latin future tense morphology has also been lost without a trace. Within the active voice, non-future morphology, however, most of the present, imperfect, and perfect categories along with most of the person-number marking system is reasonably well preserved in both indicative and subjunctive moods, and is reconstructible. This may serve to give some hint as to how much morphology might be hoped for in a reconstruction with an approximate 2500-year time depth. Koch (1996: 218–63) surveys morphological change and reconstruction with detailed discussion of methodology for recovering particular kinds of information.

8   Reconstruction at the Morphology–Syntax Interface

Case is a system for marking dependent elements for the type of relationship they bear to their heads. Nominal case is therefore most frequently a characteristic of dependent-marking languages, but pronominal case is much more widespread than nominal case. In many if not most language families, pronominals are fairly easily reconstructed. They occur in paradigms, and distinct cases often may partially share phonological shape. Person, number, and other features found in one pronominal paradigm (e.g., nominative) will normally be found in the others (e.g., accusative, dative, etc.), and reconstruction is thus facilitated. But syntactic and semantic alignment of such systems can present different kinds of reconstructive problems. In Indo-European there are numerous disagreements among languages and subgroups as to which nominal case is governed by particular adpositions. In the Siouan languages there is a split between the pronominal set used as subjects of active verbs (both transitive and intransitive) and the set used as the subjects of stative verbs and transitive objects. Siouan languages thus show active–stative (sometimes called split intransitive) case alignment, and the reconstruction of the borderline between these two categories poses interesting tests for the comparative method. The pronominal prefixes themselves have undergone phonological and analogical changes that need not be discussed here, but otherwise their reconstruction is rather straightforward (table 1.6).

Table 1.6 The active–stative borderline in Siouan languages

Person	Active subjects	Stative subjects and objects
1st	*wa-	*wį- ~ wą-
2nd	*ya-	*yį-
3rd	Ø	Ø
Inclusive	*wųk-	*wa-

Table 1.7 Simple adjectival predicates in Siouan languages

English	Kansa	Osage	Quapaw	Ponca	Dakota	Crow
'be cold'	hníčče	hnícce	sní	usní	sní	alačiši
'be blue'	ttóho	htóho	ttó	ttú	thó	šúa
'be tall'	sčéje	scéce	stétte	snéde	hą́ska	háčka

Stative verbs themselves appear to fall into about three subclasses: (i) a group that we may call adjectival predicates, which are consistently stative morphologically across the entire Siouan language family; (ii) positional verbs, which are consistently active morphologically across the family; (iii) verbs which are morphologically stative but semantically active. It is this last subclass of stative verbs that is the most interesting and that illustrates the problems faced in morphological reconstruction when Lass's (1997: 248) "genuinely morphological change takes place at the category level."[22]

Most simple adjectival predicates, those translatable into English with "to be X" and including attributes, colors, etc., are regularly stative across Siouan. There are probably hundreds of these and the class is clearly reconstructible as almost entirely stative, and this includes instances, like 'be tall,' in which cognacy is not 100 percent. In other words, this large subclass seems semantically defined (table 1.7).

A small class of exceptions is also well defined and reconstructible, namely the positionals and an existential verb. Cognacy within this set is high, and these are all intransitive and morphologically active, though semantically stative (table 1.8).

But there are numerous additional intransitives that are semantically active but morphologically stative in at least several of the languages. They present an interesting problem in morphological reconstruction because case alignment is not consistent across Siouan. In table 1.9, I eschew particular forms and note only whether the verbs are cognate (C) or non-cognate (NC) and morphologically active (A) or stative (S).

Table 1.8   Exceptions to table 1.7

English	Kansa	Osage	Quapaw	Ponca	Dakota	Crow
'be sitting'	yįkhé	ðįkšé	nįkhé	nįkhé	yąká	dahkú
'standing'	kháhe	tháhe	tháhe	tháhe	(hą́)	á:hku
'lying'	žą́	žąkšé	žą́	žą́	yųká	ba:čí
'be alive'	nį́	nį́	nį́	nį́	ni	ilí

