

THE NATURE AND USE OF PROTO-LANGUAGES

Summary

While it is true that languages related through (as one says) common descent are derived from a common ancestor language, this idiom, if unattested and available only in the form of a reconstructed proto-language, has only a limited degree of realism. The manner in which linguistic reconstruction proceeds, the lack of truly quantitative criteria in determining type and degree of linguistic relatedness, and the necessity to arrive at an entirely uniform linguistic construct are all apt to deliver a distorted or false view of the proto-language. Indeed its very existence may be questionable, especially if it cannot be supported by non-linguistic evidence; this applies in particular to intermediate reconstructed proto-languages like Proto-Italo-Keltic, Proto-West-Germanic, Proto-Ibero-Romanic, which are but way-stations on the road to the ultimate parent language. It is therefore suggested that all proto-languages be considered creations for the convenience of linguistic investigation and for the purpose of delving into an otherwise inaccessible linguistic past, but that no claim should be made for their being viewed and dealt with as real languages in any sense of the word, unless and until there accrues sufficient non-linguistic evidence for fixing them in time and place and for associating them with an anthropologically, archaeologically, or historically identifiable society. The reverse process, that is, the creation of a society to go with an unattested, reconstructed proto-language, is altogether improper.

In a recent article I proposed the view that Reconstructed Proto-Indo-European (which is unattested, almost certainly unrecorded) and Real Proto-Indo-European need not be and in all likelihood are not the same linguistic item, that indeed Reconstructed Proto-Indo-European, no matter how rich our collection of asterisked formulae may be or yet become, does not represent a real language at all under any acceptable definition of the term language¹). This charge of unreality applies to the number of reconstructions producible, since one cannot reconstruct a dead and lost language in its entirety (what is, in any event, the 'total' lexicon of a language?), and even more significantly to their linguistic shape, since one cannot be certain, hence should not assume, that one has correctly reconstructed real, synchronically existing forms from a variety of non-

¹) See Ernst Fulgram, Proto-Indo-European reality and reconstruction, *Language*, 35,421-6 (1959).

synchronic evidence. For example, if one were to reconstruct the proto-language of the modern Romanic languages, ignoring for the sake of the experiment that it is available in the form of ancient Latin of one kind or another, one could neither reconstruct the entire Latin vocabulary as we know it to have existed (a number of Latin words are not continued in any Romanic dialect), nor could one, from the evidence of the living Romanic dialects, reconstruct a language of more than three cases, or guess the existence of deponent verbs, or discover that at least one kind of Latin, the Classical Latin of metric poetry and possibly prose, had significant vocalic quantity, and so forth. Indeed the shape of every reconstructed form is entirely dependent on the type and amount of evidence available: that is, a Proto-Indo-European form reconstructed from Sanskrit and Hittite will be different from what it would be if the records came from Slavic and Germanic; also, a Proto-Indo-European formula based on ten available cognates looks different (and probably comes closer to reality) from what it would be if only two cognates were available.

In this article I am concerned not with the state of our knowledge in any one linguistic domain, but rather with a question of principle, namely: If the shape and the very existence of an unattested proto-language are sustained *entirely* by the assumption, unsupported by non-linguistic evidence, that several related dialects must of necessity be derived from, and may therefore be employed for the reconstruction of, this proto-language, what can be the degree and the kind of authenticity of such a reconstructed parent language, and of any non-linguistic corollary hypotheses based upon it? In other words, at what point in our reconstructive process, and under what circumstances, may we become prone to desert reality, linguistic or non-linguistic, and where do we fully yield to absurdity?

My line of argument will not entail a rejection of the comparative method of historical linguistics, nor generate an injunction against the reconstruction of proto-languages for purposes of historical linguistics. But I do want to call attention to the inherent limitations of these procedures, to the fallacies and delusions which are bound to attend the rash endorsement of some substantive deduction drawn from the insubstantial presence of these creatures fashioned by ourselves, often with more confidence than prudence.

When we are dealing with the recorded language of a historically

known society, including of course the audible idiom of an existing speech community, acquaintance with the speakers is implied to a greater or lesser degree since we have non-linguistic evidence to tell us more or less precisely what kind of people, having what non-linguistic attributes and history, spoke or speak Greek or Sanskrit or Turkish. Identification of the linguistic with an historic extra-linguistic entity can thus be accomplished.

At times, on the other hand, we may be fairly well informed on a language, but our knowledge of the society that goes with it is scanty or altogether lacking. Such is the case, for example, with Crimean Gothic, ancient Umbrian, or Tocharish. Naming the speakers Goths, Umbrians, or Tocharians is merely a process of tautological labeling that tells us nothing about these persons' non-linguistic identity. In instances of this type one may or may not be fortunate enough, after careful examination of the total record, to arrive at a reasonable conjecture as to who spoke these languages.

