

2 On the Limits of the Comparative Method

S. P. HARRISON

In this chapter, I explore the limits of the comparative method as a tool in comparative historical linguistics.¹ Let me be quite clear about one thing from the outset: for me, the comparative method is the *sine qua non* of linguistic prehistory. I believe that the comparative method is the only tool available to us for determining genetic relatedness amongst languages, in the absence of written records. I believe that prior “successful” application of the comparative method is a prerequisite to any attempt at grammatical comparison and reconstruction.

But the comparative method has limitations, determined by the very properties of the method that make it work:

- i *It has relative temporal limitations.* The more changes related languages have undergone (in general, a function of time), the less likely the method is to be able to determine relatedness.
- ii *It has sociohistorical limitations.* Certain historical situations can have linguistic consequences that vitiate the comparative method.
- iii *It has linguistic domain limitations.* Only certain sorts of linguistic objects can be usefully compared and reconstructed using the method.
- iv *It has limitations of “delicacy.”* Only genetic relationships up to a certain degree of precision or delicacy can be reliably determined using the method.

I discuss each of these types of limitation in turn below.

Disagreements and misunderstandings regarding what the comparative method can and cannot do are a continuing (and, some might say, distracting) leitmotif in comparative historical linguistics. The level of disagreement has often surprised me, and must be attributed to some level of disagreement regarding what the comparative method in historical linguistics actually involves, what its premises are, and what its recognized argument forms are. My first task, then, must be to outline what *I* think the method is.

In section 1, I outline what I see as the goals of comparative historical linguistics. In section 2, I describe how the comparative method serves to realize those goals. The limits and limitations of the comparative method are treated

in section 3. Sections 3.1 and 3.2 discuss the possibility of comparing and reconstructing grammar, both with and without the comparative method. Section 3.3 discusses two circumstances in which the comparative method may fail to recognize genetic relatedness. Section 3.4 is devoted to the unique problems posed by subgrouping. Section 4 considers briefly how the comparative historical linguist can survive the limitations on the comparative method.

1 The Goals of Comparative Historical Linguistics

Identifying the goals of comparative historical linguistics is not a particularly problematic exercise. They are essentially three in number:

- i to identify instances of genetic relatedness amongst languages;
- ii to explore the history of individual languages;
- iii to develop a theory of linguistic change.

Nor, of course, are these goals in practice independent. The identification of instances of genetic relatedness is likely to be a concomitant of the investigation of the histories of one or more related languages. The development of a theory of linguistic change is informed, one trusts, by investigation of the histories of individual languages and language families.

Prehistorians might be satisfied with (or, at least, most immediately interested in) results stemming from the first of these goals, and cultural historians with the second. “True” historical linguists view the third goal as the real prize, the ultimate aim of the exercise. That is certainly how I rank the goals. I want to know whether one can distinguish possible from impossible changes, or, at the very least, probable from improbable. I want to know whether or not there are any constraints on borrowing. I want to understand the engine of language change – how changes begin, and how they move through languages and linguistic communities.

The desiderata of a such a theory of language change were set out quite clearly over a quarter century ago in Weinreich et al. (1968). Some aspects of the research program they outlined have been elaborated in subsequent work. Labov and others have studied cases of language change in progress (cf., e.g., especially, Labov 1994 for discussion and extensive references). The regularity assumption (see below) has been put under scrutiny in their work, and in the work begun by Wang (1969; cf. also Wang 1977) on the so-called “lexical diffusion” of sound change. The notions “natural linguistic process” and “natural linguistic system” (and, derivatively, “natural linguistic change”) have been the focus of linguistic theory from the time Weinreich et al. (1968) appeared. More recently, scholars like Sarah Thomason have given detailed consideration to the limits of borrowing and diffusion.² But, we are still some distance away from a theory of language change.

2 The Place of the Comparative Method

A theory of the sort envisaged in the preceding section is one that, given some synchronic language state S , would tell us what immediate antecedent state(s) P_S^+ could/must have given rise to S . Such antecedent state sets for different languages could then be compared for "best fit," in order to select amongst potential antecedent state candidates (if the theory supplies more than one) and to determine genetic relatedness. In the absence of such a theory,³ however, the comparative method has served the historical linguistic enterprise for well over the past hundred years or so, *because it acts as a stand-in for, or as a first approximation to, a theory of language change.*

The comparative method does at least part of the job of a hypothetical theory of change, but in the reverse order. The primary role of the comparative method is in developing and testing hypotheses regarding genetic relatedness. Its secondary, and subsequent, role (in what might be termed "realist" comparative linguistics) is in recovering antecedent language states through reconstruction.⁴

In order to demonstrate that the members of some set of distinct linguistic systems⁵ are or are not genetically related, one must demonstrate:

- i that there are similarities amongst the languages compared, and then
- ii that those similarities can *best* be explained (or can *only* be explained, depending on just how confidently one wants to present the results of the method) by assuming them to reflect properties inherited from a putative common ancestor.⁶

What permits us to make the move from the observations of cross-linguistic similarity in (i) to the conclusion (ii) that the languages in question are genetically related is an implication (rule of inference, or warrant) that might be stated informally as follows:

The major warrant for genetic inference

If two or more languages share a feature which is unlikely to have:

- i arisen independently in each of them by nature, or
- ii arisen independently in each of them by chance, or
- iii diffused amongst or been borrowed between them

then this feature must have arisen only once, when the languages were one and the same.⁷

A genetic argument, then, consists in the presentation of a set of similarities holding over the languages compared, and a demonstration that these similarities are not (likely to be) the result of chance, nature, or borrowing/diffusion. A genetic argument is thus a negative argument, or an argument by elimination, what in classical logic is termed a *disjunctive syllogism*. One rules out all but one of the logically possible accounts of relations of similarity, so that only inheritance from a putative common ancestor remains.

2.1 *The first premise of the comparative method*

It is not unusual for scholarly papers on historical linguistic topics, and linguistics courses on the comparative method and its application, to deal with the possibilities of chance resemblances between languages, and of resemblances through borrowing/diffusion. The possibility of natural resemblance is addressed much less often. By *natural resemblance* I intend those instances of similarity between linguistic objects that are simply not surprising, and do not, by their nature, call for any account. In order to be any more precise, we must permit ourselves to be informed by insights from what can be termed “classical semiotics,” in particular, to the semiotics of the late nineteenth-century American philosopher C. S. Peirce.⁸

Peirce’s semiotics involved a number of three-way distinctions – Peircean trichotomies. The best-known is one based on a sign form’s “fitness to signify”:

- i *indexical signs*, whose forms are fit to signify by virtue of being *part of* their object;
- ii *iconic signs*, fit to signify by virtue of some similarity between the sign form and its object; and
- iii *symbolic signs*, fit to signify by virtue of some convention or agreement that their forms will stand for particular objects.

As Saussure pointed out, only in the case of symbolic signs is the sign relation arbitrary. Since indexical and iconic signs are natural (non-arbitrary), we have no reason to be surprised by their cross-linguistic similarity. It is only in the case of *arbitrary* relations between the form and the meaning of linguistic signs that comparativists ought to find cross-linguistic similarity surprising. Comparative historical linguists *only* have cause to be surprised by, and must seek explanation for, similarities between form–meaning pairings in different languages *when those pairings are symbolic*.

So the comparativist is on the safest ground by restricting comparison to those linguistic signs that are the most arbitrary and conventional – individual lexical items. One has no strong warrant to infer genetic relatedness from similarities in iconic signs – onomatopoeic forms, metaphors, compounds, or syntactic patterns – since such similarities can be explained in terms of the limited possibilities afforded by observation and analysis of the world.⁹ I will refer to the restriction of comparison to symbolic signs as the *semiotic restriction* on, or the *first premise* of, the comparative method.

It is, therefore, the first premise of the comparative method that focuses attention on the lexica of the languages compared, and not the fact that nineteenth-century linguists couldn’t do syntax, or anything of the sort. At the risk of unnecessary repetition, we have no clear warrant to compare anything other than symbolic linguistic signs, because sign similarity is only surprising when the signs are symbols. This fact does not mean that we must restrict comparison to monomorphemic signs, but it does mean that we are on increasingly thinner comparative ice the more abstract/less symbolic the signs we compare.