Table 1.9   Verb cognacy and activity in Siouan languages

English	Kansa	Osage	Quapaw	Ponca	Dakota	Crow
'fall down'	C/S	C/S	C/S	C/S	C/S	NC/S
'ache, hurt'	C/S	C/S	C/S	C/S	NC/S	C/S?
'recover'	C/S	C/S	NC/S	C/S	C/S	NC/S
'perspire'	C/S	C/S	C/S	NC/S	NC/S	NC/S
'tell lies'	C/S	C/S	C/A	NC/A	NC/S	NC/S
'die'	C/A	C/A	C/S	C/A	C/S	C/S
'belch'	C/A	C/A	NC/A	NC/S	NC/S	NC/A
'forget'	NC/S	NC/A	C/A	NC/A	C/A	NC/A
'cough'	C/A	C/A	C/A	C/A	C/A	NC/A

Stativity decreases descending the chart, but note that there seems to be relatively little correlation with cognacy of the verb roots. The distribution of the data here, along with a general lack of cognacy of the Crow forms, suggests that a morphological shift from active to stative marking of experiencer subjects has been an ongoing process within Siouan.[23] In summary, it seems probable that:

i   Adjectival predicates were consistently stative in Proto-Siouan. The only subclass of exceptions were the positionals and 'be alive.'

ii   A very few semantically active verbs may have been marked statively in Proto-Siouan. These include 'fall down, ache' and perhaps a few others with experiencer subjects.

iii   The presence of the few experiencer statives created a new model that has served to extend the category to different degrees and with different verb roots in all of the modern Siouan languages. In some cases innovations can be traced to subgroup nodes, but in many instances the switch in case alignment for a particular verb affects only single languages in diverse subgroups. While most verbs seem to have gone from active to stative, in

a few instances there is evidence of passage from stative to active.[24] Our conclusions here are rather general: specifying precisely which semantically active verbs had stative morphology in Proto-Siouan is difficult because of lack of cognacy (especially of the Crow forms) within the group. Nevertheless, comparative linguistics give us at least some perspective on this ongoing change.

9   Syntactic Reconstruction

If comparanda can sometimes be controversial in morphology, they are very much more so in syntax. Ordinarily the notion of cognacy implies structural entities that correspond regularly in both form and meaning. If either is wanting, cognacy is not achieved. In syntax there are basic problems in both domains. First of all, it is difficult to know just what to consider formal equivalents when comparing syntactic structures (see discussion in Watkins 1976). In phonology one compares phonemes (by some definition), in morphology one compares morphemes. What is the comparable unit in syntax?[25] Second, it should be obvious that the semantic relatedness criterion is simply problematic in many areas of syntax.

In most modern linguistic theories, syntactic structures are generated, not stored in memory. The structures themselves, then, cannot be comparanda in the same sense as words, phonemes, and morphemes are. "Sentences are formed, not learned; morphemes and simple lexemes are learned, not formed" (Winter 1984: 622–3).

Thus the comparative method per se has often been at an impasse in the area of syntactic reconstruction because of a lack of availability of anything like real cognates. Instead, basic typological agreements have sometimes been examined with a view to projecting their existence and accompanying congruities into the past. Central to this enterprise is the *cross-category harmony principle*, according to which head and dependent dyads tend to be arranged in either consistently head-first or consistently head-last order cross-linguistically.[26] As a general reconstructive methodology for syntax this technique cannot be judged a success, since syntactic change can affect a language one dyad at a time, and has often done just that, leaving a language or family full of cross-category disharmonies.

In the Siouan language family, virtually all members are (S)OV (dependent–head) in basic word order, and dependents normally precede their heads at other levels (noun–adposition, adverb–verb, verb–auxiliary, demonstrative–noun, genitive–noun, etc.). Adjectives follow their nouns in Siouan languages, but, as we have seen, Siouan adjectives are members of the subclass of stative verbs and may best be considered heads of their respective constructions. As can be seen below, a purely typological approach would seem to lead us to the conclusion that Proto-Siouan was an SOV language. This would probably be

historically correct, but that is really because all known Siouan languages have SOV word order.[27] If they did not, it does not seem likely that typology would give us the answers we need. Nor can it answer important questions about NP and clause marking in Proto-Siouan:[28]

Crow: iisáakši-m   háčkee-š úuxa-m dappeé-k
           y.-man-HEAD tall-DEF   deer-a kill-DECLAR
           "The tall young man killed a deer."
Lakota: koškálaka hą́ske ki (he) thá wą kté
           young.man tall   the DEM deer a   kill
           "The tall young man killed a deer."
Ponca: núžìga snéde akha ttáxti wį t?éða   biamá
           boy   tall   SUBJ deer a die-CAUS they.say
           "The tall boy killed a deer."
Biloxi: sįtó tudé ta   o   téye
           boy tall   deer shoot die-CAUS
           "The tall boy shot and killed a deer."