But we may also find ourselves confronted with an asterisked proto-language, consisting of reconstructed formulae. Unless there is good non-linguistic evidence that this proto-language actually existed, any *a priori* association of speakers with this fictitious tongue is absurd. An unattested proto-language presents us therefore with the problem of solving an equation in which both members are unknown. And that is, in mathematics at least, an impossibility. In linguistics, which is, for better or worse, not mathematics (although some linguistic phenomena are amenable to mathematical investigation), an answer may be attempted; yet it will have to remain hypothetical, even though various degrees of possibility and probability may accrue to it. But it is the non-linguistic background knowledge, in so far as it is available, which is the part of the equation that must be solved *before* any degree of reality whatever can be claimed for the reconstructed proto-language; it is this knowledge which saves us from committing errors against which linguistic evidence alone cannot always shield us ²⁾.

²⁾ I might insert here that the cautions and skeptical views that I hold on reconstructed proto-languages, together with my insistence on clothing and articulating the bare bones of historical linguistics with the facts of life of linguistic history, are due in no small measure to the fortuitous but salutary circumstance that I am familiar with Romanic linguistics. The realization of how little we should know about the Latin language and its speakers if we had only

If we knew only Latin and Italian, for example, and nothing else about the other Romanic languages or the history of the Romania, there would be nothing inherent in the records of these two languages as we have them to dissuade us from regarding them as contemporaneous related dialects rather than as mother and daughter language³). It is obvious enough that the two represent different stages of development of a proto-language, since daughter languages of different degree of removal from a parent language may exist simultaneously – such as French and Italian, where French is more different from sister Italian than is Italian from the common parent Latin. Hence we are not actually constrained to conclude, on the basis of available synchronic linguistic evidence, that Italian is a later stage of Latin. If, therefore, a linguist ignorant of the significant non-linguistic circumstances viewed Latin and Italian as sister languages, he would rightly feel entitled to reconstruct from them a proto-language. Such a reconstruction would of course be an item without any counterpart whatever in the reality which we know with absolute certainty. But nothing inherent in this monstrosity will reveal it as such. Whatever the shape of this Proto-Latino-Italian, it will be a linguistic structure: this is assured by the process of its creation; that such a one never existed will not show in its physiognomy. Anyone who finds repugnant what may seem a preposterous argument concerning Latin and Italian, may instead use, say, two African dialects completely lacking further linguistic connections and totally unprovided with a non-linguistic historical setting, but being obviously related. How can the linguist know, on the evidence he has, what the relationship of these two dialects is, whether one is derived from the other or both come from a proto-language? It is indeed conceivable that the linguistic condition of the dialects is such that it yields no clue, not even to attempts of internal reconstruction, as to which is more archaic. Certainly greater morphemic complexity (such as indeed exhibited by Latin in the modern Romanic languages to recreate anterior linguistic and non-linguistic conditions, has more than once alerted me to the possible absurdity or, at least, lack of reality inherent in the designing of various proto-languages and conjectures about their speakers.

³) Whenever I refer to the genealogical tree, or use terminology connected with it (like parent language, daughter language, descent), I am implying all the necessary reservations and restrictions discussed in my article *Family tree, wave theory, and dialectology*, *Orbis* 2,67–72 (1953).

comparison with less-inflected Italian) is no safe criterion for either chronological anteriority or greater conservatism. It if were, German could not coexist with less-inflected English among the languages allegedly derived from a Proto-Germanic parent. No one would subscribe to the view that simplification of morphology is a typological 'improvement' of an evolutionary kind that is a function of time.

One might object that I am employing for my argument extreme cases such as do not occur in reality. I need not argue the point, first, because they *could* occur, second, because I am talking here of a principle, namely, our inability to establish the reality of a proto-language without the support of a concomitant non-linguistic reality.

Someone may suggest that the details do not really matter, that every reconstructed proto-language gives us some information on the real proto-language (provided one such existed, of course) because every reconstruction is in some measure realistic (see below), that therefore reconstructed formulae are not only legitimate but may also be authentic, and that at the very least the whole procedure affords valuable exercise to student and scholar. With all this I agree, and I do not reject efforts to devise some such formulae – as long as the circumstances afford nothing better.

I may perhaps insert in this connection a few words on the critique of my 1959 article by Hall⁴). He says that I "exaggerate" (203) when I claim that realism is not possible in reconstruction. He suggests that in Reconstructed Proto-Romanic "the degree of realism attained is very high" (204), and is willing to wager "that there did exist once upon a time, somewhere, a speech community in which [a reconstructed Proto-Indo-European utterance such as cited on page 205], or one closely similar, would have made sense". (206) To this I must say: (1) I did not speak of REALISM, which means, in this context, approximation, or probability, but of REALITY (see my title); (2) realism pertains to two parameters, (a) probability of existence of the reconstructed proto-language (b) approximation in form to the real proto-language; (3) probability of existence can be demonstrated only by extra-linguistic evidence: (a) if prehistoric or historic records completely negate the existence, then the question of approximation in form is meaningless; (b) if the records but weakly support the existence,

⁴) Robert A. Hall, Jr., On realism in reconstruction, *Language* 36,203-6 (1960).

then the realism of the reconstruction is correspondingly weak; (c) if the records are altogether neutral on existence, that is, if a putative proto-language cannot be provided with a location in a time and at a place, although linguistic facts point to the existence of a proto-language some time, somewhere (and this is the case with Proto-Indo-European), then the reconstruction becomes an admissible, albeit place-less and time-less construct, to be used and understood exclusively as the best available result of probing the past, and possessed of no proven reality and a minimal degree of realism; (4) in the absence of the reality along which to measure it, greater or lesser degree of realism of a reconstruction can not be demonstrated but only surmised.