2.2 The relation cognate with

It is tempting to think of the relation *cognate with* as differing only in domain from the relation *genetically related*. The latter, defined over languages, would be in some sense the *sum* of instances of the relation *cognate with*, defined over individual linguistic expressions, grammatical rules, or whatever. But that interpretation confuses reality, what actually is the case, with demonstrability, what we can show to be the case on the basis of available evidence and “the state of the art.” Two languages¹⁰ can, in principle, be genetically related without a single cognacy relation being evident in the synchronic states of those languages. That is, those languages might be genetically related, without our being able to adduce any evidence of that relatedness. And that is precisely what instances of the *cognate with* relation are – a demonstration of genetic relatedness. If one can *prove* that even one single cognate pair holds over two languages, one has proven those languages genetically related.¹¹

Two linguistic objects σ_1 and σ_2 are cognate:

$\text{cognate}(\sigma_1, \sigma_2) [\equiv \text{cognate}(\sigma_2, \sigma_1)]$

iff both are reflexes of a single antecedent linguistic object $^*\sigma$:

$\text{reflex}(\sigma_1, ^*\sigma_1) \wedge \text{reflex}(\sigma_2, ^*\sigma_2) \wedge ^*\sigma_1 = ^*\sigma_2$

A linguistic object σ_t is a *reflex of*¹² a linguistic object $\sigma_{t'}$ if:

- i σ_t and $\sigma_{t'}$ are in temporally distinct language states t and t' (t subsequent to t') and if:
- ii σ_t is a “normal historical continuation” of $\sigma_{t'}$

Being more precise about what is meant by “normal historical continuation” isn’t easy. It must involve notions like “normal language acquisition” and “normal language change.”¹³ Although there may be some danger of circularity here, it seems to me safe to assume that historical linguists will know what I have in mind.

As noted above, comparative historical linguists must identify instances of the *cognate with* relation in order to demonstrate genetic relatedness. Even the techniques of “mass comparison” (as evidenced, for example, in Greenberg 1987; Ruhlen 1994), or any other method that begins with the mere observation of similarity, must ultimately trade in cognates. There is no other logical possibility, in the absence of written records or time machines. The comparative method is simply the principal (indeed, the only) means available to historical linguists for identifying cognates *convincingly*.

2.3 Phonological comparison and the regularity assumption

Let me stress this point again. The relation *cognate with* is independent of the comparative method. Though the comparative method is a technique for

identifying cognates, cognacy can exist without the comparative method being able to demonstrate it. That is, the comparative method has limits.

The most immediate limit on the method is the one faced by the working comparative historical linguist even before she or he sets off to hunt for cognates. The problem is *where* in language to look for cognates. One could look anywhere (a point taken up below with regard to grammatical comparison in section 3.1). But the comparative method, I would argue, is not designed to demonstrate cognacy in general, but cognacy only in the *lexicophonological domain*.

For the remainder of this section, I will assume that candidates for cognacy testable by the comparative method are (possibly morphologically complex) linguistic signs whose phonological shape is in a form no more abstract than (taxonomic) phonemic. That is, I assume we are comparing morphemes or morpheme sequences, in phonemic notation, up to the level of the phonological word.

As observed at the end of the preceding section, the comparative method is a procedure for identifying n -tuples that are instances of the *cognate with* relation, at some reasonable level of confidence. I will assume that any pair of items f and g , from different languages and meeting the domain conditions, are potential cognates. And I will use the possibility operator M of modal logic to represent potentiality. The problem of proving cognacy for potentially cognate pairs can be reduced to or recast as the problem of defining a rule of *M-elimination* that licenses the move:

$$\begin{array}{l} M \text{ cognate}(f, g) \\ \text{cognate}(f, g) \end{array}$$

The comparative method is an attempt at defining this rule of *M-elimination*. The following is an informal approximation:

M-elimination

A pair (f, g) of potential cognates is a cognate pair if:

- i they meet a *similarity condition*: that f and g are *similar* in both facets of the sign relation, in form and in interpretation, and
- ii they meet a *disjunctive elimination condition* that the similarity is not (likely to be) a consequence of *chance* or of *borrowing/diffusion*.

2.3.1 *The similarity condition*

Condition (i), the similarity condition on potential cognates, is logically prior to condition (ii), on non-genetic accounts of the similarity. After all, you have to recognize similarity before you seek to explain it! But that fact does not make the similarity condition a *precondition* (that is, a condition on *potential cognacy*), as often seems to be assumed. I choose to view condition (i) as part of the *proof* of cognacy (as part of *M-elimination*) because I believe that the definition of similarity is in fact *part of* the comparative method, at the very least, as the method was first devised.

Under this interpretation, it is the similarity condition of the comparative method that rules out natural (i.e., iconic) similarities and enforces the *semiotic restriction* on the comparative method. With the comparative method, we restrict comparison to symbols because it is only similarity between arbitrary and conventional (*symbolic*) signs that is surprising, and that could be evidence of cognacy.

Similar symbols must be similar in both form and interpretation. While it may not be entirely fair to say that comparativists have done nothing to clarify the notion “similar meanings,” we haven’t done much. Most recent work has focused on *grammaticalization*,¹⁴ the process by which reference to particular sets or relations in the world changes into higher-order reference: motion verbs to source/goal markers, object–part relations (like “top surface” or “cavity beneath”) to object–location relations (like “on” or “under”), and so forth. But we are still very much at the data-collection stage in this endeavor, and are informed in it only by vague senses of what are possible metaphors or metonymies. Sadly, we don’t really pay much attention to the meaning side of things. In general, unless a particular meaning comparison grossly offends some very general sense of metaphor, it’s “anything goes” with regard to meaning.

Comparative historical linguists have been rather more careful in stipulating what it means for linguistic symbols to be similar in form. Observe first that similarity of form must be *complete* similarity. Put rather brutally, if the front halves of two forms are similar, but the back halves aren’t, then the forms are not similar. In practice, we observe this condition by segmenting each form into its component (segmental or autosegmental) parts, and then mapping the segmented forms into a set of *correspondences* between a part or parts of one form and a part or parts (possibly nil) of the other. We need not go into the mechanics of that segmentation process here. The problem of the similarity of sign forms then reduces to the problem of similarity of objects in a *correspondence* relation. And that, as we shall soon see, is not a problem at all!

Feature (attribute-value) theories of phonological representation (and of articulatory description that precedes them) make it possible for us to measure the similarity between two representations of phonological form, in terms of shared attribute-value pairs. Phonological feature theories do not, of course, tell us precisely how many attribute-value pairs must be shared by two forms for them to be deemed sufficiently similar to be cognate. Nor is it clear how one would, in practice, begin to construct a method that makes such a determination.

2.3.2 Regularity, similarity, chance, and borrowing

The good news is that comparative historical linguists, using the comparative method, do not need any measure of relative similarity that decides when two forms are similar enough. In fact (and a fact that is not, I think, widely appreciated), comparative historical linguists don’t need, and have never really needed, a theory of phonetic similarity at all.¹⁵ What we have instead is the *regularity assumption*.

I use the term *assumption* here quite purposefully, because it is by now well demonstrated that sound change is not regular, in the usual intended sense, but precedes in a quasi-wavelike fashion along the social and geographic dimensions of the speech-community, and through the linguistic system itself. At any given point in time, a particular sound change may be felt only in a part of the speech-community and, if it affects lexical signs, only through a portion of the lexicon.¹⁶

Why then do we cling to this assumption, when it is so demonstrably false? For two reasons, it seems. First, given enough time, sound changes will *tend* toward regularity; they will continue through the community and through the linguistic system until close to all speakers and close to all appropriate sign tokens are affected. Second, and more significantly, the *assumption* of regularity stands in for a theory of (or a measure of) form similarity. The actual form of two phonological types in a *corresponds to* relation is irrelevant; all that matters is that the relation holds for *all* tokens of those two types (under any appropriate local conditions).

One function of the regularity hypothesis is to filter out chance resemblances, which are quite unlikely to be regular and, to a lesser extent, to filter out borrowings, so long as the borrowing has not been on a massive scale and, if from related languages, has not been subject to nativization rules that lend to borrowings the appearance of regularity. To be sure, the regularity hypothesis does help enforce the *disjunctive elimination condition*. But it is much more than that. To early comparativists, it was a methodological *sine qua non* of the comparative method, enabling the work of comparative historical linguistics to proceed in the absence of any theory of phonetic similarity. Indeed, many of the data on which present theories of phonetic similarity were constructed are derived from the regular correspondences of the comparative method. And even now, with our feature theories informed by 150 years of work on both synchronic and comparative historical phonology, we cannot dispense with the regularity hypothesis, because it saves us from having to determine just how similar similarity must be, in order to demonstrate cognacy.

3 The Limits of the Comparative Method

Having outlined the essential features of the comparative method, as I understand it, let me at last turn to the issue of its limits and limitations. I divide these into two rough groups:

- i limitations deriving from the interaction of language data and the method;
- ii limits imposed by the method itself.

The first group consists of those situations in which the facts of language change, in particular circumstances, can conspire against the comparative method. These are essentially situations in which the method hasn't appropriate language data on which to operate. The problems that fall within this group include:

- i the temporal limit problem;
- ii the massive diffusion problem;
- iii the subgrouping problem.

The second group consists of those linguistic domains for which the comparative method is simply not designed to operate. To discuss these limits one must address the domain problem on cognacy, in particular the issue of grammatical comparison and reconstruction.

