

Linguistic reconstruction: methods vs. interpretations*

László Kristó

vernerslaw@yahoo.com

0 Objectives

The aim of this paper is to investigate the central methods of linguistic reconstruction and the theoretical models associated with them. By “central methods” I understand the two basic ones, Comparative Reconstruction (CR) and Internal Reconstruction (IR). They can be considered central because, despite the advent of other methods such as glottochronology, they still form the nucleus of reconstructive techniques, without which no serious reconstruction can be attempted. This is because both methods, unlike other, more marginal ones, are strictly based on the absolute *sine qua non* of any historical linguistic study: the Regularity Hypothesis. Both CR and IR will be considered in the light of their alleged theoretical background, which for CR is held to be the Neogrammarian model, while IR relies on structuralism. I will come to the conclusion that this distinction is irrelevant for the method itself, because Neogrammarians and Structuralists differ not so much in the method but in the interpretation of the results of reconstruction; the difference follows from the different theoretical models of phonological change. I will point out, furthermore, that because IR is not a historical method as such but, instead, it is the historical interpretation of a basically non-historical method (from which it follows that it has serious flaws), CR still remains as the central (possibly only) really historical and exact method of reconstruction, but, interestingly, not because it is inherently historical, but thanks to the nature of the data it works with.

This paper is structured as follows. In Section 1, I provide a brief outline of the Comparative Method, followed by a summary of Internal Reconstruction in Section 2. Section 3 discusses the differences between how the Neogrammarians and the Structuralists differ in their views on reconstruction. Finally, Section 4 concludes the discussion.

1 Comparative reconstruction

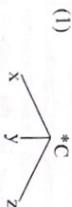
The theoretical basis of CR is the Neogrammarian doctrine known as the Regularity Hypothesis. In its strongest form, it claims that all sound change is regular in the sense that it occurs according to fully mechanical phonetically conditioned rules, operates “with blind necessity” (it is not aware of its own consequences), and due to this it admits no exception. To take a simple example, if in a language L voiced stops are lenited to fricatives intervocalically, they will always do so in that environment, and no word can escape the consequences of the change. The Regularity Hypothesis makes it possible to set up regular sound correspondences between related dialects¹. To return to the previous example, if L has a relative L' in which the given change does not take place, then any intervocalic voiced fricative of L regularly corresponds to a homorganic voiced stop in L', and vice versa.²

* My heartfelt thanks go to my professor and colleague Ádám Nádasy and my colleague András Cser, who have made extremely valuable comments on an earlier version of this paper. Of course, I am alone responsible for remaining errors (especially because I have not followed some of their suggestions).

¹ Or languages. I will use the terms *dialect* and *language* interchangeably, since there is no principled difference between them from a comparative viewpoint.

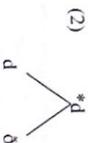
² Assuming, for the sake of simplicity, that no other change has taken place in either language as well as that there are no voiced fricatives in the proto-language, at least intervocalically.

The existence of regular correspondences is what makes CR possible. The method is a rather simple mathematical operation in its first stage, as pointed out by Lass.³ I will use this example here. Take three related languages X, Y, and Z; let $x, y,$ and $z,$ respectively, be regularly corresponding units (sounds, for our purposes) in these languages. We can create a set $C = \{x, y, z\}$, where C stands for "cognateness". This, of course, is a simple relation, which is symmetrical ($x \in C_y \supset y \in C_x$) as well as transitive ($(x \in C_y \text{ and } y \in C_z) \supset x \in C_z$). So far C is nothing but a label to give a name to the set; as a result, we have not yet made any historical statement: all we have is a static pattern. We can, however, move on to the second step and assume that C is not a label but an entity, i.e., a physically existing object which is related to $x, y,$ and z historically: it is their ancestor. The relation "ancestor of", of course, is neither symmetrical nor transitive. We can conceptualize the relation in the form of a tree as in (1):