When Hall ascribes, therefore, "realism" to a passage of reconstructed Proto-Romanic (204-5), he speaks of something unmeasurable, and he quite rightly, but inevitably vaguely, uses only such phrases as "reasonably high" (204), "reasonably realistic" (205), "closely similar" (with reference to Proto-Indo-European, 206). But he does not make it clear whether he means realism (probability) of existence or realism (approximation) of form. I dare say he does not distinguish, although he should: for if it were demonstrated in historical terms that this Reconstructed Proto-Romanic cannot be fitted into time and place, that is, cannot be provided with speakers, then all worries about realism of form become supererogatory. That "it would have been at least five-sixths comprehensible and perhaps even more so, to a Roman of ca. 50 B.C." (205) proves nothing concerning its reality; even as a measure of realism it does not carry far, since intelligibility is still possible even under conditions of extreme distortions of structure.

Hall's Proto-Romanic refers, I believe, to a koinē used all over the Roman state (it precedes his subsequent divisions into Proto-Italo-Western and Proto-Eastern Romanic) about the middle of the first century B.C. It is differently defined by Grimes and Agard, who believe that "the Romance languages are, roughly, the modern developments of what was the popular speech of central Italy about the time of the Pax Romana." ⁵⁾ I shall overlook the considerable disagreement of the Hall and Grimes-Agard dates and places for Proto-Romanic, and shall address myself to the probability of existence of such a language.

⁵⁾ Joseph E. Grimes-Frederick B. Agard, *Linguistic divergence in Romance*, *Language* 35, 598-604 (1959).

whether in the first century B.C. in the entire Romania (Hall), or in the first or second century after Christ in central Italy (Grimes - Agard).

By either of these dates Latin had spread, in the wake of Roman conquests by force of arms, to various regions of the future Romanic-speaking world, superseding in all of them non-Latin (Italic, non-Italic, and even, especially in central Italy, non-Indo-European) idioms. The Latinization of Italy itself had of course begun in pre-historic times, perhaps in the seventh century or earlier, and was still in process in Augustus' day. Other linguistically important dates for political and subsequent linguistic conquest are: Sicily, 241 B.C.; Sardinia, 238 B.C.; Cisalpine Gaul, 285-225, then lost to the Carthaginians, but definitively reconquered 198-191; Spain, begun early in the Second Punic War, completed in 205 B.C.; Liguria, 191-172; Southern Gaul, 125-121; Central and Northern Gaul, 58-51. Although we unfortunately do not know precisely with what horizontal (geographic) and vertical (social) speed and thoroughness the various native monolingualisms were replaced, no doubt through a period of bilingualism, by Latin monolingualism, it is scarcely convincing to propose, as do Grimes and Agard, that even the beginnings of Latinization in some regions of the Romania were delayed by as much as three centuries counting from the date of political conquest, or that a new Latinization (canceling the previous one?) took place with the spread of the speech of central Italy during the Pax Romana. (Note that the Pax Romana does not furnish a very precise date since it lasted from the time of Augustus through the Flavian dynasty, at least, that is, nearly a century and a half.) The fact is that we have from all these places Latin inscriptions that much precede the Pax Romana. That they do not exactly show us the kind of language that was actually spoken is a generic shortcoming of all written documents, especially those of the Roman state, of which I have spoken in detail elsewhere ⁶⁾. And it seems to me venturesome to insist, as does Hall, that there was a single Proto-Romanic spoken throughout the provinces, regardless of the date of conquest, despite the different kinds of Latin imported into the various areas at so greatly divergent dates, and despite the variety of substrata superseded by Latin. The view of such a wide-

⁶⁾ Ernst Pulgram, Spoken and written Latin, *Language* 26,458-66 (1950).

spread and long-lasting koinē (which is not the same thing as a lingua franca, or a Pidgin Latin, such as may have existed, but certainly not as everybody's speech everywhere, and not as the parent of the Romanic dialects), may be based precisely on the biased written records which I just mentioned. And if there was no real koinē, Hall's Reconstructed Proto-Romanic, for the time and place to which it is assigned, is a fiction, even though the "text is, as a whole, quite close to what we know about popular Latin from other sources . . ." (205) I should not regard it as a fiction of but minimal degree of realism if it were, not "as a whole" but in detail, not "quite close to" but identical with, not "what we know about popular Latin" but *the* popular Latin, uniform in the entire future Romania. But such a language we do not know, because it did not exist. What we need is not "more confidence" ⁷⁾ in what we are accomplishing by means of reconstructive techniques, but more evidence for its authenticity, and, lacking that, more reticence.

The method of reconstruction of a proto-language is the comparison of related dialects. The linguist decides what dialects he may compare by measuring the degree of relationship, although he is of course greatly and often decisively aided in the task of selection by extralinguistic historical facts that he knows and which prevent him from comparing, for example, English and Russian (except on the level of Proto-Germanic and Proto-Slavic). But there exist in fact no truly quantitative, scientific criteria for measuring relationship and for furnishing a clue to comparability, indeed a permission to compare, which will allow us to disregard intuitive, impressionistic judgments and non-linguistic criteria. I am aware of some recent attempts (not without precursors) to remedy this shortcoming, by Ellegård, Kroeber, and Grimes and Agard ⁸⁾. But I am not persuaded of the validity of any of the procedures. Ellegård and Kroeber base their taxonomy upon vocabulary, which to my mind is the coarsest of the ingredients of a language, and the least suitable, indeed the most deceptive, of all criteria for establishing relationship. (Certainly on the evidence of lexical composition English and French seem more closely related

⁷⁾ Hall 1960, 206.