    These sentences, most translations elicited by linguists, show closely parallel patterns that are congruent with a Proto-Siouan SOV word order. 'Kill' is a compound of 'die' plus a causative auxiliary in Ponca and Biloxi but is a lexical verb in Lakota, so the proto-language morphology is unclear there. Crow, Lakota, and Ponca require definite articles with the subject, but Biloxi does not, and the articles are not cognate across the other languages, so the origins of that morphology remain unclear also. Case marking for nouns, to the extent that it existed, does not seem to be reconstructible:

Crow: iisáakšee-š áaše kuss-basáa-k
           y.-man-DEF river toward-run-DECLAR
           "The young man is running to the river."
Lakota: koškálaka ki wakpála ektá   íyąke
           young.man the river   toward run
           "The young man is running to the river."
Omaha: núžìga akha wathíška khe   ttáðišą ttąðį biamá
           boy   SUBJ river   the.lying toward run they.say
           "The boy ran toward the river."
Biloxi: sįtó ayixyą mąkiwayą tąhį
           boy bayou toward   run
           "The boy ran toward the bayou."

    Intransitive syntax is entirely SV with postpositions, but the postpositions themselves are not cognate among the subgroups. Still, the existence of postpositions in the proto-language seems very likely. As with transitive sentences, suffixal and enclitic morphology is not cognate and therefore not reconstructible:

Crow: iisáakši-m   úuxee-š   ak-dappée-š háčka-k
           y..man-HEAD deer-DEF REL-kill-DEF tall-DECLAR
           "The young man who killed the deer is tall."
Lakota: koškálaka   wą thá ki kté ki he   hąske
           young.man a   deer the kill the DEM tall
           "The young man who killed the deer is tall."
Omaha: núžìga akha   ttáxti t̓éðe   akha snéde abiamá
           boy   the.SUBJ deer die-CAUS the tall   they.say
           "The boy who killed the deer is tall."
Biloxi: itá   té-ye   yą sįtó tudé
           deer die-caus. REL boy tall
           "The boy who killed the deer is tall."

The relative clause, *who killed the deer*, is preposed to its head in Biloxi, and that is the order expected in an SOV language. In the other languages the relative clause is postposed to its head, possibly in accordance with what typologists call the *heavy constituent principle*, by which longer, more cumbersome dependent elements are often postposed even if head-last order is expected. Nevertheless, the syntactic disagreement renders it very difficult to reconstruct a unique order for relative clauses in the proto-language. Articles and/or demonstratives (Crow -*š*, Lakota *ki he*, Omaha -*akha*, and Biloxi *yą*) serve as relativizers in all the languages, but none is cognate from one subgroup to the next, so no Proto-Siouan relativizer can be reconstructed.

Since this syntactically homogeneous language family contains 16 languages in four major subgroups, spread geographically over thousands of square miles, most Siouanists consider it likely that an SOV word order reconstruction is accurate for Proto-Siouan, probably at a time depth of over three thousand years. And Proto-Siouan probably had most of the other characteristics of OV languages. But note that this has been established by comparing entities that correspond primarily in form and only roughly in meaning. Definitizing and relativizing morphology is not cognate, nor is quite a bit of the substantive vocabulary. The comparative method requires both formal and semantic correspondence. Thus far, examining analogous (not cognate) sentence types and noting typological homogeneity, we have been able to reconstruct only the very broadest outlines of Siouan syntax.