If C is a physically existing object, it is possible (or even obligatory; see Section 3.1.) to assign physical properties to it. In case x, y, z are sounds, these will be, of course, phonetic properties. This is the final stage of reconstruction (at least on this level; we can move one level up and reconstruct morphemes, etc.).⁴

Let us take a specific example, using material illustrating a correspondence that derives from intervocalic spirantization. Standard Portuguese intervocalic [d] regularly corresponds to [ɔ] in Castilian Spanish: Portuguese *canhada, lado, cidade* ('sung-FEM', 'side', 'city') correspond to Castilian *cantado/a, lado, ciudad*, respectively. We can set up $C = \{d, \emptyset\}$, and assign C historical-ontological status, which requires (or makes it possible) to assign a phonetic value to it. Since we know that intervocalic position is a typical lenition site, we assume that $C = [d]$ in Proto-Western-Romance, the ancestor of both languages.⁵ We can represent this as in (2):



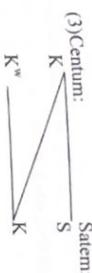
This is simple so far: we assume a change in one language. The projected proto-character (sound) is thus represented in a presumably unchanged form in one descendant. But this is not always the case: there are many instances where the comparative method requires one to reconstruct something that hasn't in fact survived. A classical example of this is what can be labelled the "back stop series" of Proto-Indo-European (PIE). For a detailed account, see any good textbook on Indo-European linguistics, such as Szemerényi (1990): I will give a simplified and rather abstract presentation here. I will use the following abbreviations: K = velar stops, K^w = labiovelar stops, S = sibilants, P = palatal stops.

³ 1993:161f.

⁴ Cf. Fox (1995:58ff) for proposed stages of reconstruction.

⁵ This stage itself represents a lenited state of intervocalic stops, since the ultimate source of these cognate items is Vulgar Latin **canhada*, **lado*, **ciuitate*, with an intervocalic voiceless stop.

The Indo-European languages have been divided into two major groups labelled "Salerni" and "Centum". Between the two groups, the following correspondences hold:



We can make three correspondence sets:

$$(4) \quad \begin{array}{l} C_1 = \{K, S\} \\ C_2 = \{K^w, K\} \\ C_3 = \{K, K\} \end{array}$$

The Neogrammarian reconstruction of the three proto-segments is as follows⁶:

$$(5) \quad \begin{array}{l} *P = \{K, S\} \\ *K^w = \{K^w, K\} \\ *K = \{K, K\} \end{array}$$

The crucial point is that no daughter language has palatal stops deriving from the proposed *P series: it is reconstructed only because there are three correspondence sets, hence there ought to be three proto-segment series. I will return to the significance of this fact in Section 3.2; let us now turn our attention to IR.

2 Internal reconstruction

The best-known early application of the method known as IR was Ferdinand de Saussure's influential *Mémoire* (1878), probably one of the most important books ever written on a linguistic topic.⁷ Saussure used the method to reconstruct the phonological system of Pre-Indo-European, i.e., the stage preceding PIE reconstructed via CR.⁸ The essence of IR is that it starts out from (*non-suppletive*) alternants within one language at a given time; assuming that the alternation (i.e. non-identity) reflects earlier identity, i.e., it arose at some point in the history of the language due to some sound change(s); it attempts to reconstruct the original single form which the alternants are derived from by regular sound changes.⁹

Take as an example the word-final devoicing of stops in German. Let T be any voiceless stop and D any corresponding voiced one, where corresponding, of course, means that the difference between T and D is in voice only. There are many stems which sometimes occur with final T (if it is also word-final), sometimes with a final D (if it is not word-final but followed by suffixal material). So, for example:

⁶ I neglect the detailed argumentation here, since it is quite immaterial for this discussion.

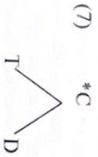
⁷ Not the first one, though, in spite of the fact that this is what textbook wisdom says. As Andreas Cser has pointed out to me, some earlier Neogrammarian laws had already been arrived at — partially, at least — by the application of IR. The most notable example is probably Verner's Law, based besides comparative evidence — on voiceless — voiced alternations within Germanic itself.