⁸⁾ Alvar Ellegård, Statistical measurement of linguistic relationship, *Language* 35.131-56 (1959); A. L. Kroeber, Statistics, Indo-European, and taxonomy, *Language* 36.1-21 (1960); Grimes-Agard, 1959.

than they actually are structurally.) Furthermore, the taxonomy of proto-languages delivers, in consideration of what I have said about them, at best a brittle and at worst an illusory result. Ellegård recognizes this: "... if statistical methods are to be used, linguistic relationship has to be split up into at least two different concepts, one of which may be termed similarity, the other interdependence of development. Both can be explicated in terms of statistical correlation coefficients. But while the coefficient of similarity demands only a knowledge of the languages to be compared, the coefficient of interdependence presupposes also a knowledge of the original [proto-] language -- a knowledge which is bound to be hypothetical in most practical cases -- and of the rate of development of the daughter languages. Now if we already know the history of the languages (i.e. the extra-linguistic history, the historic data), it will obviously be easiest and most satisfactory to describe their relationship directly on that basis (which is not statistical). Hence the measure of interdependence of development will have only a limited usefulness. In the present article I have therefore chiefly dealt with the measure of similarity." ⁹⁾ Hence the measurement of similarity (especially if based on the lexicon) does in fact not give us the measurement of interdependence of development which alone would tell us what languages may be profitably compared, in what sequence, and on what level. This realization prompts Kroeber, after having attempted a taxonomy of nine Indo-European proto-languages, to say: "It must ... be remembered that the languages are really reconstructed proto-languages, from nearly two thousand to three thousand years old. It is thus evident that if French or Italian had been tested instead of [Proto-] Italic, English or German instead of [Proto-] Germanic, and Bengali or Hindi instead of Sanskrit, the differentiating would have been even more severe." ¹⁰⁾ Consequently, Kroeber concludes, "It does begin to look as if time were beginning to justify Boas in his contention that all but the more obvious reconstructive and historic classification of unwritten languages contained elements of unreliability." ¹¹⁾ Not only unreliability but also, I should add, in some circumstances full unreality.

⁹⁾ Ellegård, 1959, 154.

¹⁰⁾ Kroeber 1960, 15.

¹¹⁾ Kroeber 1960, 19.

Grimes and Agard deal with the real Romanic dialects and seek to establish statistically what Ellegård calls the relationship of similarity. Whether their choice of criteria, the operations, and the resultant statistics are valid is not germane to the present discussion. But where the authors' endeavors concern themselves with questions on relationship of interdependence of development, that is, with the Proto-Romanic parent language, the results must be, in Ellegård's words, of limited usefulness. Since we know the external history of the Romanic languages as a matter of extensive and trustworthy historic records, it will (to speak again with Ellegård) "obviously be easiest and most satisfactory to describe their relationship directly on that basis." ¹²⁾ But this Grimes and Agard unfortunately do not do, or rather, do wrongly, when they derive the modern Romanic languages from the popular speech of central Italy in the first or second century after Christ, as I noted before ¹³⁾. Hence their statistical procedure is not suited to the creation of quantitative criteria that will determine which ones of a number of known languages must be used, in what combination and in what sequence (more of this below), in order to arrive at a proto-language of maximum degree of realism, though never complete reality.

If it were ever demonstrated, as some think it may well be, that Indo-European and Semitic are akin in such a way as to warrant the reconstruction of a Proto-Indo-Semitic language, then French and Arabic may be called, however remotely, related languages. Given the condition, then, that a linguist of the distant future, through some strange but imaginable circumstances, finds records of French and Arabic but of no other Indo-European or Semitic language, his undertaking a linguistic comparison of French and Arabic with the prospect of reconstructing a parent speech would not be nonsensical – even though no right-thinking linguist of our day would, or should, attempt it. He should not and does not because French and Arabic, apart from not looking very closely related, do not need to be compared: for the comparison of Indo-European and Semitic we shall do better, if the time comes, to employ older stages in either family, which luckily are accessible. But perhaps the most weighty consideration available to us against a French-Arabic comparison is our extra-linguistic

¹²⁾ Ellegård 1959, 154.

¹³⁾ See above, p. 23–24.

knowledge, which the future linguist of whom I spoke may lack just as we are lacking comparable information on early Indo-European and Proto-Indo-European historical circumstances, the knowledge, that is, that the non-linguistic conditions and the history of the speakers of Arabic and French are such as to divest a Proto-French-Arabic language of any degree of realism and plausibility, let alone reality.