3.1 Comparing grammatical objects

Section 2.3 above introduced what might be termed the *domain problem* for the cognacy relation. Those who use the comparative method must recognize that words or morphemes are in the domain of the cognacy relation. Cognacy between phonological units like phonemes can also be admitted (if *cognate with* is defined in terms of *reflex of*, as suggested in section 2.2 above).¹⁷ But what other linguistic objects are in the domain of the *cognate with* relation – syntactic categories, syntactic rules, paradigms? Is a syntactic rule or morphological paradigm of Portuguese, for example, to be considered a reflex of some rule or paradigm of (Vulgar) Latin, and thus potentially cognate with some similar object in French or Romanian? The quick answer to these questions is, I think, yes.¹⁸ But a qualified yes, the qualifications being that:

- i the cognacy of such objects cannot be determined by the comparative method, and that
- ii genetic relatedness cannot be determined on the basis of the putative cognacy of such objects.

Grammatical objects are different in their degree of abstraction from the lexicophonological objects on which the comparative method operates, and that difference is crucial to how we interpret those objects historically. But a slight synchronic excursus is in order, to flesh out what is intended here by the differing abstractness of lexicophonological and grammatical objects.

3.1.1 The nature of grammatical objects

An interesting insight in Head-driven Phrase Structure Grammar, at least in its early incarnations (for example, Pollard and Sag 1987), was the manner in which it generalized the notion “linguistic sign.” The term “linguistic sign” is often treated as if it were synonymous with “morpheme,” in the American structuralist sense of that term. In HPSG, it is explicitly generalized along two dimensions:

- i internal complexity;
- ii abstraction.

Any linguistic form with an interpretation and/or function is a linguistic sign,¹⁹ from the non-compositional morpheme at least up to the level of the sentence. The major difference between morphemes and sentences is that the former,

but not the latter, are paired with their interpretations in a lexical listing, while the latter are semantically compositional (in theory at least). The type of information each contains differs, of course, but that fact doesn't detract from their fundamental similarity.²⁰

Consider the following pair of pseudo-HPSG attribute-value matrices:

- a.
$$\left[\begin{array}{l} \text{cat: cn} \\ \text{syn: } \left[\begin{array}{l} \text{agr: } \left[\begin{array}{l} \text{num: sg} \\ \text{gen: masc} \end{array} \right] \end{array} \right] \\ \text{phon: /gato/} \\ \text{sem: } \lambda x.\text{cat}(x) \end{array} \right]$$
- b.
$$\left[\begin{array}{l} \text{cat: cn} \\ \text{syn: } \left[\begin{array}{l} \text{agr: } \left[\begin{array}{l} \text{num: sg} \\ \text{gen: masc} \end{array} \right] \end{array} \right] \\ \text{sem: } \lambda x.f(x) \end{array} \right]$$

Matrix (a) might be a partial representation for the Portuguese *gato* "cat," while matrix (b) is derived from (a) by abstracting away certain information (in this case, the item-specific phonological and semantic information). Matrix (b) is a representation of an abstraction, of a set of linguistic signs; in this case, set-denoting masculine singular common nouns. If (a) had been a complex sign like a noun phrase or sentence, then the corresponding abstraction (b) could be interpreted as a template or phrase structure rule for complex objects like noun phrases or sentences.

Grammatical objects, then, are abstractions on actual linguistic signs; on words, phrases, clauses. These abstract objects can still be considered signs, form–meaning pairings, to the extent that:

- i we are willing to regard *as form* the structural information remaining after actual phonological shape has been abstracted away, and
- ii it is possible to associate some meaning with such grammatical abstractions.²¹

The meanings associated with grammatical objects are of course themselves likely to be quite abstract. For example, the meaning associated with the category "cn" (common noun) in analysis (b) above is just "predicate on (or set of) individuals." But the meanings of grammatical or functional items like tense or plural markers are no less abstract than these, so their status as meanings should not be in doubt. In the following sections, I consider whether these grammatical objects can be compared, reconstructed, and used as evidence in genetic arguments.

3.1.2 *The comparison of grammatical objects*

Genetic linguistic inferences follow from the fact that, in certain circumstances, we can be justifiably surprised at similarities between different languages. The comparative method, as understood here, provides two essential tools that

make *genetic* inferences possible. In its data domain, it provides the reason to be surprised, in that similarities in symbolic form-meaning pairings cannot be attributed to nature, and are unlikely to be the result of chance. In its method, and in particular in the regularity assumption, the comparative method provides a "measure" of similarity.

Grammatical objects fare poorly as evidence for genetic relatedness under the comparative method on both these counts. On the one hand, we have little reason to be surprised at the particular form-meaning pairings observed in grammatical objects. On the other, there can be no regularity assumption for grammatical objects to provide a measure of similarity.

Observe first that there can be no regularity assumption for grammatical objects because these objects are unique. The reason is axiomatic, and thus beyond question. It is a theoretical premise in linguistics that individual simplex linguistic signs reside in a lexicon, a repository of linguistic unpredictability. We can thus speak of individual lexical items undergoing or not undergoing some sound change, because those items exist individually. Modern linguistics accepts as axiomatic that complex linguistic signs, by contrast, do not reside in some vast "grammaticon," from which they are drawn as needed in language production or reception. Rather, they exist as latent or potential linguistic signs, in the grammatical objects onto which they are abstracted. It is thus incoherent to speak of a grammatical change being regular, since a grammatical change applies in only one abstract object.

We can nonetheless compare grammatical objects in different languages, and describe the degree to which they are similar. But just how similar must two grammatical objects be for that similarity to be surprising, and thus count as evidence of genetic relatedness?²² The question is not even an interesting one, though, because similarities between grammatical objects are seldom, if ever, surprising.

Grammatical objects are templates, diagrams, or rules encapsulating what is common in sets of (simplex or complex) linguistic expressions. For the most part, grammatical objects are iconic, and not symbolic signs. This is true both for syntagmatic signs abstracted from complex linguistic signs and encapsulating combinatory linear or hierarchical information, and for paradigmatic signs abstracted from sets of lexical items and encapsulating selectional information.

Syntagmatic signs are iconic to the extent that they are compositional. If the syntagmatic information in a grammatical object, whose meaning is a function of the meanings of its component parts, is information that those parts are adjacent or overtly coindexed in some way (by agreement morphology, for example), then this information is not surprising. The "closeness" in form is iconic of association in meaning. Indeed, we would be surprised if this were not the case. And if the syntagmatic information is simply hierarchical, syntactic dominance information, there seems to me to be no question of whether or not to be surprised by association of this formal property with some semantic operation; the hierarchical association *is* the semantic operation.

In the literature on syntagmatic object comparison, semiotic considerations have run a distant second place to arithmetic-combinatoric considerations

in the more restricted domain of word order comparison.²³ Thus, a common account of the failure of the comparative method in *syntax* (read, *word order*) is the *poverty of choice* argument. In the case of comparisons of Greenbergian major clause constituent typologies,²⁴ that argument runs as follows: since there are only 2^3 (= 8) possible permutations of the major clause constituents S(ubject), V(erb), O(object), there is a 1:8 chance of any two languages sharing a (predominant) major clause constituent typology by accident, and that probability is too high to discount accident.

As compelling as the poverty of choice argument may be, in itself it is of less significance to the issue of grammatical object comparison than is the approach to grammatical theory it presupposes. What gives rise to the poverty of choice (in this case, eight possibilities for major clause constituent order) is an analysis of (transitive) clauses that assumes a limited number of major clause components (in this case, three), and a theory of grammar that permits the identification of those components cross-linguistically. It is the theories of grammar to which most linguists subscribe, and their assumptions of universality, that give rise to the poverty of choice, and deprecate grammatical object similarity as evidence of genetic relatedness. We can never be surprised by the fact that two languages share some property that is universal.

Grammatical objects need not be universal in the strong sense of the preceding paragraph for their value as genetic evidence to be questioned, as was observed above for the case of compositional syntagmatic objects. But this fact is not just true for compositional objects. Any system of grammatical contrasts is iconic to the extent that it reflects a distinctly human ontology. This is true of the systems of categorial contrast associated with X' theories of phrase structure, and is true, in exactly the same way, for inflectional paradigms.

Inflectional paradigms can be viewed as metaphors, as iconic of a highly constrained analysis of the world, given expression in the structure of language. Systems of person-number marking, for example, map onto a characteristically human manner of indexing individuals in linguistic communication – for single individuals, as speaker, hearer, or neither and, for more than one individual, as including the speaker, the hearer, or neither.²⁵

Cases like those of morphological person-number paradigms are of particular interest because, although not universal in any absolute sense (but see further below), linguists are surprised neither by their occurrence nor by their non-occurrence in the verb or common noun morphology of particular languages. For example, Mokilese and Ponapean are two very closely related Micronesian languages, verging on mutual intelligibility. Ponapean, like most Micronesian languages, has a transitive verb paradigm, with distinct suffixed forms indexing the person-number of the direct object. Mokilese transitive verbs are invariant, the person-number of the object being marked by independent pronouns when necessary. The Ponapean suffixal transitive paradigm is similar in structure to that found in Biblical Hebrew transitive verbs (and those of other modern Semitic languages). To be sure, there are differences in the structure of the Hebrew and the Ponapean paradigms; Ponapeic languages do not make gender distinctions, and Hebrew does not have the dual-plural

direct object contrast found in Ponapean,²⁶ or the inclusive–exclusive contrast. But exactly the same is true of at least one other Micronesian language, Gilbertese, with an object-indexed transitive verb paradigm identical to the Hebrew, except in not making the gender distinctions found in the Semitic paradigm. And, finally, one might observe that Mokilese and English are similar in not having object-indexed transitive verb paradigms at all.

