Indo-European and the Indo-Europeans

W
DE
G

Trends in Linguistics

Studies and Monographs 80

Editor
Werner Winter

Indo-European and the Indo-Europeans

A Reconstruction and Historical Analysis of a Proto-Language and a Proto-Culture

Part I
The Text

by
Thomas V. Gamkrelidze
Vjačeslav V. Ivanov

with a Preface by
Roman Jakobson†

English version by
Johanna Nichols

Edited by
Werner Winter

To the memory of our teachers:

Mixail Nikolaevič Peterson (1885-1962)
Georgi Axvlediani (1887-1973)
Georgi Cereteli (1904-1973)

Authors' Preface

The second half of the twentieth century has been marked, in the history of linguistics, by a growth of interest in problems of diachronic linguistics, motivated by the general evolution of linguistic thought in recent decades. Overcoming the Saussurean antinomy of diachrony and synchrony, linguistic science is moving toward a theory with greater explanatory power than purely taxonomic synchronic grammar offers.

The growth of interest in diachronic linguistics has fostered a return to questions that arose in classical Indo-European comparative-historical linguistics, questions that can be posed more clearly now with the aid of new methods of linguistic description developed by various trends in synchronic linguistics and by linguistic typology. Typology is particularly important to contemporary linguistics because it makes it possible to reveal the universal linguistic categories that characterize the deep structures of language, and also to determine the degree of diversification between various language systems. Furthermore, language is a social phenomenon and a part of human culture and therefore closely connected to other aspects of culture. Therefore, both synchronically and diachronically language must be studied together with the other aspects of culture that make up the subject matter of modern cultural anthropology.

This book presents the results of our joint comparative research into the Indo-European languages and the reconstruction of Proto-Indo-European that gave rise to the attested Indo-European languages. Indo-European is studied in this approach in typological comparison to other languages, in particular the geographically adjacent ones with which Proto-Indo-European must have interacted for a long period of time.

The first part of the book presents the results of linguistic analysis — phonological, morphological, syntactic, and areal-dialectological — of Proto-Indo-European. This does not mean that the analysis should be viewed as a systematic survey of the various branches of comparative Indo-European grammar, as is done in the standard handbooks. Rather, the first part is a study of key questions of Proto-Indo-European structure, involving a wide range of facts and yielding a relatively complete picture of this language in its dynamic development and its typological links to other language systems.

The second part gives a relatively full investigation of the Proto-Indo-European lexicon, presented by semantic groups, as well as fragments of Indo-European culture that can be reconstructed from the lexicon; it also describes the culture-historical links of the Indo-European lexicon to a number of languages of ancient Eurasia. This is properly a dictionary of Proto-Indo-

European lexemes presented not in alphabetical order (that can be found in the indexes) but in order of semantic groupings. In contrast to the well-known Indo-European dictionary of Pokorny, in our semantic dictionary each entry presents not only the formal correspondences between cognates which make possible the reconstruction of a protoform, but also the phenomena of material and intellectual culture that are connected with the root in the individual daughter traditions. On this basis a reconstruction is then given for the Proto-Indo-European level.

The final section presents the results of research into the linguistic and culture-historical data relevant to determining the Indo-European homeland and the migratory routes taken by the Indo-European tribes across the Eurasian continent to their historically attested locations.

The range of questions surveyed here should be of interest not only to linguists but also to historians, archeologists, anthropologists, and historians of culture. Given the specialized nature of some parts of the book, especially its linguistic parts, readers may wish to read chapters in various orders depending on their interests. The second part of the book, where culture-historical problems are analyzed from a linguistic perspective, and the final section on migrations can be read without reading the first part (except where explicit cross-references are made). The final section, on migrations, is essentially self-standing and can be read without the others, although the semantic dictionary will be useful for more detailed understanding of the histories of the individual words that support the historical arguments in the final section.

Linguists, on the other hand, may wish to read only the first part and skip the factual details of the second part. Still, the two parts are organically linked, as will be evident, and this is why they are covered by joint indexes. These can be used as a guide or word index by readers interested in particular questions discussed in the book.

The book is the result of joint research begun in 1970. It was produced not by assembling separate chapters written individually by one or the other author, but jointly, by laying out together the conclusions and results of many years of collaborative research that involved joint analysis of particular problems and joint formulations of results.

Throughout the time when the research was being done and the book written, publications have appeared that have been consistent with our claims. Insofar as possible we have taken these into account in the text and bibliography, in the conviction that the sheer quantity of agreement is the strongest confirmation of our analysis. The most recent literature, especially works that appeared in 1983, could generally be taken into consideration only in the Afterword, which surveys some works that directly address or respond to our claims (see also the section entitled 'Addenda and corrigenda' in the Russian original, pp. 1317ff.).

We are pleased to express our gratitude to those who have been involved in

one way or another in discussing and responding to earlier presentations of this work in seminars and reports. We are grateful, first of all, to our teachers, the late G. S. Axvlediani and G. V. Cereteli, for their constant interest in our work in its early stages, and also to V. I. Abaev, A. A. Zaliznjak, V. N. Toporov, T. E. Gudava, I. M. Diakonoff, I. M. Steblin-Kamenskij, S. D. Kacnel'son, and G. V. Stepanov. Among foreign scholars we thank J. H. Greenberg, M. Mayrhofer, O. Szemerényi, C. Watkins, J. Catford, W. P. Lehmann, E. Polomé, E. Hamp, H. Pilch, W. Winter, E. Risch, R. Schmitt-Brandt, A. H. Kuipers, H. Hoenigswald, A. Kammenhuber, A. Morpurgo Davies, M. Gimbutas, W. Cowgill, H. Birnbaum, R. Anttila, R. Austerlitz, K. Strunk, R. Schmitt, K. H. Schmidt, H. Aronson, J. Greppin, J. Nichols.

We should make special note of the extent to which we are indebted to Roman Jakobson — one of the greatest scholars of our times, one of the founders of contemporary linguistics, and the source of many of the ideas that this book is based on. During our work on the book we discussed with him our results and various problems that arose, and this greatly facilitated our analysis and exposition. His constant participation in our work is reflected in his foreword to this book, written when we had completed the manuscript. This foreword is one of his last pieces of writing.

Thomas V. Gamkrelidze
Vjacheslav V. Ivanov

Tbilisi — Cavkisi — Peredelkino — Moscow
1970-1983

Translator's Preface

This book is the first major handbook of Indo-European to be written since the discovery and analysis of Hittite, the first ever with explicit and consistent theoretical grounding, the first whole-scale Indo-European reconstruction in which typology has played a major role, the first attempt to join orthodox comparative reconstruction of Indo-European with an account of the structural and lexical resemblances in other Eurasian and Near Eastern languages, and the first reconstruction of an Indo-European homeland based on all available kinds of linguistic data. In its semantic dictionary and indexes it provides the first Indo-European lexicon of both forms and meanings. The method used is standard rigorous comparative-historical analysis, but the substance of the linguistic reconstruction and the reconstructed homeland are novel.

The Russian original is a well-written, stylistically refined exemplar of an expository tradition that has no analog in contemporary western prose and therefore cannot be captured in English. In the canon in which it is written, the expository strategy and the logical argument proceed from general to particular, seeking to ground the particular in the general. Generalizations are often implicitly treated as premises (rather than as conclusions or hypotheses) and particular facts are shown to follow from them. Hence the expository strategy may be said to emphasize deduction rather than argumentation. For instance, a strategy frequently used in reconstruction is first to show that, on logical and structural-typological grounds, one would expect such and such a structure, property, or phoneme in Proto-Indo-European, and then to show that there exist in the daughter languages forms that can be explained by tracing them back to the expected structure. Thus in I.5.3.2–I.5.3.3 it is shown that the structural typology of Proto-Indo-European is such that one expects to find alienable/inalienable possession and inclusive/exclusive pronouns; then daughter forms are presented that are consistent with a protolanguage that had those oppositions. In this mode of argumentation, a first priority is the structural and typological consistency of the reconstruction, and any conforming cognate evidence in the daughter languages that can be derived from such a reconstruction supports and confirms it. The commoner mode of argumentation in western historical linguistics in recent decades proceeds inductively, arguing that the daughter reflexes demand such-and-such a reconstruction (rather than that they derive from and confirm it), and much less priority is given to typological consistency of reconstructions, which, if brought up at all, would probably have the status of secondary observation on a reconstruction (rather than, as here, an essential logical priority).