Linguistic reconstruction, then, has no built-in control mechanism; its result betrays its own falsehood no more than it confirms its correctness. The process of reconstructing a proto-language by the methods of comparative linguistics is a mechanical one in which a series of linguistic equations is solved for the purpose of finding the common denominator, the proto-language, which will satisfy all the requirements of historic developments that produced the daughter-languages. In this method only the *procedure*, but not the *selection of the equations* with which one operates, is rigorously scientific. Even a technician who is conversant with the method but knows little or nothing of the nature and the hidden problems of his material, may successfully solve the equations, without being able to judge the validity of his premises or the merit of his conclusions. (It is of course true that a science has attained a degree of maturity if its every-day chores, be they, in linguistics, descriptive analysis or historical comparison, can be accomplished by technicians. But this does not mean that the raising of technicians is a sufficient educational goal: they by themselves will further linguistics no more than garage mechanics advance automotive engineering.) The control, however, and the critical view and the corrections are indeed provided – not by mathematics, but by the person who possesses both a sufficient knowledge of the extra-linguistic circumstances and a sufficiently educated judgment (though it can never be infallible since it deals with non-quantified items) as to which two or more languages are so constituted as to permit and invite comparison. In other words, what prevents the good historical linguist from committing grievous errors and from producing altogether fantastic proto-languages is, apart from his knowledge of extra-linguistic facts wherever they are available, his good training and his good sense, and not (for the time being, I fear) a strict, quantitative set of axioms or criteria, a really scientific, that is, self-correcting, methodology.

The Indo-Europeanist who must make a decision on the compa-

rability or authenticity of unattested proto-languages of the Indo-European family is, we must admit, in some measure and in some manner, no matter how learned and skilled he is, to be likened to the under-informed linguist of the future whom I mentioned, at least as regards his disabilities and shortcomings. Even though (let us assume for the moment) his judgment for the comparability of languages is trustworthy, and even though he has both linguistic and non-linguistic records of good quality and quantity for a fair number of centuries backward from the present, he still comes to certain points on his way toward Proto-Indo-European where he is inevitably faced with the question as to whether there was or was not a Real Proto-Balto-Slavic, a Real Proto-Italic, a Real Proto-Italo-Keltic corresponding to the customary entries on the genealogical tree.

Rather than deal with remote languages let us once more turn to a reconstructed or postulated language where the facts are better known, for example, Proto-Ibero-Romanic – a tongue which, though it is no more attested than Proto-Balto-Slavic, has the advantage of being provided with fully known mother and daughter languages, Latin and the Spanish-Portuguese dialects, respectively. A Proto-Ibero-Romanic language is conjunctured by some linguists on the familiar view that the dialects of the Iberian peninsula, which are more closely related to one another than is each to any other Romanic dialect (except Catalan, which seems nearer Provençal), must be derived from a common proto-language. This opinion seems unobjectionable enough at first sight, and the Proto-Ibero-Romanic language appears therefore justified by the methods of historical linguistics. But what does linguistic history, which is amply available in this instance, say about it?

In accordance with all we know about the early history of Spain and its conquest by Roman arms, and in the light of all we know about the behavior of languages and their spread and diversification, the Proto-Ibero-Romanic hypothetical language most likely did not exist; certainly it did not have to exist. Instead it seems highly probable and reasonable, albeit still unprovable in rigorous terms for want of complete linguistic evidence, not that Latin was imported in the Iberian peninsula and there existed for a while in a modified form called Proto-Ibero-Romanic before it was somehow decomposed into various dialects, but that Latin (and very likely not just one kind of local or social dialect of Latin, nor one Latin of a single period) was super-

imposed upon, and exposed to the substratic influence of, a variety of already existing dialects (perhaps of an 'Iberian' type – though the name has little more than geographic meaning unless Basque is a modern derivative of it). If that is so, then the multiform dialects of the area are not at all descendents or continuations of some kind of uniform Proto-Ibero-Romanic, they are not products of the disintegration of a single proto-language spoken in the Iberian peninsula. That is to say, there never were any people to whom Proto-Ibero-Romanic was a native language. As far as we know, the linguistic diversity of Iberian Romanic is due to both the diversity of the linguistic substrata (at the state of communications preceding the Roman conquest we must not assume the prevalence of a unified language all over the Iberian peninsula), and, perhaps less so, the diversity of the Latin imported (although we unfortunately know very little about Latin dialects). On the other hand, the high degree of relatedness of the dialects of Spain and Portugal in comparison with other Romanic dialects is due to, and testifies to, the relatedness of the substrata as compared with those prevailing in Gaul (Kel'ic), Italy (mostly Italic), Rumania (Thracian?), and so forth, and of course to the relatedness of the Latin speeches that superseded the Iberian stratum ¹⁴).

The term Proto-Ibero-Romanic, then, has a real content only in the sense that it is Latin itself. Latin is the only proto-language which the inevitable non-linguistic considerations allow me to postulate, a Latin, to be sure, that must be sought outside the Iberian peninsula – for as soon as it became the language of the peninsula's non-native speakers it acquired a variety of dialectal traits. I do not deny, then, that the visible relationship of the dialects of the Iberian peninsula has its source in a proto-language, but I do maintain that it is not of necessity, indeed with scarcely any probability, this so-called Proto-Ibero-Romanic, located on the peninsula and spoken after the Roman conquest, such as it appears in various genealogical trees and such as it is understood by most scholars.