As language families become syntactically less homogeneous, the necessity of using something much closer to the real comparative method clearly asserts itself. Indo-European (along with many other language families) lacks the typological homogeneity that Siouan presented: there are Indo-European subgroups with SOV, SVO, and VSO word order. And since the overall directionality of prehistoric syntactic change cannot be established simply by looking at a synchronic sample (like the Homeric poems or the Vedas) or at historical directionality over just the past two or three millennia, Watkins (1976) adopts the requirement that one compare sentences with analogous formal structure, but he adds the further requirement that they mean the same

thing. Just as we require that cognate words show equations of both form and meaning, he posits a strong requirement that comparable sentences also show equations in both form and meaning. In effect he reconstructs from *cognate sentences* in about as strict a sense as one could imagine in syntax.

And his cognate sentences tend to be from among the small set of exceptions to the general rule that "sentences are formed not learned." Some sentences, of course, are indeed learned rather than generated and are, thus, analogous to simple lexical items. These are mostly formulae of one kind or another. They may include special ways in which people or professions talk about particular subject matter (Watkins selects ancient sports events), proverbs, folk narratives, perhaps poetry (with the obvious caveat that versification often affects syntax), formal legal documents, and perhaps a few other culturally defined styles.

Like Watkins, practitioners of the typological method have also sought expressions that show archaic syntax in order to make use of the cross-category harmony principle. Among the additional sources of relic syntax that have been suggested are comparison of inequality, adpositions, numerals in the teens, pronominal patterns, and certain derivational formations (Lehmann 1976: 172ff).[29]

Both derivational and inflectional morphology are often thought to be sources of archaic syntactic structures. Givón's (1971: 413) claim that "Today's morphology is yesterday's syntax" typifies this view. The idea is that processes of grammaticalization create clitics and then affixes that attach to stems in the order in which they originally occurred as independent words. Thus frozen syntactic constructions are preserved and can be analyzed for ancient head-dependent constructions and congruities, etc. This seems to work well in certain instances; for example, future tense marking in Indo-European, Latin, and subsequently Romance. But in other cases, notably involving compounds and person-number clitics or affixes, it fails. Givón mentions that modern Spanish clitic object pronouns preserve the OV order of early Latin, but a glance at Old Spanish texts shows copious examples of just these pronouns *following* conjugated verbs in the Spanish of the eleventh century.[30] Comrie (1980) finds similar problems in Mongolian. The difficulties seem to arise during the cliticization period, when there are obviously competing principles for placement (Wackernagel's Law phenomena, unidirectionality of permitted affixation in some languages, e.g., suffixation in Turkic, etc.) that can ultimately produce ahistorical orderings. Nevertheless, morphology may be very helpful in syntactic reconstruction provided it is used judiciously and not too closely coupled to inferences derived from the cross-category harmony principle.

Harris and Campbell (1995: 355) and Harrison (this volume) discuss numerous problems associated with the notion that the order of elements within *compounds* routinely recapitulates earlier head-dependent orders. They believe compounds, as a source of information about older word orders, should be generally ruled out.

Intermediate between comparison of the arrangements of the head-dependent dyads favored by some typologists and Watkins's formulaic

"cognate sentences" are the sources of syntactic correspondences suggested by Harris and Campbell (1995: 350ff). While urging caution, they suggest translations, both literary and elicited (sometimes from bilinguals), as possible sources of generated, cognate syntactic structures. This is approximately what I have done in the Siouan sentences discussed above. While not providing "descendant" sister clauses and phrases (like formulaic utterances), such sources can perhaps provide comparable results of "sister rules."

Lehmann (1976: 172) emphasizes some of the difficulties in dealing with translations, pointing out that translations of the scriptures were used in the study of languages like Gothic, Armenian, and Old Church Slavic, but that influence from the source language, Greek in these instances, has been found to be troublesome. Obviously calques are a major problem encountered using translations, but perhaps it is one that can be overcome. Translations would certainly provide comparable material between/among closely related languages. One can easily imagine obtaining nearly identical sentences eliciting the same utterance in, say, Spanish and Italian or Slovene and Serbian. This may be of interest to linguists operating within small language families of relatively shallow time depth, but eliciting translations of the same sentence in Spanish and Irish would yield more syntactic variables than could easily be dealt with. Clearly syntax presents problems that are much more vexing than those usually faced by comparative phonologists.