Another example involves the reconstruction of the traditional plain voiced stop series of Indo-European as ejectives. In the canon followed here, typological implicational hierarchies and structural patterning in the reconstructed stop inventory — the near-absence of traditional *b, here *p'*; the relative lexical frequencies of the various stop series — are sufficient to reject the received reconstruction. To most American Indo-Europeanists, in contrast, the structural asymmetries of the protolanguage are merely interesting, and the only evidence sufficient for rejecting the received reconstruction would be a demonstration that the daughter reflexes in and of themselves demand a different reconstruction. But the daughter reflexes in themselves do not demand a different reconstruction; most of them are voiced and none are ejective except in one branch, Armenian, and even there glottalization is dialectal. The demonstration given in this book does not focus on the daughter reflexes and the reconstructed phonetics they demand, but rather adduces a great deal of information about structural properties of the comparative Proto-Indo-European reconstruction, the workings of Grassmann's Law as reconstructed both internally and comparatively, and the derivability of the daughter consonantal systems from a proto-system with an ejective Series I. Readers should be aware that the two stances on Proto-Indo-European ejectives are not a matter of individual differences of opinion or debate on phonetics and phonetic change, but rather can be thought of as a minimal pair indicating the different status given to premises and implications, or general and particular, in two different intellectual canons.

The choice of the deductive canon is a felicitous one for the task at hand. The Proto-Indo-European homeland reconstructed here is located at the very periphery or even outside of the present and historically attested ranges of known Indo-European languages (and in fact probably *all* Indo-European homeland reconstructions enjoying any currency among linguists are peripheral or external to the historical Indo-European speech territory). The structural features reconstructed here for Proto-Indo-European include some, notably the phonological system, morphophonemic canon, and word order, that are known to be strongly susceptible to areal influence; and others, such as inclusive/exclusive pronoun oppositions, alienable/inalienable possession, and stative/active verb categorization, that have a broad areal or geographical basis to their distribution. These features are predictably absent from the modern Indo-European languages because of their geographical distribution. If the daughter languages lack ejective stops, Hittite-style word order, etc., it is not necessarily because their ancestors never had them; the geography of their modern distribution — Europe, parts of Southwest Asia, the Indian subcontinent, all areas in which these features are lacking — is sufficient to predict their absence. Therefore it is probably safe to claim that, in principle, for *any* language family whose prehistory is known to have involved extensive migration, the reconstruction of the ancestral grammar should rely more

heavily on implicational hierarchies and other structural arguments than on comparison of the phonetic and grammatical substance of the daughter languages. That is what has been done in this work wherever phonological or grammatical structure is involved.

In the intellectual canon exemplified here, the scientific text itself is not so much a communicative contract between writer and reader, where the writer monitors what the reader is expected to know at a given point in the text, but rather a gnomic statement of existing knowledge. Central findings are presented not as asserted conclusions but rather as premises or presuppositions from which a range of facts follow inevitably. The gnomic text can easily be read by one unfamiliar with the tradition as obscuring the distinction between what is already known and what is being newly established, when in the gnomic canon precisely this confers validity.

The grammatical forms and categories of Russian are well suited to the gnomic canon. Definite and indefinite articles, with which English monitors what the writer expects the reader to know, are lacking. Participial and nominalized verb forms can be, and frequently are, used to presuppose (rather than assert) new findings or conclusions (where presupposition means assimilating them to general knowledge rather than to the reader's expected knowledge). For instance, in I.0.3 (p. lxxxiv of the original) we find what is literally

> (The) one-sided and restricted nature of classical historical-comparative Indo-European linguistics lay in the fact that its reconstruction of Proto-Indo-European was the result of exclusively external comparison of the separate daughter systems...

but is translated here as

> Classical comparative-historical Indo-European linguistics was one-sided and restricted, since its reconstruction of Proto-Indo-European was based only on external comparison of the separate daughter systems...

The inadequacy of the received reconstruction is the fundamental thesis of the chapter (the Introduction) and a main reason for writing the book. The close translation, like the original, does not assert the inadequacy but merely presupposes it, consistent with the gnomic canon; but the force of the argument is thereby lost to the English reader. I have therefore used freer translations with finite verbs and assertion in such examples (which are numerous).

In expository Russian the paragraph has little or no grammatical status; or, perhaps more accurately, it is not clearly a distinct level from the sentence. In this book, many paragraphs are single sentences; anaphora and other kinds of reduction hold within these paragraph-sentences but usually not between them,

in which respect they resemble English paragraphs; in the occasional instances where anaphora does hold between them, they thereby resemble English sentences but not paragraphs; the scope of certain operators can cross sentence-paragraph boundaries as it can cross sentence (but not paragraph) boundaries in English. Hence paragraphs had to be created in order to make the text readable in English. The decisions as to which paragraph-sentences would function as topic sentences of English paragraphs, which would be joined together in paragraphs, and where paragraph boundaries would be placed in the English version, were all mine and have imposed on the text a form of organization the original did not have.

The Russian text, in having many self-standing and often separately paragraphed sentences, therefore conveys (in Russian) the impression of considerable conciseness: each important claim is reduced to a single self-standing sentence. The opposite is true in a close English rendition, however. Consider the following from I.2.3.2 (pp. 100-101 in the original):

Close rendition:

> This development of Indo-European *$\hat{k}^h$ in Anatolian often coincides with the reflexes of the palatovelars in *satem* dialects. However, the coincidence in the development of palatovelars in Anatolian and the *satem* languages is only a superficial one, not the result of common internal causes.

Free translation:

> This treatment of IE *$\hat{k}^h$ partly coincides with the reflexes of palatovelars in the *satem* languages. However, the coincidence is only superficial and not the result of identical internal causes.

It is the paragraph-internal anaphoric reduction rules of English, together with the definite article, that make it possible to reduce *the coincidence in the development of palatovelars in Anatolian and the satem languages* to a simple *the coincidence*.

Since the original presents well-argued content in good style, I have attempted to convey the content accurately in English of good or at least normal style. This has meant loss of the gnomic style in favor of one that monitors what the reader knows and uses assertion in many places where the original uses presupposition. More generally, trying to put good Russian into the very different expository canon of English has weakened the rhetorical integrity of the original. The alternative, however, would be to lose intelligibility and to render good Russian in very odd English. Therefore, I will simply assure the

reader that the Russian original has an integrity and cohesion of style, expository canon, logic, and content that could not be replicated in the translation, and I take responsibility for any distortion or loss of integrity imposed by the translation.

Various substantive changes have been made to the text and forms. The Russian original contains a section of addenda and corrigenda at the end; these have been incorporated into the text and footnotes. Other authorial additions and corrections, and some editorial ones made with the authors' approval, have also been incorporated. The occasional typographical or copying error has been corrected. For most of the cited forms, glosses have simply been back-translated from the Russian glosses in the original; but for a number of them the standard sources have been consulted and those glosses used here. Graphic and other conventions in tables, figures, formulas, phonological rules, and the like have sometimes been adapted to current or more familiar western norms. The original refers to many classics of linguistic analysis in their Russian translations; these have been replaced with references to the English, French, and German originals wherever possible. The original uses good published Russian translations of Homeric citations, and I have used good published English translations (*The Iliad of Homer,* translated with an introduction by Richmond Lattimore [University of Chicago Press, 1951]; and Homer, *The Odyssey,* translated by Robert Fitzgerald [Garden City, New York: Doubleday Anchor Books, 1963]).

The Russian original had separate Cyrillic and Latin bibliographies (since the two alphabets have different alphabetical orders, they cannot be intermingled in alphabetized bibliographies); the translation transliterates the Cyrillic references and merges the two bibliographies. When the author of works in Russian also has publications in a western language, the author's name is spelled here (in all references) as it is in the non-Russian publications; otherwise the last name is simply transliterated. In the original, if one author had (say) three publications from the year 1978, they were referenced as 1978, 1978a, 1978b (rather than, as would generally be done in this country, as 1978a, 1978b, 1978c). The original reference system has been preserved in the translation (apart from changes imposed by merging the bibliographies, as when, say, an author had a 1978 publication in Russian and a 1978 publication in English). In general, I have tried to keep the bibliography as similar as possible to that of the original (apart from merging, addenda, and corrigenda).