Comparative linguists might wonder how the situation arose in which two languages as closely related as Mokilese and Ponapean are could differ in this significant respect. But no comparative historical linguist would cite the paradigm similarities between Gilbertese and Hebrew, or English and Mokilese, as evidence in a genetic argument. We are surprised neither by the occurrence nor by the non-occurrence of morphological object paradigms because we believe them to be, in some sense, latent in the human language faculty. And, indeed, this particular latency has been elevated to theory-licensed universality in recent proposals for AGR_O (an object-agreement constituent) in Principles and Parameters and in Minimalist syntax.

In summary, if one were able to identify grammatical objects that are not iconic in any of the senses considered above, but, rather, reflected some arbitrary means of mapping between the categories of language and of the world, then one could speak of comparative grammatical evidence of genetic relatedness. Radical Whorfians would have little trouble finding such cases. For most of us, though, the task would be much more difficult.

3.2 *The reconstruction of grammar*

The comparative method *sensu stricto* is a method for determining genetic relatedness amongst languages. While some aspects of the proto-language are reconstructible as a by-product of the comparative method, that is not the method's primary function. One can use the comparative method to draw genetic conclusions without reconstructing a thing!

For the reasons outlined here, I do not believe that the comparative method can be applied to grammatical objects (as described in the preceding section) to determine genetic relatedness and to reconstruct antecedent grammatical objects. But let me now temper that view by saying that I believe it is possible to compare and reconstruct grammatical objects, using other methods, *after genetic relatedness has been established*.

Once we know that two languages are genetically related, we know that at least some of the grammatical objects in those languages are reflexes of objects in their common parent, and that some of those are likely to be cognate. And once parallel separate developments and borrowings are weeded out, all that remains is to tell a plausible story about how grammatical objects in different languages developed from a single antecedent grammatical object. But such historical inferences about grammatical objects are not being guided by the comparative method, but by some other principles, because we can draw no genetic conclusions from them.

3.2.1 Undoing grammaticalization

So, not all linguistic comparison necessarily instantiates the comparative method. Nor, of course, is all linguistic reconstruction comparative. There is the “method of internal reconstruction,”²⁷ by which morphophonemic alternations are undone in putative antecedent linguistic states, and the as-yet-unnamed (and less often taught) techniques for “undoing” *grammaticalization*, by which earlier grammatical forms and constructions are inferred from synchronic observations regarding lexicon, morphology, and syntax. DeLancey (1994b) quite correctly observes that these techniques are a form of *internal* rather than *comparative* reconstruction.

A consideration of these techniques of internal grammatical reconstruction, by which instances of grammaticalization are undone, is not properly within the scope of this chapter. But these techniques are entrancing, and have yielded, for me, a number of papers, published and unpublished, on the grammatical history of Oceanic (and, particularly, Micronesian) languages. I thus cannot leave them without comment.

3.2.1.1 Typological consistency of word order

Let me first off distinguish between two quite distinct premises for undoing grammaticalization. The first is that the relative order of clitics and their hosts, and affixes and their stems, reflects the earlier order of complements and their heads or (attributive) operators and their operands. This premise allowed Givón (1971), for example, to infer historical OV constituent order from English compounds like *baby-sit* or *donkey-ride*. The technique seems to get considerable support from cases, like Romance, where the history is known. Given that Classical Latin was OV,²⁸ while its Romance descendants (and their hypothetical post-Classical ancestor, Vulgar Latin) are VO, the fact that Romance pronominal clitics are pre-verbal seems to hark back to the putative Latin situation; that is, until one observes that metropolitan Portuguese, which is apparently morpho-syntactically conservative in a number of respects, has enclitic verbal pronouns.²⁹

This use of internal reconstruction, to recover older word order, suffers from a similar problem to that of its better-established morphophonological cousin; both involve a “historical uniformity” assumption. In standard “internal reconstruction,” one assumes that phonological alternation develops from prior non-alternation; in word order reconstruction, one appears to have to assume that constituent order was typologically consistent at some point in time. The prior uniformity assumption underlying morphophonemic internal reconstruction is not particularly problematic, but the parallel syntactic premise is questionable, because it is, in fact, a much wider claim. All that is being assumed in morphophonology is that the particular alternation in question reflects the operation of conditioned sound changes on historically non-alternating forms.

We are not warranted in assuming any more in the syntactic cases; that is, we can assume that the constructions antecedent to the English N-V compounds were [N V], and that the constructions antecedent to the Romance pro-V clitic structures were [pro V] (*pace* Portuguese). What we are *not* safe in assuming is

that all (or any other) [V, NP] complement structures in either pre-Romance or pre-English were verb-final, any more than we are safe in assuming that any synchronic grammar will be typologically consistent. In short, we can infer something from synchronic word order, but not much.

3.2.1.2 *Semantic bleaching*

A second technique for undoing grammaticalization is employed on cases of “semantic bleaching.” These are instances in which morphemes have much of their particular semantic content abstracted away. For example, relational common nouns (like ‘bottom’ or ‘surface’) develop into thematic-role markers. Motion verbs and modals come to have temporal marking functions, demonstratives become articles or complementizers, and so forth. This phenomenon has been recognized in the literature for some time (see, e.g., Benveniste 1968; Givón 1975).

One argument form commonly employed to recover instances of semantic bleaching begins with observations of polysemy/homonymy in a language. A particularly transparent case is that of Gilbertese *nako*, which has three functions:

- i a motion verb ‘go’
 Nako mai.
 go hither
 “Come here.”
- ii a directional enclitic ‘away’
 E matuu nako.
 3s sleep away
 “She or he fell asleep.”
- iii a preposition ‘to(ward)’
 A boorau nako Abaiaang.
 3p voyage away Abaiaang.
 “They travelled to Abaiaang.”

Using the premise (the second mentioned above) that polysemy/homonymy is likely to be the result of semantic change, one postulates a single form and function for sets like *nako*, and constructs a plausible history to account for the observed polysemy/homonymy. The technique is clearly a form of internal reconstruction, in which the alternation being eliminated is semantic rather than phonological.

The case of Gilbertese *nako* is not only a transparent one, but also one for which there is no obvious *synchronic* analysis of the observed polysemy/polyfunctionality.³⁰ As is doubtless true of most historical grammarians, I have been tempted over the years to resolve other, less trivial cases. For example, in Harrison (1982), I used both internal arguments and comparative evidence in a historical resolution of the Gilbertese agentless passive suffix *-aki* and a particular transitivizing suffix *-akina* restricted to motion/stance and some psychological state verbs. The subsequent publication of Burzio’s (1986) observations regarding the unaccusativity of a similar semantic class render that resolution much less fanciful than it may have appeared at the time.

3.2.1.3 *Grammar and the comparative method*

Yes, comparative evidence is used in reconstructing grammatical items, but this is not the comparison and reconstruction of grammatical objects as defined in section 3.1. Much of what is called grammatical reconstruction in the literature is just the plain vanilla comparative method applied to morphemes in the usual way.³¹ The main difference is that the morphemes have glosses like ‘to,’ ‘present,’ and ‘ergative marker,’ rather than ‘sun,’ ‘wind,’ and ‘fire.’

When abstract “grammatical” items are compared, it is often the case that the formal phonological relationships between the items compared are less an issue than are the functional semantic relationships. A comparativist who pays little attention to the glosses of putative cognates, as long as they are in the right semantic neighborhood, will often become much more demanding regarding grammatical items. A case in point: Proto-Micronesian **fanga-ni* ‘to give’ is easily reconstructed on the basis of cognates in Gilbertese and Trukic. My suggestion (Harrison 1977) of a Ponapeic cognate in Ponapean *-eng* and Mokilese *-oang* has not been universally accepted by other Micronesianists. The historical phonology is perfect. The problem is that the Ponapeic form is a verb enclitic marking dative/goal arguments.

This may be healthy skepticism in general, because the only limit on the language-internal or comparative cognacy of grammatical items is our sense of metaphor and of possible semantic relation. And some historical linguists can be very imaginative indeed. But one shouldn’t be too skeptical of this endeavor, because what those engaged in the comparison and reconstruction of grammatical items are doing (albeit in rather circumscribed domains) is something the field as a whole should have been attending to all along – the comparison of meanings.

3.2.1.4 *The role of morphology and the significance of oddity*

Meillet is credited with the assertion that “morphological” evidence is stronger evidence of genetic relatedness than is mere phonological correspondence. The claim seems to derive from a discussion in Meillet (1948), where he states (pp. 24–6, given here in translation):

From the principle underlying the [comparative] method, it follows that, within the domain of comparative grammar, the probative facts are idiosyncrasies, and they are so much the more convincing as, by their very nature, they are less suspect of being attributable to a general cause. This is only natural: given that what is at issue here involves positing, via comparative procedures, the historical fact of the existence of a particular language – that is to say, of a thing which, by definition, arises due to a series of diverse circumstances which have no necessary connection with one another – it is these characteristic idiosyncrasies alone which must be taken into consideration.