The primary comparanda of comparative syntax are still being debated, but we should not be surprised to find that different language families and different historical circumstances place different demands on the comparativist. The relative uniformity of the Siouan language family (with its relatively shallow time depth), coupled with the relatively greater syntactic homogeneity found in SOV languages generally, makes comparative syntax there relatively straightforward. In Indo-European, however, with much less syntactic homogeneity to work from (and considerably greater time depth), Watkins (1976) sees a necessity for greater stringency in selecting comparanda. As difficulty increases, he properly tightens his requirements. Some linguists loosen their methodology when faced with difficult problems, voicing the complaint that by sticking to old-fashioned standards one might never make new discoveries. This is basically the position that *necessity confers legitimacy*. But in science necessity does not confer legitimacy.

9.1   The problem of naturalness in syntax

As we have seen, one of the factors that makes phonological reconstruction possible is our fairly thorough understanding of the directionality of sound change in particular environments. We expect sound change to be phonetically natural, at least at the outset, and we expect it often to affect entire natural classes. This frequently makes reconstruction a matter of working backward along well-established trajectories. Our understanding of naturalness in syntactic

change is far less well developed (see chapters by Harris, Lightfoot, and Pintzuk in this volume, as well as those on grammaticalization by Bybee, Fortson, Heine, Hock, Joseph, Mithun, and Traugott). And, in fact, there is little reason to believe that we will ever reach comparable levels of understanding in syntax, because phonetic change is physiologically shaped and constrained by the configuration of the vocal organs and by perception, while syntactic change is not.

The best bets for syntactic reconstruction at this time would seem to be the use of relic constructions, if such can be identified. Working backward along well-established grammaticalization clines and/or syntactic change trajectories may be helpful, again, if sufficient numbers of these can be identified with certainty. Harris and Campbell (1995: 361ff), for example, identify postpositions → case suffixes, modal auxiliaries → modal suffixes, passive → ergative, ablative → partitive as "one-way" morphosyntactic changes. In some instances it may also be possible to take advantage of certain, unambiguous cross-category harmonies. Harris and Campbell concentrate on restricted parts of the word order typology, especially the few apparently conservative characteristics that are consistently SOV-related. These include (pp. 364–6) relative clauses preposed to their heads, and the order Standard–Marker–Adjective in comparisons of inequality. They first establish syntactically corresponding patterns so that reconstruction becomes a matter of determining which pattern is older. Then they concentrate on a single strong argument of the sort mentioned just above.

10   Proto-Language as a Repository for Regularities as Opposed to Irregularities

Most linguists prefer to reconstruct only those features that can be shown to have been systematic in the proto-language. Returning to the Siouan cognate set translated "throw" (table 1.2), we see that no Winnebago cognate was given. In fact there is a Winnebago word, *gu:č*, that closely resembles the cognates in the other languages. Except for the fact that the form begins with *g-* instead of *k-*, it is precisely what we would expect in this set. Most comparativists would judge this exception to be too small to justify reconstructing anything but **hku:te* for the set. Since there are no other examples of this correspondence, and we lack parallel cases with *p/b* or *t/d*, we assume that some interesting but irrecoverable development occurred in Winnebago alone and do not reconstruct a third stop such as **gh* or the like because of this set. We assume the anomaly is internal to Winnebago and not that Winnebago retains something lost everywhere else. The difference between our treatment of Winnebago 'throw (= shoot)' and the problem of the two rhotic phonemes, **r* and **R*, is one of degree, however. There are too many instances of **R* without an explanatory environment for us to ignore them, even though we suspect there may have been only a single **r*, with **R* arising in certain kinds

of clusters. We make a conscious decision to exclude a single Winnebago form that contains a unique sound correspondence, preferring to reconstruct only what is systematic.

Of course inconvenient residue can be very important and should never be dismissed out of hand or simply hidden away. The celebrated case of Verner's Law illustrates clearly the fact that a closer examination of residual cases that seem to be exceptional can lead to important discoveries that serve not only to explain the data of particular languages or language families but also to reinforce our understanding of basic sound change regularity.