⁸ Szemerényi 1990:86–97, 127–137.

⁹ Lass 1997:232–41, Anttila 1989:264–73.

- (6) Ra[ɫ] 'wheel-NOMSG' ~ Ra[ɫ]dɛs 'wheel-GENSG'
 We[ɫp] 'woman-NOMSG' ~ We[ɫ]bɛs 'woman-GENSG'
 Ta[ɫk] 'day-NOMSG' ~ Ta[ɫ]ɛ 'day-NOMPL'

We can now set up a correspondence set $C = \{T, D\}$ as the first step. As the second step, we interpret C as the ancestor of both T and D, as in (7):



As the final touch, we assign phonetic properties to *C. We can safely assume, on both theoretical and language-internal grounds, that *C = D, i.e., historically, such occurrences of voiceless stops derive from voiced ones via the regular sound change of Word-final Devoicing.

Anyone familiar with phonological analysis will have noticed that this is the same as what one does in a process-based paradigm (such as the SPE model) when analyzing synchronic alternations. Indeed, Anttila says, "Internal reconstruction (...) is *exactly* the same as morphophonemic analysis" (emphasis mine).¹⁰ Please note the word *exactly*: it implies that there is no difference between the two things. Indeed there isn't: the method is exactly the same. But then, where is the difference? After all, synchronic analysis is *not* internal reconstruction. The answer is that the difference lies not in what one does but how one interprets the results. In synchronic analysis, we set up C as a set and assign a theoretical status to it, and we may as well stop there, but we can go on and claim that the alternants are actually derived from it (if we believe in phonological processes); in other words, we can regard C as an underlier. Internal Reconstruction is none other than assigning historical status to C, that's where IR is, for the historian, more than simply a synchronic analysis: that's why it's something historical. In other words, set up an alternation, label it, and whether you do IR or synchronic analysis depends on the content you give to your set: in synchronic analysis, it is "alternates with"; in IR, it is "cognate with".¹¹ I will return to this point later, but now let us see the limits of IR and its fundamental dependence on CR.

As Anttila says, IR is but morphophonemic analysis as far as the method goes. In fact, I take the opportunity to correct Anttila here: IR is not necessarily based on *morphophonemic* alternations, although this is indeed the majority case: any purely *phonologically governed* alternation is liable to such an interpretation. (See below for such an analysis of Non-thoticity.) Second, as alluded to above, I must disagree with Anttila in equating IR with synchronic analysis: IR is not the same as synchronic analysis: it is a historical interpretation of the same data as used for synchronic analysis. Or, to put it differently, IR = synchronic data + historical interpretation.

¹⁰ 1989:264.

¹¹ "Cognate with" is understood here, of course, in a non-comparative sense (roughly, "having the same ancestral form").

Nonetheless, IR has serious flaws. I illustrate this with two examples. First, take English Spirantization, illustrated by pairs like *defend* ~ *defens(ive)*, *omit* ~ *omission(s)*, etc. At first sight, we might be tempted to use IR to reconstruct an earlier single stem form underlying the present-day alternation. But we know that these words are Latinate borrowings, in which the alternation is already present: in other words, English borrowed the alternation hand in hand with the words. It would be wrong to assume a Spirantization Rule as a sound change in the history of English: a synchronic rule, then, is not necessarily a historical change.