Transliteration has been adjusted to current western norms for a few languages (notably regarding the graphies $u̯, i̯$ in Anatolian and Proto-Indo-European). The original uses boldface for all reconstructed Proto-Indo-European forms in the text, but italics in footnotes; the translation uses boldface throughout (correcting occasional other inconsistencies in the original). The original writes ***p**[*h*], ***t**[*h*], ***b**[*h*], ***d**[*h*], etc. to indicate that aspiration was a

phonetically relevant but not distinctive feature of the stops traditionally recon-
structed as **p, *t, *bh, *dh*, etc.; the translation uses ***pʰ, *tʰ, bʰ, *dʰ**, etc., thereby
saving two keystrokes and three font changes per token as well as giving the
reconstructions a more familiar graphic appearance.

Section and subsection numbers are as in the original (the translation prints
the entire section or subsection number, beginning with the chapter number but
not including the volume number; the original leaves out the chapter number
and also refrains from showing chapter numbers in running headers). Cross-
references are to sections or subsections, not (as in the original) to pages.
Footnotes are numbered consecutively within chapters (the original numbers
them page by page) and include some addenda and corrigenda.

For all of these reasons, the translation can be taken as an updating of the
original and an improvement in the graphic and editorial quality of the original,
but the Russian prose of the original is still to be considered authoritative as
regards wording and precise details of argument. For these same reasons, a
reader looking into the Russian original may find it difficult to pinpoint a
Russian sentence corresponding exactly to a particular English sentence.

The authors would have liked to change and expand the text in various ways
and respond to reviews and other published commentary of the past ten years,
the editors wished to add comments of their own, and I myself would have liked
to make annotations and adduce further supporting evidence at some points.
Such requests have been turned down, since the publisher's original plan was to
produce a translation and not a revised second edition. Readers should keep in
mind, therefore, that the translation reflects the authors' thinking as of about
1983.

Some of the research into forms and glosses, and innumerable questions on
the translation of technical terms from various philological traditions, botanical
nomenclature, English-language titles for ancient texts, and the like, as well as
standard transliterations of various languages and standard English renditions of
place names, personal names, names of some languages, etc. lay far beyond my
own expertise, and I owe a large debt of gratitude to those whose expertise and
willingness to answer questions have made this translation possible. Edgar
Polomé and Winfred Lehmann edited the first drafts of Parts II and I respec-
tively; Werner Winter edited the final draft. Gary Holland has answered
countless questions on every aspect of Indo-European. Thomas Gamkrelidze has
answered many questions, discussed many technical points, made available
proofs and advance copies of the Russian original, and offered institutional
hospitality while I worked in the Oriental Institute of the Georgian Academy of
Sciences in Tbilisi. Martin Schwartz and Calvert Watkins read and commented
on earlier drafts of some chapters. The scholars who have answered my
questions on the languages and areas of their expertise are too numerous to be
listed, so I thank them all anonymously. Orin Gensler did most of the typeset-

ting and in addition has done proofreading, general troubleshooting, and extensive checking of forms and glosses, as well as assisting with Afroasiatic and Celtic. Marie-Louise Liebe-Harkort kept the project alive with her organizational skills, professional knowledge of every aspect of linguistic publishing, and enthusiasm. Richard Rhodes served as her local representative, organizing the production of final copy including the final formatting and some of the typesetting, down to the actual design of some special font characters. He and Orin Gensler produced the indexes. Thomas Hedden checked botanical and zoological nomenclature, verified and updated bibliography entries, merged the (separately alphabetized) Cyrillic and Latin bibliographies of the original, and helped with various linguistic and philological problems. Mary Rees and Ellen Rosenbaum assisted with computerization, bibliography, and a variety of research tasks. Kenneth Whistler converted disk copies to Macintosh format and made computerization possible in the early stages. Margaret Kabalin and Ruth Shields did technical typing of specialized characters and assisted with the production of the final draft. Joseph Schallert assisted with the first stage of translation. Peter Chang and Richard Rhodes produced the line drawings. My native-speaking Russian colleagues, as always, were generous with their time and their finely-honed intuitions and stylistic and textual sense, and I thank Arkady Alexeev, Olga Astromova, Boris Gasparov, Olga Hughes, and Igor Mel'čuk for their help. My debt of gratitude is enormous in the case of Edgar Polomé and Gary Holland, whose time and expertise were given so unstintingly and so far beyond the call of duty.

Though I could not have done the translation without the help of these colleagues and assistants, the decision to seek out their expertise and to use or not use their advice was always mine. Therefore, I take all responsibility for the felicity and accuracy of the translation and more generally for the scholarly qualities of the English text. I also emphasize that none of the editors and consultants has reviewed all forms cited from his or her language of expertise.

Some of the research assistance was supported by the Center for Slavic and East European Studies and by the Committee on Research, both of the University of California, Berkeley. Some of the translation and consultation with the authors were done in Tbilisi when I was a participant in the 1984 Exchange of Senior Scholars between the American Council of Learned Societies and the Soviet Academy of Sciences, sponsored by the International Research and Exchanges Board.

The Russian original was written in about ten years, and spent another five years in press. The translation and its typesetting have taken ten years because of the size and complexity of the text, the number of specialized symbols and diacritics, the decision to set the text from my draft disk copy, and the range of philological and linguistic expertise required to transliterate, check, gloss, or even simply reproduce the forms. And this is for translating a text that already

existed. Meanwhile, the authors wrote the entire work from scratch, including working out the argumentation, selecting the evidence, and assembling the linguistic data (I call the reader's attention to the fact that citations of Hittite and Luwian data are generally referenced not to secondary sources but to the primary texts and are usually transliterated directly from the cuneiform spelling), as well as overseeing production of a typescript, in the same amount of time it has taken to complete and typeset the translation, and they did it under the highly disadvantaged conditions in which serious scholars worked in the Soviet Union. Furthermore, their published text systematically used three alphabets — Cyrillic for the main text and forms cited from Cyrillic orthographies, Greek for Greek forms, and Latin for all others — in addition to a good number of special symbols, while this translation essentially uses only the Latin alphabet and the special symbols (plus Greek, for lengthy Homeric quotes only); and their published text is remarkably free of errors, typographical and other, for any publication and especially a Soviet one. A preface usually ends with an expression of personal gratitude, but instead of gratitude I will take the opportunity to express my admiration to Thomas Gamkrelidze and Vjačeslav Ivanov for their erudition, their dedication, the magnitude and depth of their research for this project, and the intellectual quality and polish of the Russian original.

Johanna Nichols,
Berkeley, July 1994

Foreword

Among the favorite themes and main tasks of linguistics from the last century to the early years of this one were questions of the reconstruction of Proto-Indo-European, and in the world's universities the chief, and usually the only, linguistics department was a department of comparative Indo-European linguistics. It was that epoch whose efforts are summed up in the classic handbooks, efforts directed at revealing the diverse particulars of the common protolanguage underlying the genetically related members of what is known as the Indo-European linguistic family.

During the course of the twentieth century a change in the basic mission of linguistics has made itself felt — on the one hand in the increasing frequency with which technical means and methods developed in Indo-European linguistics were applied to other language families in both the Old and the New World, on the other hand in an increased enthusiasm for the strictly descriptive approach to individual languages without historical comparison either to earlier stages of that language or to related languages. These two spheres of research interests inevitably led to verification and critical reassessment of the inherited methodology.

On the one hand, comparative-historical inquiries concerning the ancestors and interrelationships of various language families deepened and enriched the problematics of linguistic reconstruction, while on the other it was descriptive linguistics that raised fundamental questions about the linguistic system and its regular structure, and in particular laid the groundwork for systematic inquiry into the relation of sound structure and meaning.

A process of integration naturally arises between the expansion of comparative-historical problematics and the commitment of descriptive linguistics to the discovery of systematic structure: the necessity of restricting the tasks of comparative linguistics to strictly genetic comparison falls away, while questions of systematic structure finally go beyond the bounds of descriptive linguistics and find welcome application to the historical past of attested and reconstructed languages.

The first step in this process is to acknowledge the inseparability of the regular system and its changes, which are also regular. The limits of linguistic comparison shift considerably, and new tasks accrue to the study of the general patrimony of linguistic families. The commonalities acquired by the phonological and grammatical structures of languages that are spatially adjacent and enter into areal relations can now be explained. Then there arises the possibility and even the necessity of comparing different linguistic (and chiefly phonological) systems without regard to their genetic or geographical closeness.

As a consequence of the comparative analysis of all these linguistic systems, systematic typological classification grounded on rational principles is now

feasible. In these efforts the facts of living languages, supported by documentation of historical languages, make it possible to check the plausibility of proto-systems reconstructed by the comparative method, and they conclusively identify the most appropriate solutions to difficult problems of reconstruction. In a word, typological comparison renders salutary aid to comparative-historical procedures.