Meillet then continues with an example from the paradigm of ‘to be’ in a number of Indo-European languages. Teeter extrapolates from that discussion the claim that “knowing that German has a verb ‘to be’ with a third singular

ist and third plural *sind*, and that Latin has one with a third singular *est* and a third plural *sunt*, is all by itself sufficient to guarantee the relatedness of German and Latin" (Teeter 1994b). This Meillet–Teeter conjecture is not a claim that the structure of the morphological paradigm (i.e., a grammatical object, in the sense of section 3.1) is evidence of genetic relatedness, but that the presence of particular fillers in particular slots of the paradigm is evidence of genetic relatedness.

Let me make two points about this issue. The first is merely to reiterate my views about the status of grammatical object similarity as evidence for genetic relatedness. The fact that Polish and Lithuanian both have a common noun paradigm that distinguishes two numbers (singular and plural) and seven cases (nominative, genitive, dative, accusative, vocative, locative, and instrumental) is not evidence that the languages are genetically related. It only becomes evidence when the phonological shapes of the characteristic markers (of some significant number) of those paradigm slots are also similar,³² as the comparative method would require.

The second is to question the claim that *ist/est* and *sind/sunt* have privileged status as evidence of genetic relatedness. Teeter claims their special status derives from the fact that they are "grammatical lookalikes, guaranteed to prove genetic relationship because grammar (short of learning a language) is exempt from borrowing" (Teeter 1994c). It is not clear what a "grammatical lookalike" is, but it is clear that two putative cognates are not exempt from the usual strictures of the comparative method just because they happen to be members of a high-frequency morphological paradigm. And, as Thomason and Kaufman (1988) point out, *nothing* is exempt from borrowing.

Teeter's motivation seems clear to me, because it is at the heart of the comparative method. Like many of us, he wants some sort of evidence that is guaranteed to satisfy the disjunctive condition of section 2 – something odd, outstanding, or irregular. The principal virtue of the comparative method is just that its logic doesn't demand that we seek out oddities, but regularities.

Manaster Ramer (1994) points to examples of what he regards as odd syntax, and suggests that their oddity alone makes them reconstructible. His principal example is the singular verb agreement of neuter plural nouns in Old Iranian and Ancient Greek.³³ Since he seems to be suggesting that such syntactic oddities are unlikely to have arisen by chance or been borrowed, then it would appear to follow that he regards them as evidence of genetic relatedness. But the whole argument rests on the premise that a certain sort of grammatical object is odd. A principled definition of "grammatical oddity" is desirable, before one can accept such evidence.³⁴

3.3 False negative results from the comparative method

The comparative method was not designed to operate on non-lexical data. There are at least two situations in which the comparative method fails on lexical data, in not recognizing genetic relatedness amongst languages that are genetically related. These are:

- i very long absolute time depth for the proto-language;
- ii massive diffusion of lexical items across a multilingual domain.

3.3.1 *Time depth*

Time is both parent and adversary to the comparative method: without change through time, there is nothing to compare; with enough change over enough time, comparison yields nothing. That is the most basic lesson in comparative linguistics. The more time that elapses from the initial break-up of some ancestral language, the more difficult it will become to demonstrate the kinship of its descendants.

The effect of time has nothing whatsoever to do with any putative upper limit on the comparative method. It has to do with the availability of evidence. The more time, the more change, the more lexical replacement, the fewer cognates: end of story. The limit is a practical (and statistical) one, not a temporal one. When the number of putative cognates and/or correspondence sets approaches a level that is not statistically significant (i.e., that might be attributable to chance), the comparative method has ceased to work.

Johanna Nichols (1992a), among others, muddies the waters somewhat by stating the restriction in terms of absolute dating (8000–10,000 years). In a thread of discussion on the time-boundedness of the comparative method, she qualifies quotes like: “But the comparative method does not apply at time depths much greater than about 8000 years (this is the conventional age of Afroasiatic, which seems to represent the upper limit of detectability by traditional historical method)” (Nichols 1992a: 2–3) by saying that one arrives at such absolute limits not by analysing the comparative method, but by examining the “oldest uncontroversial genetic groupings” (Nichols 1994b) and, one assumes, using the oldest date amongst those (which she suggests is that for Afro-Asiatic).

As others rightly asked in the subsequent discussion: where do those dates come from? Only two places, so far as I am aware. One possibility is from the archeological record, if there is some reason to associate a particular datable assemblage with a particular node on a genetic linguistic tree. For example, many Austronesianist prehistorians have sought to associate the Oceanic node on the Austronesian family tree with the Lapita pottery culture.³⁵ The other source of dates is glottochronology, in one guise or another. For glottochronology, one must make some assumption about the rate of lexical replacement/retention. The constant *r* usually cited is 14 percent replacement (86 percent retention) per millennium. As has often been pointed out, Bergsland and Vogt’s (1962) paper should have put paid to the notion that there is such a constant, but it seems that each new generation of comparative linguists must learn this lesson anew.³⁶ I side with Jacques Guy (1994) on this one, when he says: “Short of datable documentary evidence – such as lapidary inscriptions, clay tablets, etc. – there is no way to date putative ancestors, no way at all.”

What interests me most of all is why so many historical linguists feel drawn towards absolute dating. Sure, it would be nice to know when, but

the comparative historical enterprise doesn't stop because that question can't be answered. It seems to me that the obsession with dates, like the obsession with family trees, is at least partly the result of "prehistorian envy." Too many comparative historical linguists want to dig up Troy, linguistically speaking. They consider it more important that comparative historical linguistics shed light on prehistoric migrations than that it shed light on the nature of language change. I can only say that I do not share those views on the focus of comparative linguistics. I do not consider comparative historical linguistics a branch of prehistory, and I sincerely believe that if we cared less about dates, maps, and trees, and more about language change, there'd be more real progress in the field.

3.3.2 Diffusion

In a number of papers, Grace (1981, 1985, 1990) reports the results of research conducted on the languages of southeastern New Caledonia over a 20-year period beginning in the mid-1950s. Grace's intention was to place these languages more accurately within the developing tableau of genetic relationships amongst the Oceanic languages. The problem had been that these languages were what Grace terms "aberrant," in that their phonologies did not correspond to the general Oceanic pattern. This historical accident, Grace reasoned, was what was obscuring their Oceanic genetic heritage. Grace also reasoned that if one reconstructed from those languages *alone*, the resulting reconstruction would undo much of what was aberrant about the southeastern New Caledonian languages, and facilitate comparison with other Oceanic languages.

Grace was able to collect extensive material on two SE New Caledonian languages, Canala and Grand Couli. An initial inspection of these data suggested some nine hundred possible cognate sets between these two languages. But, far from reducing the degree of "aberrancy" (relative to other Oceanic languages) of the New Caledonian languages, the results Grace obtained by applying the comparative method to these languages only made matters worse.³⁷

Both Canala and Grand Couli have identical inventories of 24 consonants and 18 vowels (oral and nasal). Grace identified 140 consonant correspondences (56 with more than 5 tokens) and 172 vowel correspondences (67 with more than 5 tokens). Nor was there much evidence of conditioned change to reduce the number of reconstructed segments. These results do not demonstrate genetic relatedness, even though it is obvious that the languages in question *are* genetically related. On one interpretation, the correspondences are simply not regular; on another, the reconstructed inventory is not that of a natural language.

Grace (1990) suggests two possible explanations for the situation observed in SE New Caledonia. The first challenges the regularity assumption. Under that account, a change begins, affects a few tokens, and stops. Another change begins, affects a few tokens, and so forth. As stressed earlier, attacking regularity is beating a dead horse. The falsity of the regularity assumption, as an account of how language change takes place, is evident. The assumption is a methodological,

not an empirical, necessity. In those cases in which it is grossly violated, as here perhaps, nothing can be done, because the method won't work.

But it is not clear that that is the better of Grace's two explanations. His second account relies on the sociolinguistic situation in southern New Caledonia. In that area, marriage is outside the local community, often (if not typically) into a community with a different language – whatever that might mean; for Grace also asserts that our European monolingual view of the world may not apply to this situation, because languages have “mixed” to the point that the notion of “pure” distinct languages might not make any sense.

If time is one great adversary of the comparative method, prolonged socio-economic intercourse amongst small-scale (genetically related) linguistic communities is another. Language contact and borrowing are a normal occurrence, and make comparative linguistics interesting. But most instances of borrowing can be recognized as such, and factored out. Even cases of massive borrowing (as a consequence of some cataclysmic event like invasion) can often be teased out. There is, for instance, the classic Oceanic case of Rotuman, as reported in Biggs (1965), where two distinct sets of correspondences ultimately revealed themselves, one native and one imposed from outside.