Second, although an alternation may point to a historical change, it may do so in the wrong way. Consider those non-rhotic accents of English which have obligatory, full R-liaison, i.e., both Linking and Intrusive R, such as London English. This means that a set of words ends either in a non-high vowel or a non-high vowel + R, as in *car* /ka: ~ kar/, depending on what follows the word. In a synchronic analysis, we can assume a rule of R-insertion to handle the alternation. But we know that historically, there are two distinct processes: (1) R-dropping, (2) R-insertion. If we did not have any historical information at our disposal, we could not choose which process to assume: it is due to the testimony of other accents (as well as orthography, grammatical descriptions, etc.) that we know what happened. For example, take the words *spa* and *car*: both have R-ful and R-less alternants, in exactly the same environments; there is no difference between the two words. Historically, though, one of them is R-ful, the other R-less: but based on the present-day language alone, we can't tell which is which. The appearance of Intrusive R results in what we can regard as a kind of merger: the historically distinct categories -V# and -V# merge, yielding a situation where they have become context-dependent variants. This reflects a basic problem one must face when doing IR: unconditional merger, which renders previous contrasts unrecoverable for the method. To sum up, IR requires comparative backup, and therefore, it is insufficient to solve this particular problem.

This much has often been said. Yet, we must be careful here, because CR is not almighty, either. Consider another type of accent, in which there is no R-liaison whatsoever, such as Southern US English (SUSE). Here, it would not even occur to anyone to reconstruct anything, because we have no alternation: *car* is always pronounced /ka:/. Let us now imagine the situation that all we have access to is SUSE and London English. In an analysis of the latter, we are faced with the problem described above: but would SUSE provide the necessary comparative backup? It would not. We are still faced with the same problem, because the difference between two accents can still be accounted for in two ways: either by assuming that SUSE is conservative and London E innovates (via R-insertion) or that SUSE is innovative (R-dropping). We need even further comparative support, either from thotic accents or from ones which have linking but no intrusive R (if there are any such accents left, maybe conservative RP speakers have it). The possibilities of thoticity and R-liaison are summed up in (8):

- (8) > London E: /ka: ~ kar/, /spou: ~ spou:
 > Southern US E: /kar/, /spou:
 > GenAm E: /kar/, /spou:
 > ? Cons RP: /ka: ~ kar/, /spou:/

To sum up this lengthy discussion: the fact that IR is not flawless is not in itself an argument against it or in favour of CR, because CR is not flawless either.¹² The point is that when we have access to both IR and CR, and the two disagree, CR takes precedence, but this is a logical consequence of the fact that CR works with data from several dialects. It is in this sense only that CR is superior.

3 Comparative reconstruction: the method vs. its interpretations

In Section 2, we saw that IR is not an independent historical method, but simply the historical interpretation of a fundamentally non-historical procedure. Furthermore, it has been pointed out that, whenever possible, it must be checked against results provided by CR. In this section, I will examine CR against the two theoretical models widely used in reconstruction: the Neogrammarian and the Structuralist models. I will argue that the two theoretical frameworks differ not so much in their reconstructive methods but rather in how they interpret the results. First, however, I will discuss two differing views on the status of reconstructed entities: the "idealists" versus the "realists" positions.

In Section 1, I established three steps of reconstruction: (1) setting up the correspondences; (2) assigning historical status to the set label; (3) assigning phonetic properties to the reconstructed item. Step 1 is the basis for any reconstruction whatsoever, but what about the rest? Do we really need to go on? In fact, why assign phonetic reality to our hypothesized entity? Why not omit Step 3? Or, why not omit Step 2 as well, and say, with Meillet that "the reconstructions are merely symbols with which we express the correspondences in an abbreviated form."¹³ Indeed, this stance has been taken by many linguists including Meillet and it has been labelled the "idealist" or "formalist" position (as opposed to a "realist" stance).¹⁴ For an idealist, then, there are either no "proto-segments"; reconstructed forms are just set labels; or, a bit less abstractly maybe, "proto-segments" are not labels, but quite abstract (past) entities, whose phonetic content is immaterial; what counts is that the entity underlies the correspondence set. Few (if any) historians would nowadays take the rather extreme position formulated by Meillet, and it must be pointed out that Meillet himself was much more a "realist" in practice. I'll return to this point below, but first I will consider CR as a historical method; then I compare the Neogrammarian and the Structuralist models as far as their reconstructive techniques are concerned.