All of these newly discovered or at least newly rethought linguistic principles now confront each concrete linguistic work with inevitable and inescapable demands. Gamkrelidze and Ivanov's *Indo-European and the Indo-Europeans* fulfills in every respect the goal promised in its subtitle, 'A reconstruction and historical typological analysis of a protolanguage and a proto-culture'. The deep shifts and transformations that characterize the stage now attained in linguistics, and in which no small creative role has fallen to these two authors, lie at its methodological foundation. The approaches to particular problems of Proto-Indo-European linguistic antiquity taken by researchers from around the world are brought to bear here, and an appealing answer is given to the various theses that entered scientific currency at the turn of the century. This work stands out not only for its unusual answers to old questions, but in the very way it poses questions and the unprecedented breadth of its thematic horizon.

Consistent with the dialectic removal of the dichotomy of synchrony and diachrony and with the parallel inclusion of spatial diffusion among internal linguistic factors, the book naturally transforms the time-honored, spatially and temporally uniform view of Proto-Indo-European and creates a model of dynamic synchrony which fully comprehends the foundations of the protolanguage, its evolutionary shifts, its internal, regional differentiation, and its recurrent intersections with neighboring linguistic areas. It is the questions of mutual interactions among the dialects of Proto-Indo-European and the relations of the protolanguage to neighboring protolanguages that have given rise to the authors' richly promising work on the geographical definition of the (Southwest Asian) Indo-European homeland and the early migratory routes followed by the various branches of Proto-Indo-European.

The widened range of treatments of two concepts — comparison and system — in modern linguistics is linked to a consistently progressive relativization of all of linguistics and the steady transformation of linguistics into a science of language-internal relations, while the attention of linguists, especially Ivanov and Gamkrelidze, is concerned primarily with the unbreakable mutual connection of parts and whole, especially the central notion of the entire complex problematic: the relation of invariant and variation, the essential theme of all contemporary scientific thought. The dependency of variation on a diversity of contexts becomes all the more clear with the development of the main thesis of contemporary linguistics, which opposes context-free languages, i.e. artificial formal systems, to context-sensitive natural language. Here, of course, variation of form and meaning plays an essential role: both on the sound plane and at various levels of grammatical meaning the systematic extraction of invariants

grows into a central linguistic task.

This entire methodological program is realized in the reconstruction of Indo-European. In breaking the phoneme down into its minimal distinctive components the notion of context, formerly limited to the temporal sequential context of phoneme combinations, has expanded to include simultaneous combinations, and this double assessment of phonological combinations continues to reveal new, previously unstudied typological regularities both within and between the two kinds of combinations. The authors have made notable typological contributions on favored and disfavored combinations of differential components on the axis of simultaneity (see Gamkrelidze's chapter in *Problemy lingvističeskoj tipologii i struktury jazyka*: Institute of Linguistics, Soviet Academy of Sciences, 1977) and on the varieties of symmetrical relations that Ivanov has shown to lie at the foundation of linguistic structure. In particular, the picture of Indo-European consonantism gains novel, internally convincing shape in the work of these authors.

As the problematics of context is developed, the simplistic treatment of stylistic variants as free variants yields to an understanding of style as a context of its own, and the conditions imposed on language by various speech functions are incorporated into the general understanding of context. We are indebted to the authors' initiative in including Indo-European poetics, in particular metrics and the questions of anagrammatic tradition raised by Saussure, among the tasks of linguistic reconstruction.

It is no accident that reconstruction of protolanguage and reconstruction of proto-culture are treated together here as connected parts of a single whole; a consistently holistic approach requires that the reconstructed proto-lexicon be analyzed into semantic fields and the corresponding prehistoric realia be reconstructed through the prism of the Proto-Indo-European lexicon. The notion of the lexicon as a structural system, which has usually lagged behind the phonological and grammatical planes in linguistic work, is reliably grounded here, and in such areas as mythology and ritual it shows the way to systematic application of the comparative method.

In the number and magnitude of the questions it asks and answers it proposes this work occupies a unique place. Fully consistent with the highest standards of contemporary theoretical work, the book in turn will certainly provide valuable impetus not only to linguistic analysts of all schools, but also to specialists in related fields, for instance ethnographers, culture historians, and archeologists. A great deal of fruitful discussion will come forth in international science as a result of this momentous work.

Roman Jakobson

Contents

Volumes I and II of the text of the original Russian edition have been combined in the English version as Part I; the Bibliography and Indexes are published as Part II.

I

Part One
The Structure of Proto-Indo-European

Section One: The Phonological System and Morphophonology of Proto-Indo-European

Section Two: The Grammatical Structure of Proto-Indo-European

Section Three: The Areal Organization of Proto-Indo-European

Part Two

Semantic Dictionary of the Proto-Indo-European Language and Reconstruction of the Indo-European Proto-Culture

Section One: Semantic Dictionary of Proto-Indo-European

Section Two: The Chronology of Proto-Indo-European. The Indo-European Homeland and Migration Routes to the Historical Territories of the Indo-European Tribes

I I

Detailed Contents

Volumes I and II of the text of the original Russian edition have been combined in the English version as Part I; the Bibliography and Indexes are published as Part II.

I

Part One
The Structure of Proto-Indo-European

Introduction — The linguistic system and the premises of diachronic linguistics

Section One: The Phonological System and Morphophonology of Proto-Indo-European

Chapter One — The three Indo-European stop series: Paradigmatics and syntagmatics

**Chapter Two — The Indo-European points of stop articulation
and the Indo-European sibilants: Paradigmatics and syntagmatics**

**Chapter Three — The vowel system and the theory of morphopho-
nological alternations. Sonants and laryngeals in Indo-European**

Chapter Four — The structure of the Indo-European root

Section Two: The Grammatical Structure of Proto-Indo-European

Chapter Five — Proto-Indo-European as a language of the active type

Section Three: The Areal Organization of Proto-Indo-European

Chapter Seven — The differentiation of the Indo-European linguistic region

Part Two

Semantic Dictionary of the Proto-Indo-European Language and Reconstruction of the Indo-European Proto-Culture

**Introduction — Methods for reconstructing the semantic dictionary
of a protolanguage and the linguistic paleontology of culture**

Section One: Semantic Dictionary of Proto-Indo-European

Chapter One — The living world: Gods, people, animals

Chapter Two — Indo-European conceptions of wild animals, and names for them

Chapter Three — Indo-European terms for domestic animals. The economic functions of animals and their ritual and cultic role among the early Indo-Europeans

Chapter Four — Indo-European plant names. Uses of plants; their ritual and cultic functions in ancient Indo-European culture

Chapter Five — Geographical environment and climate. The Indo-European terms for heavenly bodies

Chapter Six — Economic activity, material culture, crafts, transportation

Chapter Seven — The social organization, economy, and kinship system of the ancient Indo-Europeans

Chapter Eight — The connection of ancient social organization with intellectual constructs and the mythological view of the world

Chapter Nine — Reconstruction of Indo-European rituals. Legal and medical conceptions. The afterworld and burial rites

Chapter Ten — Reconstruction of Indo-European text fragments. Fragments of poetic speech; Indo-European metrical schemes. The counting system and number symbolism

Section Two: The Chronology of Proto-Indo-European. The Indo-European Homeland and Migration Routes to the Historical Territories of the Indo-European Tribes

Chapter Eleven — Proto-Indo-European in space and time, based on linguistic and culture-historical data

**Chapter Twelve — The migrations of the Indo-European-speaking
tribes from their Near Eastern homeland to their historical ter-
ritories in Eurasia**

II

The languages (Indo-European and non-Indo-European) and their written sources

1. Indo-European languages

1.1. Anatolian

The Anatolian languages of ancient Asia Minor — Hittite (written in cuneiform), Luwian (cuneiform and hieroglyphic), and Palaic — are attested in documents from the Boghaz-Köy archive of the second millennium B.C. The oldest Hittite cuneiform texts can be dated by paleographic features (their ancient ductus) to the period of the Old Hittite Kingdom (first half of the second millennium B.C.) and include royal inscriptions: the oldest, that of King Anittas (18th century B.C.: Neu 1974), the Annals of King Hattusilis (17th century B.C.: Imparati 1965), and the Will of King Hattusilis I (Sommer and Falkenstein 1938); there are also historical inscriptions (that of Zukrashi: Otten 1953) and others from KUB and KBo (see Laroche 1971). Also composed in the Old Hittite period is the early variant of the Hittite Laws (Friedrich 1959; Imparati 1964). An archaic form of the language similar to that of Old Hittite is found in many ritual and mythological texts: the Myth of Telepinus (KUB XVII 10; XXXIII), the metrical text about the god Pirwa (Bo 6483: Otten 1951), a burial song (also metrical) (KBo III 46), hymns to the sun (KUB XXI 127-34, XIV 74), a building ritual (KUB XXIX 1), royal burial rituals (Otten 1958), the Prayer of Mursilis II in a Time of Plague (Goetze 1929), and others. Archaic features are also preserved in Middle Hittite texts such as the Text of Madduwattas (Otten 1969): see Heinhold-Krahmer et al. 1979. In the Late Hittite period (14th-13th centuries B.C.), Hittite texts show significant influence from spoken Luwian (for instance, Luwian words marked with the special cuneiform Glossenkeil sign ⦥).