¹⁴) I do not wish to give the impression that I regard the substratum as the only source of linguistic diversification of an original more or less unified dialect; many linguistic and non-linguistic (political, geographic, demographic, economic) factors play important roles, and are often difficult to disentangle from one another. But I do consider the substratum a very important force among all these.

Since we have no truly quantifying criteria that measure the relationship among languages, the constitution of subgroups of more closely related idioms for the purpose of reconstructing intermediate, partial proto-languages, such as Proto-Ibero-Romanic or Proto-Italo-Keltic, derives from value judgments, from educated guesses, on the part of the investigator. But here it must be noted that divergent decisions on the first level of comparisons, that of known dialects, must lead to divergent results on the final level, that of the unknown proto-language. I may illustrate this by the following example dealing with an imaginary set of related languages, *a, b, c, d, e, f*, declared to belong to one linguistic family. Let us assume that we know only these, the present generation, so to speak, of this family, such as is indeed often the case with Amerindian or African dialects. In attempting to arrive

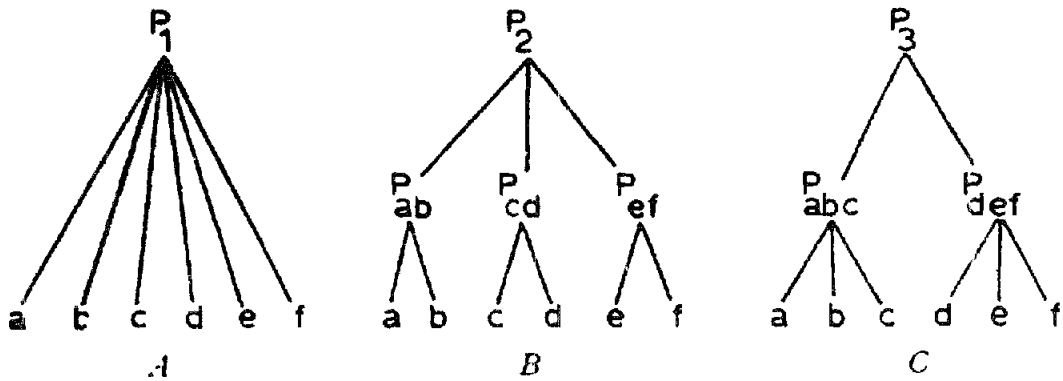


Fig. 1

at a proto-language the linguist may choose to follow one of several possible paths toward the solution of this series of linguistic equations, as shown in Fig. 1. Under *A*, he compares all dialects at the same time and derives the proto-language P_1 directly. Under *B* and *C* he connects certain dialects because they seem more closely related and appear to form a sub-family. Also other designs are of course feasible. But owing to the absence of quantitative criteria on degrees of relationship, and of historic information on either the languages or the speakers, not all linguists need necessarily agree on the grouping of dialects for the purpose of reconstruction, and even disagreement from the majority view may mark neither the fool nor the maverick. Be it also noted that any grouping preceding the comparison implies or anticipates a certain historic relationship which, one should imagine, ought to be the result rather than the premise of reconstruction. In any event, the reconstructed proto-languages, P_1 , P_2 , P_3 , can hardly come out

identical because they were fabricated by different procedures, while of course the real proto-language, *P*, if ever there was one in the first place, is unchangeably fixed. Not only that, but also the intermediate proto-languages under *B* and *C* (P_{ab} , P_{cd} , P_{ef} , P_{abc} , P_{def}), which are not the same, and of which there is none under *A*, may or may not have existed. Whether schema *A* or *B* or *C* is representative of a reality of linguistic and non-linguistic history can therefore not be argued on the strength of precarious and debatable choices pertaining to the arrangement of dialects ¹⁵).

In some linguistic families, notably Amerindian and African, pre-history is but a few decades distant. Any thrust into the past will involve the linguist in reconstruction. He will therefore often find himself in a position such as exemplified by Fig. 1 above. If what I have said so far in this paper is true, it follows that the pyramiding of reconstructed series of proto-languages which are meant to probe into the past, cannot but suffer from rapidly and radically dwindling realism on both counts, authenticity of form and reality of existence. By the time the Amerindian or African linguist has reached, speaking in terms of the genealogical tree (which he is likely to construct in order to illuminate chronology and family relationship – and, one may hope, no more), the third or fourth generation, which perhaps carries him backward no farther than a century, he faces a proto-language of his own making that has an exceedingly small degree of verisimilitude. And if furthermore he wishes to assign dates to these proto-languages by means of linguistic calculations, glottochronologic or other, the realization becomes inescapable that such a chronology, based largely (as it is) on fictional material, is highly questionable if not worthless.