Step 2 of (any) reconstructive technique, i.e., the assignment of some "thing" status to a set label, yields tree diagrams such as the one in (1), repeated here as (9) for convenience:



This is what we get either by IR (= synchronic analysis) or by CR. In both cases, take a correspondence set, label it, and call the label an object, which "turns into" other

¹² In fact, no scientific method is flawless, because we can only use what we have access to; as formulated so beautifully and appropriately by the British physicist John D. Barrow in *The world within the world*, our universe is what we can see; what we cannot is beyond it.

¹³ Fox 1993:7-17, Lass 1993, 1997:270ff.

objects. In synchronic analysis, the "object" is an underlying unit from which surface forms are derived by (ordered) rules; in IR and CR, it is a historical (past) object, from which attested (present) ones derive by chronologically ordered sound changes. The method, then, is no different, and we have seen this already in the case of IR and synchronic analysis. We can now see that CR is based on the same method. There is, however, a crucial difference: *x, y, z* are, for CR, from different languages. This is very important, because *C*, whatever it is, cannot be sensibly interpreted as a common "underlying representation"; the only sensible interpretation is genetic, i.e., historical: the relation "cognate of" is reinterpreted as "ancestor of". Alternatively, we are left with the possibility of not interpreting it at all, or rather, not even calling it an object: this is basically the idealist position. If one prefers not to be an idealist, one must interpret *C* as a historical object; as it is a historical object, from which physically existing present-day objects are derived, it must have physical (phonetic) form. Note, however, a very important point: CR as a method does not start out as a historical method; either: the basic procedure is the same as for synchronic analysis or IR. What makes CR historical is the nature of the data it has to work with: that the data are from related languages, assumed to be related because they derive from a common ancestor: relatedness equals ultimate monogenesis. CR is not inherently historical: it is forced to be historical. A synchronic alternation (within one dialect, of course) in itself does not force one to do a historical analysis. But if one does choose to do so, it is called IR. We can sum up the similarities and differences in a table, as in (10):

(10)

	Sync. analysis	IR	CR
Set up correspondence?	Yes	Yes	Yes
Content of set?	"Alternates with"	"Cognate of"	"Cognate of"
Interpretation?	"Underlier of"	"Ancestor of"	"Ancestor of"
Phonetic content?	Yes/No	Yes/No	Yes/No

The position of a full interpretation up to assigning phonetic content, is basically the position of the Neogrammarians, for whom the question of "idealism" would not have occurred. As opposed to them, the Structuralist school found it not only acceptable, but often desirable (of course, not necessarily everyone!). Of the two central reconstructive techniques, CR and IR, the former is said to be a Neogrammarian invention, while IR is considered to be a structuralist one. We have seen that the two methods are fundamentally the same in their procedures: it is the nature of the data they differ in. IR, then, is none other than CR applied to a different type of data, and as such, it does not constitute a principally new method.¹⁵

Furthermore, let us recall the Neogrammarian (comparative) reconstruction of the "back stop series" of PIE (cf. (5) above). Why are there three different series reconstructed? After all, no daughter language has so many; they have two series only. The answer is, because there are three different correspondence sets. There's no direct phonetic proof for the "palatal" series at all; the whole system is typologically suspicious (it is at least highly unusual for a language to have palatal *and* velar *and* labiovelar stops); moreover, the [K,K'] correspondence is extremely rare. Why did the Neogrammarians reconstruct such a system? Because they applied the comparative formula in a rigorous manner. In this sense, they were more structuralist than the Structuralists themselves. It is highly enlightening to consider Meillet's

¹⁵ Anttila 1989:229.

position on the PIE stop system: he (and many other linguists) represents the view that PIE had only two back series, *K^w and *K; the Centum group preserves both intact, while the Satem group palatalizes the velar (*Kj) series (which then ultimately turn up in the atested languages as sibilants) and de-labializes the labiovelar series. Meillet uses typological and phonological-historical arguments against the Neogrammarian position, pointing out that the paucity of {K,K} correspondences may be an indication of "deviant" developments. Interestingly, Meillet is, in fact, much less "structuralist" than the Neogrammarians (and, moreover, a realist in practice).