Luwian is attested in cuneiform texts, chiefly rituals, of the time of the Hittite Kingdom as well as in later hieroglyphic inscriptions of southern Asia Minor and northern Syria, written in their own hieroglyphic script (Laroche 1960:I; Meriggi 1966-1975).

Palaic is known from fragments, chiefly of mythological and ritual texts, found among the Hittite cuneiform inscriptions (Kammenhuber 1969a).

The later Anatolian languages of Asia Minor are descendants of Hittite and Luwian: these are Lydian and Lycian, attested in alphabetic documents of classical times (Houwink ten Cate 1961, Gusmani 1964, Heubeck 1969, Neumann 1969, Laroche 1974; cf. also Zgusta 1964a).

1.2. Indo-Iranian (Aryan) languages

The earliest evidence of an Indo-Iranian dialect is Mitannian Aryan, attested in the form of Indo-Iranian words and deity names contained in Hittite texts, chiefly texts about the training of horses (Kammenhuber 1961, Mayrhofer 1966).

The oldest actual texts in an Indo-Iranian language are the Old Indic texts of the Rigveda, written down in Indian syllabary in the first millennium B.C. but composed much earlier, probably in the second millennium B.C. The oldest hymns of the Rigveda were probably composed before the entry of Indo-Aryan tribes into northwest India (Aufrecht 1955, Elizarenkova 1972, 1982). The archaic Vedic language that represents the earliest form of Indic is also used in the Atharvaveda (M. Bloomfield 1899, Elizarenkova 1976). The later literary form of the language is often simply called Sanskrit (*saṁskṛta-* 'perfected').[1]

Sanskrit literature in the broad sense also includes archaic prose texts of religious and philosophical content, later than the Vedic texts: the Brahmanas and Upanishads (see Van Buitenen 1962, Satya Shrava 1977). The ancient juridical tradition is reflected in numerous texts, of which the best known are the Laws of Mana (*Mānava Dharma Śāstra*): see Nārāyaṇ Rām Āchārya 1946. Sanskrit continued to be used as a literary language parallel to the spoken Middle Indic Prakrits from which today's Indo-Aryan languages evolved: Hindi, Bengali, Panjabi, Sindhi, and others (Bloch 1934).

A separate branch of Indo-Iranian is the Kafir, or Nuristani, languages spoken in the mountainous part of Afghanistan (Nuristan, earlier Kafiristan). These are unwritten languages: Kati, Ashkun, Waigali, Prasun (Strand 1973, Grjunberg 1971, 1980; for individual languages see Morgenstierne 1929, 1949, 1954, Fussman 1972). Related to the Kafir languages are the Dardic languages, which fall into two subgroups, Central Dardic (Dameli, Pashai, Gawar-Bati, Shumashti, and others) and Eastern Dardic or Dardic proper (Phalura, Kashmiri, Shina, Garwi, and others): Èdel'man 1965.

The other branch of Indo-Iranian is Iranian, of which Avestan and Old Persian are attested in ancient documents. The hymns of the Avesta, written in Avestan (which has Eastern Iranian dialect traits), were composed in the second and first millennia B.C. but written down in alphabetic writing much later, in the first millennium A.D. The four major parts of the Avesta that have come down to us are the Yasna — which includes the Gathas, hymns attributed to Zarathustra (Humbach 1959) — the Vispered, the Videvdat (or Vendidad), and the Yashts (hymns), as well as fragments (Geldner 1886-1895). Old Persian, which represents the Western Iranian dialect type, is known from cuneiform documents of the Achaemenid period (6th to 4th centuries B.C.), historical in

---

1. [Sanskrit citations in this book are largely from Vedic. Hence they are usually identified as simply Sanskrit, abbreviated Skt. — JN.]

content (Herzfeld 1938, Cameron 1951, Brandenstein and Mayrhofer 1964). Median is the language of the Medes (8th to 6th centuries B.C.), reconstructed from individual words (personal names and tribe names) found in Assyrian and Greek sources and in Old Persian inscriptions (see Mayrhofer 1968).

The Middle Iranian languages are divided into eastern and western groups. Eastern Iranian languages include Khotanese Saka (texts of the first millennium B.C. from Central Asia: Bailey 1945-1956, 1951), Sogdian (Benveniste 1940, Henning 1940, Mackenzie 1976, Livšic and Xromov 1981:347-514), Khwarezmian (Frejman 1951), and Bactrian (Humbach 1966-1967, Steblin-Kamenskij 1981:314-16). Western Iranian languages include Middle Persian, or Pehlevi, attested in numerous texts from Sassanid Iran (Henning 1955), and Parthian (texts from Central Asia: Diakonoff and Livšic 1960, Gignoux 1972, Diakonoff and Livšic 1976).

Modern Iranian languages of the eastern group include Ossetic of the Caucasus, with two dialects, Iron (eastern) and Digor (western), considered to be derived from Scythian, which is known from individual words and personal names in the writers of classical antiquity (Abaev 1949); Yagnobi in Central Asia, a direct descendant of Sogdian; Pashto or Afghan; Munji (and its Yidga dialect); and the Pamir languages: Shugni, Rushan (and dialect Khuf: Sokolova 1959), Bartang, Oroshor, Sarikoli, Yazgulami, Ishkashim (and dialect Sanglechi), and Wakhi (see Paxalina 1959, Grjunberg and Steblin-Kamenskij 1976). Western Iranian languages include modern Persian, Tajik, Kurdish, Baluchi, Tat, Talysh, Ormuri, Parachi, and several dialects of central Iran (see Morgenstierne 1929-1938).

1.3. Armenian

Classical Armenian (Grabar) is known from numerous texts going back to the fifth century A.D., including Bible translations and original texts such as the History of Armenia by Moses of Khorene, both of which include fragments of older texts from the prehistoric period of Armenian (Schmitt 1981:215ff.). There are two main groups of Armenian dialects: eastern (in the Transcaucasus) and western (Meillet 1936).

1.4. Greek

The earliest form of an ancient Greek dialect is Mycenean, known from Cretan and Mycenean documents in the Linear B script dating to the 15th-13th centuries B.C. (Morpurgo 1963, Ventris and Chadwick 1973). The oldest texts of the next period are the Iliad and the Odyssey of Homer. The main dialects of

ancient Greek are Attic-Ionic (including the Ionic dialects of Asia Minor and the Cyclades as well as Euboean and Attic), Achaean, including Aeolic and northern Achaean (Thessalian, Boeotian, and Aeolian of Asia Minor and Lesbos), and southern Achaean (Arcadian on the Peloponnese, Pamphylian or Cypriot on Cyprus; this dialect is close to Mycenean); Doric and Northwest Greek (spoken in Laconia, Messenia, the Argolid, the islands of Aegina and Crete and other Aegean islands, and Corinth and Megara). See Buck 1910, Bechtel 1921-1924, Schmitt 1977; also Risch 1979a.

1.5. Phrygian

Phrygian is known from inscriptions from the first half of the first millennium B.C., from northwestern Asia Minor. A few of these are early inscriptions dating back to the seventh century B.C., written in an archaic script similar to the early Greek alphabet; most are later inscriptions of the Roman period, written in the ordinary Greek script of the time (Gusmani 1958, O. Haas 1966, Diakonoff and Neroznak 1977, Neroznak 1978).

1.6. Tocharian

The Tocharian languages are attested in texts from the second half of the first millennium A.D., at the easternmost periphery of Indo-European speech in Eurasia, in Eastern Turkestan [Xinjiang]. There are two languages, conventionally called Tocharian A (East Tocharian) and Tocharian B (West Tocharian). Most of the texts are translations of Sanskrit Buddhist documents, but there are a few original texts: business letters, monastic administrative texts, accounting documents (Sieg and Siegling 1921, 1949-1953, W. Thomas 1964).