Grace's New Caledonian case is not like that. It appears to have been the result of a slow but relentless dissolving of lexical resources into a common pool. The effect on comparative historical method is profound too. We “know” the languages are related, but can't demonstrate that they are by using the logic of the comparative method. Nor is this case an isolated one. Though I am not an Australianist, from what I have come to know second-hand about the situation in parts of northern Australia (Arnhem Land, for example), a situation parallel to the New Caledonian one holds there. The languages are grammatically quite similar, often admitting of morpheme-by-morpheme translation. The lexica look comparable. But the method doesn't work.

3.4 *The special case of subgrouping*

3.4.1 *Simple genetic arguments and subgrouping arguments*

The subgrouping problem is different from what I might term the simple (or *in vacuo*) genetic problem with which the preceding sections of this chapter have dealt. The simple genetic problem is to determine, for some set of languages $L (= \{L_1 \dots L_n\})$, whether or not the members of some subset of L share a period of common history. Using the comparative method, one does that by finding regular sound correspondences over sets of putative cognates. The subgrouping problem is a *tree selection problem*. One has already determined, using the comparative method, which members of L are genetically related (as descendents of some $*L$). The subgrouping task is to select, from amongst all possible trees T (with no non-branching nodes, to keep things finite!) with root $*L$ and leaves L , the one tree $T \in T$ that best represents the genetic history

(order of speciation) of L. Put somewhat differently, a simple genetic argument demonstrates that there is (or is not) a tree whose leaves are some subset of the languages compared; *if* there is a tree, subgrouping arguments are used to decide which tree. In a real sense, then, subgrouping is logically subsequent to determining genetic relatedness via the comparative method.

Subgrouping is not just comparative reconstruction of a small number of languages from a larger sample. The raw data for both simple genetic and subgrouping arguments are the same – sets/patterns of (partial) similarity in the form of linguistic expressions – but the propositions that one seeks to prove about those raw data are not precisely the same. In a simple genetic argument, one seeks to show that the patterns of similarity are a consequence of retention of properties of a common antecedent state, and not of diffusion or (natural or incidental) accident. In a subgrouping argument, one seeks to show that the patterns of similarity are not a consequence of retention from an antecedent state, but of a *unique* event (or change) common to the histories of all the languages in the subgroup.

To obviate any misunderstanding, let me make this last point a bit differently. In a simple genetic argument, we don't care whether the observed similarity is the result of some earlier change (in the history of the proto-language), or whether it reflects a situation going back to the dawn of time. In a subgrouping argument, it is crucial that the similarity be a shared innovation of the period of common history of the subgroup, an event/change that took place before the subgroup began to speciate, but after speciation at the immediately higher level in the tree.³⁸

It is also significant that subgrouping arguments must make crucial reference to changes (events). When we seek to rule out borrowing or iconic or accidental similarity in simple genetic arguments, using the comparative method, we are talking about the borrowing or chance similarity of linguistic signs. In subgrouping arguments, we are talking about the diffusion or chance independent repetition of linguistic changes. The canons of evidence in evaluating changes and signs are not necessarily the same.

3.4.2 *The practice of subgrouping*

Let's restrict attention here to two sorts of subgrouping evidence:

- i evidence from lexical identity;
- ii evidence from phonological similarity.

In order to demonstrate, in such cases, that the observation of similarity/identity is the outcome of a single act (of lexical coinage or sound change), one must demonstrate that the similarity/identity is unlikely to have been:

- i retention from an earlier state, and not change, or
- ii independent change in the languages sharing the form, or
- iii diffusion of the change across language boundaries.

In Harrison (1986), I identified six heuristics (in the form of implications) guiding the subgrouping enterprise. Two that are relevant to the evaluation of single correspondence sets (as subgrouping evidence) depend on the following premises:

- i The fact that any change takes place at all is remarkable. (The act or occurrence of a change is of itself a remarkable event.)
- ii Some changes are more remarkable than others. (Changes can be, and indeed are, ranked in terms of relative naturalness.)

from which one can conclude:

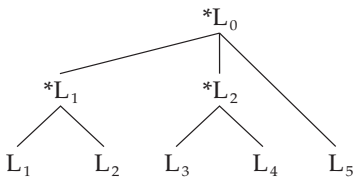
- i' Since the act or occurrence of a change is of itself remarkable, identical outcomes are likely to reflect a single act of change.
- ii'a A tree that entails a relatively unnatural change is a poorer candidate as a diagram of genetic relationship than one that does not entail that change.
 - b Unnatural changes are less likely to be repeated independently than are natural changes, and so are stronger evidence for subgrouping.

Heuristic (i') is essentially an appeal to simplicity; trees that represent a history with fewer change events are to be preferred over those that entail more change events. Note that (i') seems to vitiate (ii'b) somewhat, since (i') doesn't demand that we consider the content of the change at all.

Let's try to make all this a bit more concrete, by considering how to evaluate, as subgrouping evidence, a single hypothetical sound correspondence for a set L of five languages:

L_1	L_2	L_3	L_4	L_5
p	p	f	f	∅

If we assume, for the moment, that all the outcomes in this set represent change from *L then, by (i'), we would want to draw the tree:

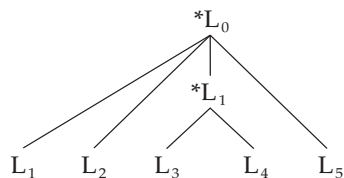


in order to minimize the number of actual events in the history. That history can be further simplified under the assumption that one of the outcomes reflects retention, rather than change. In the case in question, simplicity and simple arithmetic cannot be used to decide which outcome is the most likely retention, because at most one act of change is eliminated regardless of the choice made. But an appeal to (ii'a), through our linguists' understanding of the facts of change, does give an answer.

If we restrict attention to possible histories in which each language has undergone at most one change, the choices are:

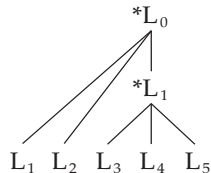
- a *p > f; *p > ∅
- b *f > p; *f > ∅
- c ∅ > p; ∅ > f

Choice (c) is likely to be ruled out immediately as just too unnatural an unconditioned change. Of the remaining choices, most phonologists and historical linguists would probably select (a), on the grounds that lenition is more common/natural than fortition. In that case, we have the tree:



in which L_1 and L_2 are assumed to have undergone no change.

We could stop there, but one might reason, by (ii'b), that the change $p > \emptyset$ in L_5 is unlikely to have proceeded in one step, and that a two-stage lenition process (with an intermediate fricative stage) is more likely/natural. Since L_3 and L_4 show that fricative stage, and rather than assume two occurrences of $p > f$, we can simplify the history by subgrouping L_3, L_4 , and L_5 , yielding:



3.4.3 Evaluating subgrouping arguments

That is how subgrouping is done, from the perspective of single correspondences at least. Observe, first, that heuristic (i') (called the *strong act of change* warrant in Harrison 1986) addresses the possibility of identical independent change only by denying it, and provides no guidance in ruling out either retention or diffusion. It is rather like what the comparative method would be, stripped of the restriction to symbolic data, and without the regularity assumption. By itself, (i') provides relatively unmotivated subgrouping hypotheses.

Given some theory of (sound) change by which changes are ordered for plausibility, heuristics (ii'a) and (ii'b) (together called the *fact of change* warrant in Harrison 1986) ought to filter out at least some cases of shared retention and of identical independent change. But these heuristics are far from unproblematic. *First*, the goals of eliminating retentions and identical innovation are often in conflict. When faced with a putative unusual change, like $f \rightarrow p$, does one conclude that it is strong subgrouping evidence or that it is so unlikely that the

/p/ forms are shared retentions? *Second*, if the only good subgrouping evidence is evidence from unusual, unnatural changes, then, by that very token, such evidence will be in short supply, and it will be impossible to construct good subgrouping arguments simply because the evidence won't be there! *Third*, premise (ii) does not entirely rule out the possibility of unnatural change. There is very little to guide us in recognizing when an unnatural change actually *has* taken place. *Fourth*, and most damaging of all, is premise (ii) itself. There is, in fact, no theory of phonology or of sound change by which changes can be ordered for naturalness. Modern phonological theory, in a diachronic guise, can be interpreted as an exercise in motivating all observed phonological alternation and sound change. Our notions regarding naturalness are grounded in nothing more than vague intuition and anecdote. In the absence of a true theory of relative naturalness, the use of premise (ii) in subgrouping arguments is, quite literally, unmotivated.

In simple genetic arguments using the comparative method, accidental similarity and borrowing, as accounts of similarities between forms, can be eliminated for the most part by restricting data to symbols and by the regularity assumption, respectively. There are no parallel means for eliminating diffusion and identical independent change, in a principled fashion, as accounts of shared *changes* in subgrouping arguments. Diffusion, it seems to me, is never going to be easy to rule out, except in cases in which the putative subgroup is geographically discontinuous (but see further below). To rule out identical independent development, we must rely on premises (i) and (ii) above, and they are far from unproblematic.