The two schools, then, do not really differ from each other so much as far as reconstruction goes. The main difference is in the interpretation of the results. We can now ask where the differences come from. The answer lies, as far as I can see, in the different conceptions of phonological change. Neogrammarians thought in terms of sound change; Structuralists concentrated on changes in the phoneme system. For them, a sound change in itself was not a linguistic change if it did not alter the phoneme system. For example, if in a language [u] is fronted to [y], which is a novel decrease, only the phonetic realization of one phoneme alters. This, of course, is an abstract view that few if any phonologists would nowadays take.¹⁶ But it explains the origins of the idealist position: it is phonemic differences that count, so we must reconstruct phonemes (we cannot reconstruct allophones anyway, unless we have some special clue). Phonemes, however exactly one conceptualizes them, are abstractions. This is what makes the idealist position possible (but not obligatory: that is a further step). Neogrammarians dealt with sounds, hence idealism was impossible for them.

4 Conclusion

To sum up, the basic difference between Neogrammarians and Structuralists is that for Neogrammarians, all three steps of reconstruction are self-evidently compulsory; Structuralists can omit the last step(s). As far as IR is concerned, it is but a historical interpretation of a non-historical method. This is not to deny its significance: there are many instances where comparative evidence is not available, either because the language has no (close) relatives or because we want to reconstruct earlier stages of a proto-language (as Saussure did). But the central core of reconstruction is still the comparative method: IR is the application of CR, a method devised to handle data from a historical perspective, to data which need not be interpreted historically. Furthermore, IR — generally associated with Structuralism — is not a Structuralist invention: the 1870's had seen several examples. Linguistics may have undergone revolutionary changes during the twentieth century, but in reconstructive techniques, we still use what our Neogrammarian predecessors invented. This note is not meant to devalue the work of twentieth-century historical linguists (including Structuralists — they have done an excellent job, especially in our understanding of language change) and to imply that we have no reason to be satisfied, but to turn the reader's attention to the extraordinary achievement of the old nineteenth-century masters — who deserve the highest respect and whom we can take pride in.

¹⁶ In fact, Structuralists were also interested in the phonetic content of oppositions, so I am being somewhat unfair, my excuse is that I use these *extreme* examples to refute an *extreme* position.

References

- Antila, Raimo (1989) *Historical and comparative linguistics*. 2nd ed. Amsterdam and Philadelphia: John Benjamins.
- Bynon, Theodora (1977) *Historical linguistics*. Cambridge: Cambridge University Press.
- Comrie, Bernard (1993) 'Typology and reconstruction' in Charles Jones (ed.) *Historical linguistics: Problems and perspectives*. London and New York: Longman.
- Fox, Anthony (1995) *Linguistic reconstruction: An introduction to theory and method*. Oxford: OUP.
- Lass, Roger (1993) 'How real(ist) are reconstructions?' in Charles Jones (ed.) *Historical linguistics: Problems and perspectives*. London and New York: Longman.
- Lass, Roger (1997) *Historical Linguistics and Language Change*. Cambridge: CUP.
- Lehmann, Winfred P. (1992) *Historical linguistics: An introduction*. 3rd ed. London: Routledge.
- Meillet, Antoine (1964) *Introduction à l'étude comparative des langues indo-européennes*. University of Alabama Press.
- Saussure, Ferdinand de (1878) *Mémoire sur le système primitif des voyelles dans les langues indo-européennes*. Leipzig: Teubner.
- Scemeryni, Oswald (1990) *Einführung in die vergleichende Sprachwissenschaft*. 4th ed. Darmstadt: Wissenschaftliche Buchgesellschaft.