1.7. Albanian

Albanian, in the western Balkan peninsula, as attested from the 16th century A.D. There are two main dialects: Geg (northern, in Albania and Kosovo) and Tosk (southern): see Desnickaja 1968, Solta 1980.

1.8. "Ancient European" languages

The term "Ancient European" will be used to include the dialectally and areally related Indo-European languages of Europe from the end of the second millen-

nium B.C. to the beginning of the first millennium B.C. (Krahe 1951, 1954, 1959, 1962).

1.8.1. Italic languages. This is a family of languages of the Apennine Peninsula of ancient times, including the Latin-Faliscan and Osco-Umbrian subgroups, and attested in documents from the first millennium B.C. (Vetter 1953, Solta 1974):

Latin (originally the language of Latium and Rome) and the closely related Faliscan (the dialect of Falerii in southern Etruria) are attested in inscriptions going back to the sixth century B.C., written in a Greek-derived alphabetic script. The oldest Latin is attested in several inscriptions (Ernout 1950:274ff.),[2] Saturnian verse, and the works of early Roman authors (Plautus, Terence).

The Osco-Umbrian subgroup includes the dialects of the Oscans (inscriptions in Samnia and Campania), the Volsci, the Umbrians, and other Sabellian tribes. The most important Umbrian document is the bronze Iguvine tablets (from Iguvium): Poultney 1959, Vetter 1955, Ernout 1961. At the beginning of the present era the Osco-Umbrian languages yielded to Latin.

Venetic. A distinct Indo-European dialect known from brief inscriptions of the fifth to first centuries B.C. in northeastern Italy (Beeler 1949, Krahe 1950, Untermann 1961).

Illyrian. Known from brief Messapic inscriptions from Calabria and Apulia (southern Italy). Illyrian is also known from onomastics of the Italic peninsula, the northwest Balkan peninsula, and adjacent regions (Krahe 1955, Mayer 1957-1959; cf. Tronskij 1953:57-59).

1.8.2. Celtic languages. The Celtic languages fall into two groups: continental Celtic and insular Celtic. On the evidence of Celtiberian inscriptions, speakers of continental Celtic dialects lived in Iberia from the first half of the first millennium B.C. (Lejeune 1955a, Tovar 1961, Untermann 1961a), and in Gaul (modern France) in Roman times, from which there are a number of short inscriptions in Gaulish. Other groups of Celts lived in central Europe: southern and western Germany, the Alpine regions, Pannonia, Italy, and the Balkan peninsula. Insular Celtic, found on the British Isles, includes the Goidelic and Brythonic dialect groups. Goidelic includes Old Irish, attested in Ireland in the Ogam script from the fourth century A.D. and in later texts in the Latin alphabet, as well as Scots Gaelic and Manx. The Brythonic group includes Welsh (Old Welsh is attested in glosses and written documents from the 11th century A.D.), Cornish, and Breton, brought to Brittany by immigrants from Britain in the 5th century A.D. (Old Breton glosses go back to the eighth to eleventh centuries: Jackson 1953).

---

2. The Praenestine fibula (allegedly from 600 B.C.), formerly considered to be the oldest Latin inscription, has recently been shown to be a forgery created in the 19th century (see Guarducci 1980, Pfister 1983).

1.8.3. Germanic languages. The Germanic languages are usually divided into three groups: Scandinavian (or North Germanic), East Germanic (Gothic), and West Germanic. North and East Germanic are put in a single Gothic-Scandinavian group by some investigators. The earliest attested Scandinavian language — the language of the ancient Scandinavian runic inscriptions — is still close to Proto-Germanic (Makaev 1965). The oldest texts in Old Icelandic (Old Norse) are collected in the Elder Edda (Neckel 1962), a parchment manuscript miscellany of old songs compiled in Iceland in the 13th century but composed much earlier. Old Icelandic also has an extensive prose literature and skaldic poetry. In about the middle of the second millennium A.D. Old Icelandic (Old Scandinavian) split into West Scandinavian (Norwegian and Icelandic) and East Scandinavian (Swedish and Danish) (M. Steblin-Kamenskij 1953).

The primary representative of East Germanic is Gothic, whose oldest document is a Bible translation done by the bishop Wulfila in the 4th century A.D. Gothic was spoken in eastern and southeastern Europe in lands belonging to the Byzantine sphere of influence. There is evidence for the presence of Goths in Byzantium itself, as well as on the Crimea (in the form of a Crimean Gothic word list: Žirmunskij 1964:85-102).

West Germanic (or South Germanic if opposed to a Gothic-Scandinavian branch) includes Old English (or Anglo-Saxon, with texts from the 7th century A.D.), Old Frisian, Old High German (texts from the 8th century A.D.), and Old Saxon (texts from the 9th century), the earliest representative of the Low German dialects. English and Frisian make up an Anglo-Frisian subgroup, opposed to High German dialects, while Low German is intermediate between these two. The earliest poetic texts in West Germanic languages include the Old English epic of Beowulf (Irving 1968, 1969, Wrenn 1973), the Old High German Song of the Nibelung (Körner 1921), and the Old Saxon epic Heliand (Behaghel 1933). The descendants of these languages are modern English, German, Flemish, and Dutch (the latter from Low German dialects).

1.8.4. Baltic languages. There are two groups: West Baltic, represented by Old Prussian (with written texts from Prussia in the 14th-18th centuries A.D.: Mažiulis 1966-1981), which subsequently yielded to German, and East Baltic, including Lithuanian and Latvian (texts go back to the 16th century; see Zinkevičius 1980:I.15-18).

1.8.5. Slavic languages. There are three groups: East Slavic (Russian, Ukrainian, Belorussian, with early documents going back to the 11th century); West Slavic (Polish, Slovincian-Kashubian, Czech, Slovak, Sorbian, and extinct Polabian); and South Slavic (Bulgarian, Serbo-Croatian, Slovene, and others), with early texts in Old Church Slavic (translations of Greek texts beginning in

the 11th century A.D.: the Zographensis and Marianus Gospel translations in the Glagolitic alphabet, and in the Cyrillic alphabet the Savvina Kniga of Gospel readings, the Sinai Psalter, the Suprasliensis collection of lives of saints and prophets, and others).

2. Non-Indo-European languages in adjacent parts of Eurasia

2.1. Ancient Near Eastern languages

Hattic. The non-Indo-European language of the indigenous population of northeastern Asia Minor, attested in the form of fragments in Hittite texts. Extinct since the early second millennium B.C. (Kammenhuber 1969).

Sumerian. A non-Indo-European language of Mesopotamia, attested in early pictographic and cuneiform texts going back to the 4th and 3rd millennia B.C. Subsequently yielded to Akkadian. There is evidence for different periods in the history of Sumerian, which point to changes in phonetics and grammatical structure.

Elamite. A non-Indo-European language of southwestern Iran (the mountain valley of the eastern Tigris, modern Khuzistan), with hieroglyphic and cuneiform texts going back to the beginning of the second millennium B.C. Yielded to Old Persian by the first millennium (Diakonoff 1967, Reiner 1969).

Hurrian-Urartean. A non-Indo-European language family, attested from the third to first millennia B.C. in various parts of Southwest Asia: upper Mesopotamia, northern Syria, Asia Minor, the southern Transcaucasus, and northwest Iran. The earliest appearance of the Hurrians in the Near East (specifically, northern Syria) was in the mid-third millennium B.C., at which time in the vicinity of Ebla (modern Tell Mardik) there flourished a powerful Semitic-speaking state with Hurrians as one of its ethnic components. The earliest Hurrian texts date from the second half of the third millennium. The Nawara inscription, from the valley of the Diyala, an eastern tributary of the Tigris, was written in Old Akkadian for the Hurrian king Arižen (or Ari-šen or Atal-šen): see Diakonoff 1967:114, Wilhelm 1982). The oldest Hurrian document in Hurrian proper is an inscription from Urkish in northern Mesopotamia, also written in Old Akkadian cuneiform of the same kind (see Diakonoff 1967:6-7, Haas et al. 1975:24). In the second millennium B.C. there are many Hurrian texts from Mari, Boghaz-Köy, and Ugarit (the latter using two different writing systems: a syllabic logographic type and a consonantal-syllabic one of the ancient Semitic type). Hurrian was the main language of the Mitannian kingdom of Mesopotamia in the mid-second millennium B.C. The El Amarna archive yielded an extensive Hurrian text in the form of a letter from the ruler Tushratta to the Egyptian Pharoah Amenhotep III from the early 14th century

B.C. (Speiser 1941, Laroche 1978).