Eliminating "shared retention from an earlier antecedent state" as an account of similarities in outcome is a problem unique to subgrouping. The comparative method can give us no guidance, so we must again depend on heuristics like those following from premises (i) and (ii). As an example of the problems involved, consider the case of the Romance verb "to eat":

Portuguese	comer
Spanish	comer
Catalan	menjar
French	manger
Italian	mangiare
Romanian	mînce

For convenience, I label the two roots in question C and M. It would appear at first glance that, for lexical data like this, we can at least rule out the possibility of identical independent change. And, for the sake of this argument, I ignore the possibility of diffusion. Three possibilities remain:

- i C is a retention, and M an innovation (of subgroup {Cat, Fre, Ita, Rom});
- ii M is a retention, and C an innovation (of subgroup {Por, Spa});
- iii both C and M are innovations (and evidence of two subgroups).

The “right” answer is iii, more or less. Both C and M are reflexes of Latin verbs: *comedere* ‘to eat out of house and home’ and *manducare* ‘to chew.’ So both forms are in fact retentions. The innovation is the loss of the original Latin verb *edere* ‘to eat,’ and its replacement by two distinct alternatives from the common Latin lexical stock. The act of replacement involved a semantic change in the replacing forms.

We know enough about the history of Romance to be able to recover the right answer in this case, and it is not obvious how one would use these data as subgrouping evidence otherwise. It might be objected that in “real” subgrouping, one has access to a large number of correspondences, and that this quantity of evidence affects the quality of the resulting argument. In other words: *the more numerous are the changes shared by a set of languages, the more likely that set is to be a subgroup.* For lexical data, this reasoning is valid. If we had 10 cases like the C/M case above, all distributed the same way, we would still not be able to distinguish the innovating subgroup from the remaining languages retaining the proto-forms. We might want to rule out (iii) (rightly or wrongly), on the grounds that 20 changes in two subgroups are less likely than 10 changes in one subgroup and 10 retentions. This reasoning may not be acceptable since, by the same token, one change and one retention is better than two changes. But we wouldn’t be that much farther ahead in any case.

I chose a lexical example to highlight the problem of identifying shared retentions. Sound correspondence data don’t fare particularly better. For sound correspondences, we can rule out the possibility of *both* forms being retentions, but the problem of distinguishing retention from innovation remains. Two sorts of argument are often used in such cases. One, exemplified in the hypothetical sound correspondence above, is that incorrect identification often leads to postulating unnatural changes. I won’t reiterate the difficulties associated with the notion “natural change,” except to note that this example was not entirely hypothetical, but is drawn from the correspondence set from which Proto-Micronesian *f has been reconstructed (see, for example, Jackson 1983: 352ff), and that the reconstruction entails the “unnatural” change *f > p in the Ponapeic languages.

The other is the quantitative argument noted above for lexical data, and it fares no better for sound changes. It might, however, be argued that the quantity of changes is some help in ruling out diffusion and independent innovation, from the premise that the more shared changes there are, the less likely they are to have diffused or arisen separately. However, the use of “more” in this subgrouping heuristic is problematic. Exactly how many shared changes does it take to make a subgroup? This question is not entirely a facetious one, if one considers a situation in which each of the subsets of the languages concerned shares some number of changes. Short of a “subgroup constant,” this heuristic seems to imply that subgroup membership is relative; that is, that we use a wave model of relatedness, rather than a family tree. And in that case, the subgrouping issue becomes moot.

What is perhaps the least problematic basis for subgrouping is also the least linguistically interesting, and that is geography. *A historical outcome shared by L₁*

and L_2 is more likely to be a shared retention if L_1 and L_2 are geographically distant, and more likely to be a shared innovation if they are adjacent. That heuristic has traditionally had a role in the hypothesis that identifies the locus of change with an “innovative core.” While the logic of the geographic premise appears faultless, you really don’t need to know much linguistics to subgroup on that basis.

I despair for the subgrouping enterprise, then, because good subgrouping evidence is very hard to find and motivated subgrouping argument forms virtually impossible. Given this bleak scenario, it is unfortunate that comparative historical linguists cannot restrict themselves to simple genetic arguments, and just ignore subgrouping. Many comparative linguistics view their principal goal not to be demonstrating genetic relatedness, but producing a complete genetic history for some language family, in the form of a tree. I do not suffer terribly myself from “Darwin envy,” but I am interested in using the comparative method to do realist reconstruction of aspects of the grammar of a proto-language. One cannot select a proto-phoneme or a proto-lexical item, in any but the most trivial cases, without some subgrouping assumptions.

Indeed, I make subgrouping assumptions in my own work, though not without at least a twinge of guilt, because those assumptions are often not well motivated, and may often not be justified. But maybe I’m being too hard on myself; as important as it is to know what can be done, it is equally important to appreciate what it might not be possible to do.

4 Some Concluding Thoughts on Subgrouping and Method

Any historical enterprise is by nature limited, since time leaves only a very imperfect trace of its passage for subsequent generations to read. Modern comparative historical linguists are perhaps luckier than practitioners of other historical disciplines, though. Linguistic theories may change, but the majority of linguists, unlike our earlier nineteenth-century progenitors, do not believe that the essential nature of language has changed over the timespan with which comparative historical linguistics deals. In that respect, we may still have more in common with geologists and geomorphologists than with sociopolitical historians, many of whom in the present intellectual climate appear to feel constrained (or liberated!) to interpret history only in a contemporary context.

And we have the comparative method, from which genetic conclusions can be inferred from evidence of acceptable quality. Practitioners of other historical disciplines, archeologists for example, envy us that method and are often led to seek guidance from us as a result, in the mistaken view that comparative historical linguists can answer many of the questions that archeology cannot. The shoe is less often on the other foot.

But historical linguistics is not the comparative method. Much can be done through internal reconstruction, or with techniques that have as a premise just

the demonstration of genetic relatedness, without either subgrouping or comparative reconstruction. Much historical grammar is done that way.

Subgrouping has always been, for me, the soft underbelly of comparative linguistics, for the reasons outlined above. Subgrouping is not only methodologically problematic, but factually so as well, since we know that changes diffuse through the linguistic landscape, and give rise to the patchwork of isoglosses rather than the discreteness of trees. The status of subgrouping in comparative linguistics is similar to that of regularity; it is in fact questionable but in practice necessary. Subgrouping is necessary not for genetic inferences themselves, as pointed out above, but for realist lexical reconstruction. This is so because the phonetic content one reconstructs is a function of subgrouping assumptions (and assumptions about subgrouping like those considered in section 3.4.2). Whether or not one is interested in homelands and migrations, or in any similar issues in general prehistory, one must subgroup in order to reconstruct.

In section 3.2 it was observed that, though sound change is not regular, given sufficient time depth it gives the appearance of regularity. The same may be true for subgrouping in that, with a sufficiently long period of relative homogeneity and/or contact, a set of shared innovations (or, at least, the *appearance* of a set of shared innovations) may arise. But the number of actual cases for which that is demonstrably the case does not appear to be as large as those in which time yields the appearance of regularity.

As a consequence, if we want to do realist lexical reconstruction, it is standard practice to make subgrouping *assumptions*. If the views on subgrouping elaborated here are in any sense deviations from this standard practice, it is only in recognizing that subgrouping arguments are very seldom more than assumptions. But there's no shame in that. It is a mature discipline that has evidential standards, and that recognizes its own limitations.

ACKNOWLEDGMENT

I would like to thank Alan Dench and Brian Joseph for comments on an earlier draft of this chapter and suggestions that I hope have improved this one. The usual disclaimers apply.

NOTES

- 1 For a detailed explication of the comparative method per se, see Rankin, this volume.
- 2 See Thomason, this volume, for discussion of this point.
- 3 An explanatory “retroductive” theory of change, one that tells us how language states could/must have arisen, is probably a chimera, given that particular changes do not,

in fact, *have* to happen. My point is only that, if we had such a theory, we wouldn't need the comparative method.

I might also note the existence, since the nineteenth century, of a partial theory constructed along these lines, and used in conjunction with, or as a preliminary to, the comparative method. I refer, of course, to *internal reconstruction* (see Ringe, this volume), the technique of synchronic morphophonemic analysis in its historical interpretation. Internal reconstruction tells us that synchronic morphophonemic alternation is the result of conditioned change applied to antecedent non-alternating forms. We need only infer the precise changes involved to undo the alternations and recover the antecedent state. It is, after all, a partial theory!

- 4 I myself am a realist as regards reconstruction from the comparative method, *pace* such criticisms as those in Lightfoot (1979). I believe that we can use the comparative method for reconstruction, and that such reconstructions have the status of best approximations to antecedent historical states.
- 5 I will refer to these systems as *languages*, rather than use some less sociolinguistically charged term like *lect*.
- 6 The term *genetically related* is frequently paraphrased as "sharing a period of common history." Though I am not above using that paraphrase myself, it is dangerously vague in that it covers both relations through a common ancestor and relations through diffusion/contact/borrowing. A paraphrase like "having a common ancestor" is, strictly speaking, more accurate.