Concerning the extra-linguistic deductions often derived from such

¹⁵ I have noted (see above p. 28) that the reconstructing comparativist is solving a number of linguistic equations. But when I speak of equations in this context, it should be borne in mind that they differ significantly from mathematical ones. In mathematics, given a number of unknowns and a number of equations, it does not matter in what sequence the unknowns are gradually eliminated by the process of substitution: the result is always the same. If it is not, an error was committed somewhere. In linguistics, however, where neither the knowns nor the unknowns can be provided with a quantitative value, more than one solution is possible, and it is the progress toward it, that is, the sequence in which the equations are solved, which in considerable measure determines the final result.

a genealogical tree, I need scarcely say more than has already been said in passing. Having societies or persons migrate from one place to another, along roads that are represented, somewhat flimsily and recklessly, by the branches of the genealogical tree, which becomes a geographical map, in order to accommodate displacement and diversification of languages, all that is bad enough even if the tree is a map of known territory. But it is preposterous to pretend that one may pursue the progress of imaginary societies along the branches of a hypothetical tree of dubious reconstructed languages, and to provide with ethnic names various vague tribes and nations which go with the more than doubtful linguistic items. The linguist who thus has perilously nurtured a genealogical tree and topped it with a shaky proto-language, now triumphantly descends the same puny growth, at the head, as it were, of bands of lusty migrants and invaders who disperse their spectral selves and their mute idioms, leaving behind them, at fleeting halts and along conjectured paths, a faint and untidy trail of arrowheads or pots or battleaxes or cephalix indices, or perhaps of none of these – but always of alleged proto-languages. This is not linguistics, and it is not prehistory or history ¹⁶).

¹⁶) A portion of my book *The tongues of Italy: prehistory and history* (Cambridge, Mass. 1958) attempts to demonstrate that in determining the linguistic aspect of the prehistory of a certain region it is necessary to use not linguistic arguments alone, nor those pertaining to any other single branch of science alone; and I tried to show, by word and example, that prehistory, a branch of human knowledge which by its very nature deals with the recondite, must be based upon all the evidence and all the methods accessible to the modern scholar, whether they come from linguistics, ethnology, antropology, archacology, history, mythology, geology, or any other available source. (Significantly, two reviewers of my book, both linguists, were unable by inclination or education to grasp my intent, and they lamented a circumstance that they should have welcomed, namely, the vast amount of non-linguistical material which I adduced in my argument.) I cannot recommend warmly enough to anyone really interested in this question, Luigi Pareti, *La comparazione scientifica e la preistoria*, in: *Studi minori di storia antica*, I: *Preistoria e storia antica*, 1–22 (Rome 1958). Anyone who has but cursorily informed himself on the problems and principles involved in prehistoric research cannot help perceiving the amateurishness and frivolity inherent in the ventures of some linguists to paint a picture of prehistory with purely linguistic means – with the aid of reconstructed proto-languages, if you please!

Cf. Rouse, in Raymond H. Thompson, ed., *Migrations in New World cultural history* 67 (Tucson 1958): "One of the values of the present volume, in my

If I hold such views on proto-languages in general, especially the intermediate proto-languages, what do I do with the queen of all proto-languages, Proto-Indo-European? If I deny an *a priori* existence to Proto-Ibero-Romanic, should I not deny it also to Proto-Indo-European, that is, should I not in effect be saying that Proto-Indo-European may not have existed at all, especially since I have argued that it is not really a 'beginning' any more than is Proto-Ibero-Romanic, but part of a larger genealogical tree of which we do not know the rest? ¹⁷⁾ But there is a difference. In Romanic we are entitled, if we so choose, to go beyond the proto-language, that is, Latin, and penetrate to older stages of the family; in Indo-European we cannot, at least not at the present state of our knowledge, go beyond Proto-Indo-European. Consequently Proto-Indo-European is, in terms of currently practicable historical linguistics, though surely not in terms of real linguistic history, a terminal item; it therefore belongs into a different category from all those other proto-languages, real or reconstructed, that lead up to it.

It follows that, while the existence of a real parent language for the Indo-European dialects can be questioned as little as that of a real parent language for the Ibero-Romanic dialects, it may not be something that we shall, if we ever obtain further information, continue to call Proto-Indo-European, certainly not if the term connotes, as it does to some, a number of geographic, ethnic, and indeed linguistic corollaries. For such corollaries, pertaining to realia, cannot be deduced from the reconstructed language in any event. If we should ever discover the shape of the larger genealogical tree of which 'Proto-Indo-European' is a branch, then we may also find it useful, or indeed obligatory, to revise the name: Proto-Indo-European may turn out to be language X of the larger tree just as Proto-Ibero-Romanic turned out to be Latin. I should of course require, in accordance with what I have been saying, that the case of the genetic connection of the

opinion, is that it illustrates the necessity of combining archaeological, ethnological, physical anthropological, and linguistic evidence in drawing inferences of migration'. To this adds the reviewer (Newman, *Language* 35,717 [1959]): "None of the symposium contributors, however, attempted to utilize evidence outside his own specialty".

¹⁷⁾ See my earlier article, *Language* 35,425 (1959). See also above, p. 23, where I put Proto-Indo-European under (3c), making it an admissible though unreal construct.

Indo-European with another language family not only be stated on purely linguistic grounds, but also be accompanied and supported by such non-linguistic evidence as would make the physical contact of the two language families, their geographic contiguity if not congruence at some time, a fact or at least a very strong probability. But if that is accomplished we shall also be in a much better position than we are now to place 'Proto-Indo-European', or whatever we shall prefer to call it, in time and place, and to associate it with a culture and with speakers – just as, thanks to the historical information we do possess, we could find a home, an *Urheimat*, if you will, and a culture and speakers to go with 'Proto-Ibero-Romanic', that is, Latin.