A later Hurrian dialect is Urartean of the Kingdom of Van in the southern Transcaucasus, written in Assyrian cuneiform. The texts date to the first half of the first millennium B.C. (Melikišvili 1960, Diakonoff 1971).

2.2. Semitic languages

A family of Near Eastern languages attested in written texts from the 2nd millennium B.C.:

East Semitic comprises Akkadian, with Babylonian and Assyrian dialects, spoken in Mesopotamia and adjacent regions. The following stages are distinguished for Akkadian: Old Akkadian (24th-22nd centuries B.C.); Old Babylonian (in southern Mesopotamia) and Old Assyrian (in the middle Tigris and the Cappadocian tablets from Asia Minor), from the beginning of the 2nd millennium B.C.; Middle Babylonian and Middle Assyrian (16th-11th centuries B.C.); Neo-Babylonian and Neo-Assyrian (10th-8th centuries B.C.). The Akkadian inscriptions used a cuneiform script of Sumerian origin, with some changes adapting it to Akkadian dialects.

West Semitic languages from the eastern Mediterranean area (Palestine and Syria). These include the Canaanite subgroup: Paleo-Canaanite or Eblaite (spoken in northern Syria and attested in recently discovered inscriptions from Ebla dated to the mid-third millennium: Pettinato 1975, Gelb 1977); Moabite, represented in one lengthy inscription of the king Mesha (11th century B.C.); Phoenician, with inscriptions beginning in the second half of the second millennium B.C.; (ancient or Biblical) Hebrew, with documents from the end of the second millennium B.C.; and Aramaic, attested in many inscriptions beginning in the first millennium B.C. Several different Aramaic dialects are distinguished: West Aramaic (Nabatean, Palmyrene, Judeo-Palestinian, Samaritan, and others) and East Aramaic (Syriac, Mandaic, and others). The oldest inscriptions in the West Semitic languages are in a linear script of the consonantal-syllabic type going back to an early Semitic prototype.

The Ugaritic language of ancient Ugarit (modern Ras-Shamra, Syria) forms a separate branch within West Semitic. Its texts, dating to the mid-second millennium B.C., are written in a distinct consonantal-syllabic cuneiform.

South Semitic languages include (epigraphic) South Arabian dialects (with texts from the first millennium B.C.); their modern forms Mehri, Shahari, and Soqotri; classical Arabic; Geez (or Classical Ethiopic) and the modern Ethiopic languages Amharic, Tigre, Tigrinya, Harari, and others.

2.3. Ancient Egyptian

The language of ancient Egypt, attested in hieroglyphic documents from the end of the fourth millennium B.C. on: Old Egyptian, Middle Egyptian, and Late Egyptian (the latter from the mid-second millennium B.C.). The continuation of Egyptian is Coptic, the language of the Egyptian Christians, written in a script derived from Greek uncial writing; it is an extinct language, preserved only in liturgical contexts. Egyptian, together with Semitic and several language families of Africa — Berber, Cushitic, and Chadic — makes up the Afroasiatic family (see M. Cohen 1947, Diakonoff 1965, 1975).

2.4. Caucasian (or Paleo-Caucasian) languages

Kartvelian languages. This is a language family of the southern Caucasus, consisting of four related languages: Georgian (with texts going back to the fifth century A.D.), Mingrelian, Laz (or Chan), and Svan.

Abkhaz-Adyghe (Northwest Caucasian). Languages of the northwestern Caucasus, including Abkhaz (with Abaza), Adyghe, Kabardian, and Ubykh.

Nakh-Daghestanian (Northeast Caucasian). Languages of the eastern Caucasus. There are two branches. The Daghestanian branch consists of the Avar-Andi-Tsez group (Avar, Andi, Botlikh, Godoberi, Karati, Akhvakh, Bagvali, Tindi, Chamali, Tsez, Khvarshi, Ginukh, Bezhta, Hunzib); the Lezghian group (Lezghi, Tabassaran, Agul, Rutul, Tsakhur, Archi, Kryz, Budukh, Xinalug, Udi); Lak; and Dargi. The Nakh branch consists of Chechen, Ingush, and Batsbi (or Tsova-Tush).

2.5. Dravidian languages

This is a family of 23 languages spoken for the most part in the southern Indian subcontinent. There are seven major languages: Telugu, Tamil, Kannada, Malayalam, Gondi, Kurukh, and Tulu. Tamil, Telugu, Kannada, and Malayalam have written traditions going back to the beginning of this era. Before the arrival of the Indo-Aryans in India the Dravidian languages were spoken over most of the Indian subcontinent, including the northwestern part; they were displaced by Indic, with which they were in contact for a long period (see Burrow and Emeneau 1961).

2.6. *Munda languages*

A group of languages spoken in India (chiefly its central part) and forming a branch of Austroasiatic, a large family of mostly Southeast Asian languages.

2.7. *Uralic languages*

A family with two branches, Finno-Ugric and Samoyedic. The Finno-Ugric languages are divided into Ugric (Hungarian in central Europe, and Ostyak and Vogul [also sometimes known as Khanty and Mansi] in western Siberia) and Finnic, the latter with subgroups: Finno-Permian (Komi [or Permiak and Zyrian] and Votyak [Udmurt]); Balto-Finnic (Finnish, Veps, Vote, Estonian, Livonian, and others in northeastern Europe); Volgaic (Moksha Mordvin, Erzja Mordvin, Hill Cheremis [Mari], and Meadow Cheremis [Mari]); and Lapp [Saami]. See Hajdú 1975.

2.8. *Altaic languages*

Many scholars group the Turkic, Mongolian, and Tungusic languages together as Altaic.

The Turkic languages are attested in written texts from western Siberia (the Yenisei-Orkhon inscriptions) and Central Asia dating to the seventh century A.D. and written in a Turkic runic script derived from Sogdian (Kononov 1980) and reflecting the Old Turkic (or Old Uigur) language. Turkic languages fall into several subgroups: Chuvash; Southwest or Oghuz Turkic; Northwest or Kipchak; Kirghiz-Altay; and Southeastern (Uigur).

The Mongolian languages include Mongol proper as well as Buriat, Kalmyk, and others such as Oirat, Dagur, and Monguor.

The Tungusic languages include Manchurian and Jurchen (the Manchurian or southern branch) and Evenki, Even, Negidal, Solon, and the languages of the Amur subgroup (the Tungus or northern branch).

2.9. *Paleoasiatic (Paleosiberian) languages*

This is a grouping of genetically unrelated languages of indigenous ethnic minorities of Siberia. Four language families are included in Paleoasiatic: Chukchi-Kamchatkan, Eskimo-Aleut, Yukagir, and Ket-Assan (Yeniseian). The latter group includes several languages of western Siberia extinct since the 17th

to 19th centuries (Kott, Arin, Assan, and others) as well as Ket, spoken on the middle and upper Yenisei.

Also included in Paleoasiatic is the language isolate Nivkh (Gilyak).

2.10. *Chinese*

Chinese is attested in texts written in pictographic script from the second millennium B.C. The oldest texts come from northern and central China. The phonetic forms of Old Chinese words have been reconstructed by comparative analysis of the modern dialects and analysis of words borrowed into other Asian languages — Japanese, Korean, Vietnamese, and others (Karlgren 1940). Old Chinese forms are cited in the transcription of Karlgren 1923, 1940. Modern Chinese forms are in *pinyin*.

The territory of Chinese speech forms the easternmost limit of the non-Indo-European languages which have come into contact with Indo-European dialects in the course of the Indo-European migrations.

Transliteration of languages with non-Latin writing systems

In this work, examples from languages with their own traditional scripts are usually cited not in phonetic transcription (unless explicitly indicated) but in Latin transliteration with some additional diacritics and special symbols. The transliteration systems used are generally standard ones for the languages concerned. When necessary, the phonetic content of transliteration symbols for ancient languages is described below in terms of the phonetic and phonological status of the phonemes they represent. Aspiration of voiceless stops, where not indicated in standard transliteration, is described in this section.

Hittite and other Anatolian languages

The transliteration system for Hittite and the other Anatolian languages (Luwian, Palaic) is a one-to-one system rendering each syllabic sign of the cuneiform (or, for Luwian, hieroglyphic) script with the standard combination of Latin letters used in Assyriology (for Hittite see Friedrich 1960), each sign set off by hyphens. In addition, standard unhyphenated transliteration is also used: thus *akkanteš* or *kuenzi* in addition to *ak-ka-an-te-eš* or *ku-en-zi*. This unhyphenated transliteration makes no claim to phonetic or phonological accuracy but is used for economy and convenience.[3] Thus, for example, the cuneiform symbols for *š* and *z* in Hittite render [s] and [ts], but will be transliterated here as *š* and *z*. Actual phonetic properties of Hittite sounds will occasionally be explicitly indicated by phonetic or phonemic transcription in square or slant brackets.