Comparativists are sometimes accused of reconstructing completely uniform proto-languages – agglutinating languages without morphophonemic alternations, without variation, and without irregularities. This is simply not a serious criticism; the shape of our reconstructions is most often a consequence of our preference for regarding proto-languages as repositories for systematicity, not idiosyncrasy, but it is also a consequence of insisting on pushing internal reconstruction as far as possible.[31] This does not mean that we believe in the perfect uniformity of proto-languages. Every serious comparativist understands that, doubtless, there were older irregularities, morphophonemic alternations, and dialects; we simply reconstruct as far as we can and no farther. Proof of older fusion, variability, or idiosyncrasy is simply beyond our reach at some point.

11   Temporal Limits on the Comparative Method

The above discussion does raise an interesting question. Both phonological and analogical change erode languages constantly. Over time, reanalysis and extension can alter the most basic syntactic patterns, and an SOV language may take on an entirely different word order and set of accompanying cross-category harmonies. Lexicostatistics has shown that basic cognates shared by pairs of languages undergo attrition at a relatively common rate.[32] These factors, taken together, will tend over time to render our methods of reconstruction less effectual and finally ineffectual. If cognate attrition takes place at somewhere in the vicinity of 20 percent per millennium, and we depend on cognates for lexical and phonological reconstruction, the comparative method will be useless for recovering information within a family of languages in a period of something less than 20,000 years. Adding other phonological and morphosyntactic change to cognate loss, we may count on significantly less than this amount of time. Just how much is a matter of debate. There is no consensus on just what the temporal limits really are, but well-studied language families such as Indo-European, Uralic, and Afro-Asiatic suggest that our methods may be valid to a time depth of at least around 10,000 years.[33]

The productivity of the method simply trails off as availability of comparanda declines over time. At some point linguistic relationships may yet

be recognizable, because of retained idiosyncratic morphological patterns of the sort that Meillet (1925) delighted in, or multidimensional paradigmaticity of the sort discussed by Nichols (1996), but the ability actually to reconstruct may be lacking. We find this situation to a degree in Algonquian-Ritwan (Goddard 1991), where there is strong paradigmatic evidence for genetic relationship and a certain number of clear lexical cognates but little possibility of fleshing out details of the proto-language.

Overall, however, the comparative method is arguably the most stable and successful of all linguistic methodologies. It has remained essentially unchanged for over a century. This is not because comparative linguistics has faded from view or is less important than it was a hundred years ago. Quite the opposite: its principles have withstood the tests of time and the onslaughts of its critics. The reconstruction of Proto-Indo-European stands as a monument to the very best of nineteenth-century intellectual achievement. In the twentieth century, the comparative method was shown by Bloomfield and others to be equally applicable to non-written languages in diverse parts of the world. Much linguistic reconstruction remains to be done, and if we maintain the integrity of the comparative method, we will be able to do it.

NOTES

1   Here I refer only to reconstruction. Grammatical correspondences have often been the feature that first established genetic relationships beyond doubt. For example, Sir William Jones's oft-quoted statement about Sanskrit, Greek, and Latin refers to the systematic correspondences in their grammars.

2   I do not mean to imply that archeology cannot contribute outside of areas of material culture, only that linguistics is a complementary and often superior tool in the non-physical domains. I have also ignored here the increasingly important contributions of physical anthropology in the study of prehistoric movements and relatedness of peoples, determination of their diet, etc. A synthesis of linguistic, archeological and physical anthropological information is ultimately necessary.

3   See also Hopper and Traugott (1993) and chapters by Bybee, Fortson, Heine, Hock, Joseph, Mithun, and Traugott in this volume.

4   Since, with imitative vocabulary, there is never a necessary historical connection between the onomatope at one stage and the ostensibly "same" one at a later stage. Onomatopes can be reinvented at any time and by any generation.

5   A detailed discussion of sound change is found elsewhere in this volume (see the chapters by Guy, Hale, Janda, Kiparsky, and Ohala). There are a dozen different definitions of the term *sound change*, however, so I feel it is important to include a brief discussion of the phenomenon here. Much of the ink that has been spilled debating the nature of sound change could have been saved simply by not applying one linguist's definition to another

linguist's work, especially if they were not contemporaries.

6   Schuchardt (1885) in fact claimed that most of what the Neogrammarians saw as sound change was "rein lautliche Analogie," purely phonetic analogy, which affected single words or environments at a time (Keith Percival, pers. comm.).