- 7 This characterization of the major warrant for genetic inference in comparative linguistics is a modification of that given in Anttila (1972: 302).
- 8 Many linguists might be tempted to turn off at this point; such is the discomfort conjured up by the very mention of the word "semiotics" in polite linguistic company. Permit me a slight departure from convention in presenting a very short anecdote that serves to demonstrate the power of ideology in modern linguistics, and the strength of the prevailing ideology's disdain for anything connected with the term "semiotics."

Some years ago I had the opportunity to give a graduate course I titled "Historical Grammar" to about a dozen students in an American linguistics department. One of those students was a recent transfer from a quite prestigious east-coast linguistics department. He was taking the course under some duress, to prepare himself for the historical linguistics component of the Ph.D. qualifying exam. I began the course much as I've begun this chapter, with a discussion of the goals of comparative historical linguistics and of the nature and limitations of the comparative method, particularly as regards investigation of the history of non-lexico-phonological aspects of language. In the course of that discussion, lecture 2 I think it was, I introduced aspects of the semiotic theories of Charles Sanders Peirce, in an undisparaging manner. At that point, the aforementioned student rose and left the room, never to return. He didn't pass the historical linguistics section of the qualifying exam that semester either. I returned to Australia shortly thereafter, and

- have no idea of his subsequent history.
- 9 For a contemporary view of iconic linguistic signs, see Haiman (1985a, 1985b). Of course, no onomatopoeic form and no metaphor is purely iconic; all have some measure of conventionality about them. But few linguists, I think, would want to argue that the sign 'moo' is as arbitrary as the sign 'cow,' though I am prepared to listen to any such argument! Indexical signs, in the sense I have in mind (as distinct from that in which deixis is indexical), do not seem to be relevant to natural language.
- 10 I will speak of genetic relatedness and cognacy as binary relations, but intend that the relations be generalizable to *n*-ary. I don't want to buy into the "binary comparison" issue (see DeLancey 1994a), except to say that I'm not convinced there's an issue.
- 11 The emphasis on *prove* is deliberate; saying two objects are cognate, and proving that they are, is not the same thing.
- 12 My choice of the indefinite article is deliberate, in allowing for the possibility of more than one reflex of the same antecedent object coexisting in a single language state. Possible examples are: French *le* 'the' and *le* 'him,' English *an* and *one*, or Spanish *muy* 'very' and *mucho* 'much'. And how does one talk about the relation between such items? Are they, for example, cognate?
- 13 As observed, for example, by Thomason and Kaufman (1988), the sort of acquisition and change involved in the pidginization phenomenon is not "normal" in the intended sense.
- 14 See the chapters by Bybee, Fortson, Heine, Hock, Joseph, Mithun, and Traugott in this volume for discussion, plus Janda (2001).
- 15 That's not to say that such a theory is not heuristically useful; only that it's not necessary.
- 16 This is the view of sound change suggested by Labov in published work as early as Labov (1972) and, more recently, in Labov (1994).
- 17 A problem like that of multiple reflexes of the same historical segment is no worse for this view of cognacy than is the problem of multiple reflexes of the same lexical item, noted in n. 12.
- 18 These issues were the subject of a thread of discussion begun by Fritz Newmeyer on 30 November 1994 (see Newmeyer 1994) and dealing with "the applicability of the comparative method to syntax." As is often the case in such discussions, there was some confusion regarding exactly what was, or should have been, at issue. Many of the contributors were concerned as much or more with the proper delimitation of the question as with the answer. Should the term "syntax" in this context refer just to constituent order, should it include category systems, paradigm structure, and so forth? However, I was particularly struck by the view put by Karl Teeter: "If one can include a section on syntax in a grammar, one can apply the comparative method in syntax" (Teeter 1994a). As my remarks above might suggest, I have seldom come upon a methodological assertion with which I disagree more. On the other hand, I have strong sympathy for his assertion that "when I do linguistic history I write a grammar of a protolanguage" (Teeter 1994d), if what he means is that one must aim at reconstructing a coherent

- fragment, however small, of a possible natural language.
- 19 This insight is made particularly salient in the fact that the same attribute-value matrix representations are used in HPSG for signs of all types.
- 20 One might be tempted to stress that sentences, and other syntactically complex signs, have information about their component parts. But the same is true of morphemes too; it's just that for the latter the information is "phonological," while for the former it is (more critical) "syntactic" information. I'll do my best to avoid that minefield here.
- 21 Such associations of grammatical form with meaning were long deprecated in "standard" generative grammar, it seems to me, as a consequence of Chomsky's strong insistence, in the past, on the "autonomy of syntax."
- 22 There is perhaps a paradox, not often noted, in the fact that some linguistic objects are reconstructible without counting as evidence of genetic relatedness. The limiting case for such objects is linguistic universals. If one believed, for example, that all languages have a categorial distinction between nouns and verbs, then one has licence to reconstruct that distinction in any proto-language. But since such reconstructions do not depend on evidence, or depend on evidence that holds equally over unrelated languages, it is of no value in determining genetic relatedness.
- 23 Any reference to the semiotic properties of syntagmatic objects is rare in the historical linguistic literature. An exception is Anttila (1972: 195), who points out that "rules are largely iconic," but does not elaborate.
- 24 I use the example of Greenbergian clause typologies because of its importance in the literature on word order change. Of course, the facts of word order are often more complex than can be accommodated by simple statements that, in L, transitive clause order is one particular permutation of S, O, and V. "Fixed word-order" languages often show more than one order of major constituents in transitive clauses, under grammatically well-defined conditions. Such observations have no direct bearing on the issues I raise here, but the same is not true of the problem of identifying subject and object in ergative languages. The universality of the subject and object relations is the core of the problem – see below.
- 25 Many languages admit a fourth possibility in the plural, in distinguishing those speaker-inclusive groups that include the hearer from those that don't.
- 26 Classical Arabic has distinct dual pronouns in the second and third persons masculine. The same was apparently true of Ugaritic (see Pardee 1997: 133–4), which had an additional distinct first dual suffixed pronoun as well. The only modern Semitic languages with dual pronouns are Eastern South Semitic languages like Mehri and Soqotri. These forms do not appear to be cognate with those of Classical Arabic, however.
- 27 See Ringe, this volume, for discussion of this method.
- 28 With more than a little justification, Brian Joseph (pers. comm.) objects that it is perhaps more accurate to describe Classical Latin as having had "free" word order. One could always consult the statistics on word order in the Classical Latin prose corpus to help decide whether or

- not OV was the unmarked order. I've not sought out those statistics, since I offer this example for illustrative purposes only.
- 29 Similar observations can be made regarding the English compound data. Brian Joseph (pers. comm.) points out that compounds like *pick-pocket* and *turn-key* are instances of a non-productive, and thus perhaps archaic, mechanism for forming verb-object compounds in VO order in English. It is the OV order that is productive.
- 30 In preparing drafts of a grammar of Gilbertese, I endeavored to construct just such an analysis, but ultimately gave up the attempt.
- 31 In my own linguistic area, Oceania, I might note the pioneering work of Pawley, of Clark, and of Chung on Polynesian articles, prepositions, and verb morphology, and some of my own efforts in Micronesia.
- 32 DeLancey makes the same point (1994b).
- 33 This same phenomenon is found as well in Hittite and in Vedic Sanskrit.
- 34 In a reply to Manaster Ramer, Valiquette (1994) suggests that the Iranian/Greek oddity might not be all that odd, but is a consequence of the generalization of a collective interpretation for neuter plurals. Since I'm not an Indo-Europeanist, I can't comment.
- 35 See Pawley and Green (1984: 139–42) for some discussion.
- 36 Rate of change may itself be the "problem" for the comparative method. If some language or set of languages changes very quickly, then it is that fact, rather than the absolute time since the onset of differentiation, that trips up the comparative method. A rapid rate of change may lead some language(s) to be underrepresented in reconstructions, as Grace (1985) suggests has been the case in the reconstruction of Proto-Austronesian and its descendants over the last century. Though I have felt personally slighted in the past because the Micronesian languages on which I was working were largely ignored in reconstructing Oceanic, on reflection it would seem that there is logic in putting greater emphasis on languages that (are believed to) have changed least. It is the same logic used when one puts greater emphasis on Greek and Sanskrit (or, perhaps, Icelandic and Lithuanian) than on Romanian and Afrikaans in reconstructing Proto-Indo-European (PIE).
- 37 I might note that the same problem had been recognized for the Micronesian languages. I was privileged to be part of a group at the University of Hawaii that applied the same logic to integrating Micronesian languages into Oceanic. In our case, however, the logic worked.
- 38 As Brian Joseph has reminded me (pers. comm.), Sihler (1995: 7) makes a similar point about the importance of shared innovations as opposed to shared retentions by means of an analogy, noting that subgrouping is rather like club membership: "Members of a club have something in common – they joined the club; but the people in the community who are not members of the club do not constitute a second de facto club."