Indeed even now we do in fact abrogate the existence of Proto-Indo-European as *the* proto-language of all dead and living Indo-European dialects, if we follow the reasoning of those who see in Hittite not a daughter but a sister language of Proto-Indo-European and who postulate a Proto-Indo-Hittite parent language for both¹⁸). It is true that we still need a proto-language for all the many dialects that must be related by descent; but it is also true that Proto-Indo-Hittite, if we postulate a reality for it since it is now at the practically accessible apex of the pyramid (or, if one prefers, at the base of the genealogical tree), detracts from the probability of the real existence of Proto-Indo-European: for Proto-Indo-European is now, like Proto-Ibero-Romanic, an unattested intermediate proto-language whose reconstructed formulae do not, in the absence of non-linguistic corroboration, imply or warrant the presumption of its reality. We must not think or pretend, therefore, that the acceptance of the Proto-Indo-Hittite hypothesis, or perhaps of an even more conjectural Proto-Indo-Semitic, or of any other farther removed proto-language, betokens merely a rearrangement of the genealogical tree without affecting the nature and the reality of what we have been calling Proto-Indo-European. For as soon as we have relegated Reconstructed Proto-Indo-European to any place other than the ultimately accessible apex of the pyramid of family reconstruction, have placed it in any but

¹⁸) Ellegård 1959, 146, seems to come to this conclusion statistically. But I am reluctant to accept a calculation based on vocabulary alone. I have more confidence in views obtained without statistics and mathematics as long as they come from a first rate historical linguist who *knows* Hittite, such as Sturtevant. (Perhaps Ellegård does too, but he makes his case on statistic grounds.)

the terminal position, as some day perhaps we shall have to do, we shall have put it on a spot where, apart from its congenital lack of authenticity in form and substance, the very existence of a Real Proto-Indo-European in the customary sense of the word is no longer implied.

What I have said in this article concerning a few of the unattested reconstructed languages of Indo-European applies, *mutatis mutandis*, to all of them, as it does of course to those of other linguistic families. That is to say, as regards Proto-Italo-Keltic, Proto-Italic, Proto-Gallo-Romanic, Proto-West-Germanic, and many more, with or without proto- in their names, no matter how convenient and even indispensable they may be as items in the procedure of historical linguistics, the linguist must in each case test and support their reality by the non-linguistic facts of linguistic history.

Postscript. When I had finished this article in its first rough draft, I received a letter from a student of mine in Romanic Linguistics, Mr. Herbert Izzo, then in Vietnam. In it, discussing mainly possible subjects for a doctoral dissertation, including the reconstruction of Proto-Romanic, Mr Izzo states an opinion concerning reconstruction and historical linguistics and linguistic history that can serve as a lucidly worded resumé to a portion of my own article, and that contains arguments and conclusions gratifyingly close to mine. I shall now cite, with the author's permission, the pertinent passage so as to give Mr Izzo due credit, and also to allow our views to support one another by the very fact of their convergence.

"...I have doubts about the validity of several aspects of reconstruction. One question is, if we were to compare two stages of a single language while supposing them to be different dialects, wouldn't we be very likely to set up a proto-language for them instead of deriving the later from the earlier form? And, although I believe that the Romance languages derive ultimately from a nearly uniform base and that all the Indo-European languages do likewise, I cannot believe in the reality of reconstructions (better, constructions) like Proto-Germanic, where there is (so I believe) no historical evidence but only linguistic similarity to show that a single proto-language ever existed. Still less can I believe in the reality of such languages as Proto-Gallo Romance, Proto-Ibero-Romance, etc., where what is known about the area in question makes it seem (to me at least) *most* unlikely that

there could have been any 'period of common development' before differentiation. Of course, I am more or less echoing your remarks in 'Linguistic expansion and diversification' [in *Studies ... What-mough* 239-52 (The Hague 1957)], but the contrary is implied in [Robert A.] Hall's 'Reconstruction of Proto-Romance' [*Language* 26,6-27 (1950)] and his papers on the Oaths of Strasbourg [*Language* 29,317-21 (1953), 35,24-5 (1959), the second in reply to remarks by John A. Rea, *Language* 34,367-9 (1958)]. Now I think perhaps reconstruction may be compared to the solution of a series of simultaneous equations in that one might compare *any* two dialects and set up a proto-dialect for them, then compare this proto-dialect with any other similarly deduced proto-dialect *or* with any third dialect, and so on until a final single reconstruction (the 'solution' of all the 'equations') had been made. Then the final reconstruction approximates an actual stage of the language, but the intermediate 'reconstructions' are merely convenient fictions corresponding to no actual linguistic stages except by coincidence.

But if all this is true, it is impossible to know without extra-linguistic information whether a reconstruction represents a relatively real or a completely fictitious proto-language. (For example, the existence of Proto-Hispano-Portuguese seems somewhat more plausible than the existence of Proto-Hispano-Roumanian.)

But if it is not true, i.e., if related languages cannot be compared at random, but only those which have an immediate common ancestor can be compared and only relatively contemporaneous stages can be compared with each other, then we are even worse off; for we need still more extra-linguistic information to proceed. . . ."

University of Michigan

ERNST PULGRAM