7   After more than thirty years of redefining dialect borrowing as "sound change" (Labov 1963, esp. 1965: 272), Labov (1994: 440ff), citing Hoenigswald (1978), acknowledges this truth about the Neogrammarian position. See also Lass (1997: 134) for discussion of this issue. A particularly good example of "straw man" discussion of the Neogrammarian position is Postal (1968: 231–60).

8   Hoenigswald (1960: 73) went so far as to say that "viewing sound change as a special case of (total) dialect borrowing . . . does no . . . violence to (the) facts; it accounts both for the suddenness of phonemic change and for its regularity and requires few particular assumptions beyond that of the existence of subphonemic variation in the speech community – an assumption in perfect keeping with observed data." This view characterizes the better-elaborated position taken later by Labov (1963, 1965). Labov (1994: 470f, 541ff) clarifies his earlier position and tries to sort out contexts in which regularity operates according to Neogrammarian principles and those in which lexical diffusion is more likely to be found. Labov (1994) is probably the best and most complete discussion of the problems (and pseudo-problems) to date.

9   For example, Malcolm Ross, in lectures given at the 1997 LSA Linguistic Institute, divides much of Austronesian into (i) those languages within a subgroup whose speakers migrated (generally eastward) across the Pacific and can be accommodated fairly easily in a family tree and (ii) what he calls the "stay-at-home languages" whose speakers remained in close contact with each other, forming complex interrelationships that are very difficult to sort out.

10   See Fox (1995: 122–36) for a history and discussion of the pros and cons of the allegedly polar views.

11   Ross (1996: 181ff) presents particularly good examples of these sorts of problems. Although he confines his discussion mostly to Austronesian languages of Papua New Guinea, the model and developments he postulates for PNG are probably not far from what happened in Europe and much of the rest of the world as today's national languages were forming.

12   See also Labov (1994), Fox (1995: 195), Lass (1997), Janda (2001), and the introduction to this volume for further discussion.

13   Discovery and/or confirmation of relatedness is considered an integral part of the comparative method by some linguists. The problem of establishing genetic relationship has become important enough in recent years to require a separate chapter, however. See Campbell's excellent discussion in this volume.

14   It is probably not an accident that the study of lectal variability was perceived as being increasingly important as phonology became more abstract. Until the early to mid-1960s dialect data were often subject to analysis and presentation in terms of surface phonemes. This had the effect of reducing the visibility of variation and probably

of de-emphasizing the social dimension that it presents. It is reasonable to phonemicize comparative data, however. Here it may be looked upon as a form of low-level internal reconstruction.

15   Lass (1997: 250n.) makes this point nicely, but more in relation to morphosyntactic reconstruction (where it is just as valid).

16   I am grateful to Eric Hamp for discussion of the issue of abstraction in choosing comparanda. The importance of the surface phoneme in historical linguistics was recognized fairly early in the generative period by Schane (1971).

17   Siouan languages are native North American languages spoken originally in a broad band extending from the foothills of the Rockies in Canada southeastward to the mouth of the Arkansas River with several outliers as far east and south as Mississippi, Alabama, and Virginia. There are about sixteen Siouan languages documented to various degrees. About ten are still spoken by at least a few persons, about five of these by more than a few hundred. At least six are extinct. These cognate sets are taken from Carter et al. (forthcoming) and some of the discussion recapitulates Rankin et al. (1998). Interpretation of these data is my own.

18   It is important to note that the correspondence sets that comparativists work with are often the "compressed" result of many individual changes.

19   It is worth mentioning here that Allen (1994: 639) recommends also considering what he calls *subfamily typology* when reconstructing. He is referring to what amounts to particular, often recurring, phonetic "drifts" present in individual families or subgroups that may not

be as common outside that group. This might include such persistent processes as palatalization in Slavic or nasal spread in Siouan, for example. Lass (1975) referred to such drifts as "family universals," a term with implications broader than what I wish to convey here.

20   Some Siouanists have preferred to reconstruct *l for this set. Phonetically there is probably little reason to favor one over the other. Several languages have shifted from rhotic to lateral resonants during the historical period, however, so *r is perhaps the better choice. I would like to thank Dick Carter, Wes Jones, John Koontz, and David Rood for their many useful observations on Siouan reconstruction.

21   A possible exception may be Ofo. In the transcription of John R. Swanton (1912), Ofo aspirates seem to alternate, with aspiration often disappearing in unaccented syllables. Swanton only recorded about six hundred words of Ofo, and little was included in the way of verb paradigms that would tell us whether the alternations were systematic. And even if some such alternation is found in Ofo, it may represent an innovation rather than a retention, since even the most closely related languages lack any sign of an aspiration alternation.

22   I wish to thank Fr. Randolph Graczyk, John Koontz, and David Rood for their protracted discussion of these matters with me via electronic mail. They have provided numerous insights, although any errors are my own. Kathy Shea and Parrish Williams provided fresh Ponca data, Randy Graczyk provided Crow data, Quapaw data are from the James Owen Dorsey collection at the National Anthropological Archives of the

Smithsonian Institution, Osage data are from Carolyn Quintero (pers. comm.), and Quintero (1998), Kansa data are from †Maude Rowe. This work has also benefited from exchanges with Regina Pustet about her statistical analyses of this split in Siouan.

23   Other verbs in my sample with mixed active/stative marking across Siouan include "get lost, stumble, lack, tremble, have a cramp, possess, arise, itch, pant, suffer, bleed, get dizzy, shrivel, swell up, tumble, lose, bow head, snore, twitch, stagger, open eyes, remember, have a chill."

24   This treatment avoids discussion of additional, often phonological, mechanisms affecting this change. In Crow, for example, only the pronominal prefix vowel serves to differentiate actives from statives, and these vowels are often assimilated in vowel-initial verb stems, leaving the distinction only in 1st pl. forms (Graczyk, pers. comm.). In Biloxi, a language not dealt with in this section, the active/stative distinction is only maintained in the 2nd person and is phonologically difficult even there. So a number of linguistic factors contribute to some of these category changes. In Dakotan, Omaha-Ponca, Kansa, and Osage, conditioning does not seem to involve much phonology, however.

25   The notion of the *tagmeme* has surfaced from time to time, but there is little if any agreement about its nature among syntacticians. The putative existence of such a unit should, however, serve to underscore the theory-dependent nature of some of the arguments about comparative syntax. Lehmann (1976: 171) emphasizes that there is no agreement on units or their interrelationships at the syntactic level.

26   There is an entire literature on this subject. For a recent survey and discussion of the consequences of using such methodology, see Harris and Campbell (1995: 140, 195ff).

27   Actually, the Dhegihan subgroup of Siouan shows OVS word order rather often, perhaps 10–12 percent of the time (Catherine Rudin, pers. comm.).

28   Fr. Randolph Graczyk and David S. Rood provided me with Crow and Lakota data respectively and helped clarify my understanding of them. Ponca examples are composed from Dorsey (1890), and the author's own Omaha and Kansa language notes were also consulted. These examples may be taken as representative of what one finds in the larger text collections. The Biloxi examples are composites, with certain vocabulary replaced, of more than one sentence from among those found in Dorsey and Swanton (1912). Such composition is not a technique I would recommend in actual reconstruction, but Biloxi is long extinct, and it seemed advisable to use examples containing approximately the same lexemes.

29   Comparing numerals in the teens in language families such as Romance or Siouan, one is hard put to perceive a clearly archaic pattern. And there are competing patterns among other Indo-European subgroups also. Adpositions in Latin were preposed even though the older language tended strongly to SOV word order.

30   There are dozens of examples of this *finite verb+object pronoun* order in just the first couple of hundred lines of the *Poema de mio Çid*.

31   Performing internal reconstruction on a reconstructed proto-language

yields a result that may, of course, represent a collapsing of many centuries of development.

32   One need not embrace the tenets of glottochronology (this writer does not) to accept lexicostatistical demonstrations of fairly regular attrition. It has been shown that some languages are indeed more conservative in retaining basic cognates, while others, of course, have undergone complete relexification. Whatever the rate, loss is continuous.

33   Nichols (1992a: 2–3), for example, posits a practical limit on the comparative method of about eight thousand years or a bit more.