Mycenean and classical Greek

Transcriptions of Mycenean Greek inscriptions, either syllabic or unhyphenated, follow standard practice for rendering Linear B (see Morpurgo 1963, Ventris and Chadwick 1973). Classical Greek is transliterated in standard fashion, except that citations of long verse passages have been left in the original alphabet.

---

3. It has recently been established that plene writing (or doubling of vowels) represents word stress in Hittite, so that forms such as *e-eš-zi, a-ša-a-an-zi, pa-ta-a-na*, etc. are to be interpreted phonetically as having not length but stress, i.e. *éstsi, asántsi, patán* (see Hart 1980, Carruba 1981). In unhyphenated writing the accent is not indicated (unless otherwise noted), nor is the vowel doubling indicated with a macron (as in *ēszi, asānzi, patāna*).

Sanskrit

Vedic and other Sanskrit forms are cited in the standard transliteration (see Renou 1930, Mayrhofer 1953). The Old Indic writing system (*devanagari*) was a syllabic one organized by phonetic and morphological paradigmatics of the sounds:

Short syllabics:	*a, i, u, r̥,* (*l̥*)
Long syllabics:	*ā, ī, ū, r̥̄,* (*l̥̄*)
Vowels:	*e, o* (functionally, these are diphthongs with short first elements: **ai*, **au*)

Diphthongs with long first element: *ai, au* (functionally, **āi*, **āu*)
Non-syllabic sonorants: *y, v, r,* (*l*)
Stops:

	Voiceless	Voiceless aspirate	Voiced	Voiced aspirate	Nasal
Palatal	*c*	*ch*	*j*	*jh*	*ñ*
Velar	*k*	*kh*	*g*	*gh*	*ṅ*
Cerebral (retroflex)	*ṭ*	*ṭh*	*ḍ*	*ḍh*	*ṇ*
Dental	*t*	*th*	*d*	*dh*	*n*
Latial	*p*	*ph*	*b*	*bh*	*m*

Sibilants: *s, ṣ* (cerebral), *ś* (palatal)
Aspiration: voiced spirant *h* and voiceless *ḥ* (*visarga*), an allophone of *s* in word-final position. The nasal vowel *anusvāra* is indicated with *ṁ*.
Palatal affricates: *c, ch, j, jh* are compact (hushing) sounds, phonetically *č* (= [tʃ]), *ǰ* (= [dʒ]), etc.

Tocharian

Tocharian texts are written in Indic script and transliterated as for Sanskrit, with the addition of *ä*, a distinct front vowel, and *ts* for [ts].

Avestan, Old Persian, and other Iranian languages

Avestan is written in a script derived from Aramaic-Pehlevi with some additions and changes, and is cited in a normalized standard transliteration (see Reichelt 1909, Morgenstierne 1942):

Vowels: $a, \bar{a}, i, \bar{i}, u, \bar{u}, e, \bar{e}, o, \bar{o}, \mathring{a}, \mathring{\bar{a}}$
Reduced vowels: $ə, \bar{ə}$
Nasal vowel: $ą$
Stops:

     Labial: p, b
     Dental: t ($t̰$ word-finally in certain positions), d
     Velar: k, g

Spirants:

     Labial f, w; dental $θ, δ$; velar $x, x^v, γ$

Sibilants:

     Hissing $s, z;$ compact (hushing) $š$[4], $ž, ç$

Aspiration: h (corresponding to several different signs)[5]
Affricates: compact (hushing) $č, ǰ$
Sonorants:

     Nasal: $m, n, ŋ, (ń)$
     Oral: y, v, r

A simpler transliteration system is standard for Old Persian, which was written in syllabic cuneiform (see Kent 1953, Brandenstein and Mayrhofer 1964):

Vowels: $a, \bar{a}, i, \bar{i}, u, \bar{u}$
Stops: b, d, g
                             p, t, k
Affricates: $ǰ, č$
Spirants: v, f
                             $þ$
                             z, s
                             $š, ç$
                             $ḥ$
Aspiration: h
Sonorants: m, n, y, r

Forms from Middle Iranian dialects such as Sogdian, written in a consonantal script derived from Aramaic, are transliterated without vocalism.

---

4. This symbol renders several signs of the Avestan script. According to Morgenstierne, they correspond to three distinct phonemes: /š/, /š´/, /ř/.

    5. In addition to h there is also a sign for a palatalized variant $h´$.

Classical Armenian

The transliteration system used here is that of Meillet 1936, 1937, 1938:45, with some later corrections in the aspirated affricates (Schmitt 1972, 1981):

Stops:
  Voiceless: *p, t, k*
  Voiceless aspirated: *p', t', k'*
  Voiced: *b, d, g*
Affricates:
  Voiceless: *c, č*
  Voiceless aspirated: *c', č'*
  Voiced: *j* (= [dz]), *ǰ* (= [dž])

The phonological opposition of voiceless unaspirated to voiceless aspirated affricates is best understood not as an opposition like that of Indic *č* to *čh* but as an opposition of glottalized (unaspirated) to unglottalized (aspirated), like that of *c'* to *c* or *č'* to *č* in Caucasian languages.

There are distinct letters for two types of non-nasal sonorants: *l* and *ł* (the latter velarized), and *r* and *ř* (the latter a trill which can be interpreted as long or geminate: see Bolognesi 1962, Schmitt 1972:303). *ē* transliterates a closed vowel reflecting a former diphthong and found only in stressed syllables; *ə* is a short tense vowel found in unstressed syllables.

Gothic and other Germanic languages

The Gothic writing system, based on late Greek uncial script, is transliterated with Latin letters and some additional characters: *þ*, a voiceless dental or interdental spirant (like English [θ]), *q* for the labiovelar stop [k°], and *ƕ* for the labialized fricative [h°]. [w] and [j] are transliterated *w* and *j*. The same symbols are used for transliterating runic letters, with *R* indicating a voiced spirant similar to a vibrant.

In Icelandic orthography, the acute accent marks vowel length: e.g. *ý* indicates [ü]. The fricatives [β] and [ð] (and in Old English also [θ]) are written with the barred letters *b* and *đ* (or *ð*).

In Old High German, *z* indicates the affricate [ts].

Slavic languages

Old Church Slavic forms, whether originally written in Glagolitic or Cyrillic, as well as words from modern Slavic languages written in Cyrillic (Russian, Belorussian, Ukrainian, Bulgarian), are written in standard academic transliteration: *y* is a high central or back unrounded vowel, *ě* an open or low front vowel, *ĭ* and *ŭ* short (or reduced) high vowels (where these letters, known as *jers*, are retained in later orthographies including Old Russian, they are transliterated with single and double apostrophes: ' and "); *ǫ* and *ę* are nasalized vowels; *c* is [ts], *č* is [tʃ], and *x* is the voiceless velar fricative. The modern jotated vowels are transliterated with letter sequences: *ja, ju*.

Italic languages

The Italic languages Faliscan, Oscan, and Umbrian, together with other ancient languages of Italy including Messapic and Venetic, are transliterated with ordinary Latin script (in which *c* indicates [k] and *u* indicates [w]). Oscan *í* and *ú*, with acute accent, indicate [e] and [o].

Old Irish

Irish words written in Latin letters are spelled as in the original. The combination *th* indicates a voiceless dental spirant [θ], and *ch* is voiceless velar [x]. Vowel length is shown by an acute accent: *é*, *ó*, etc.

Kartvelian languages

Words from Kartvelian languages are transliterated as follows, with the apostrophe marking glottalization. Voiceless stops are aspirated, but this is not indicated in transliteration.

Stops and affricates:

	Voiced	Voiceless (aspirated)	Glottalized
Labial	*b*	*p*	*p'*
Dental	*d*	*t*	*t'*
Velar	*g*	*k*	*k'*
Postvelar	—	*q*	*q'*
Alveolar	*ʒ*	*c*	*c'*
Palatal	*ǯ*	*č*	*č'*

Fricatives:

Alveolar	z	s
Palatal	$\check{z}$	$\check{s}$
Velar	γ	x

Semitic languages

Words from Semitic languages are transliterated using the standard system of Latin letters with some additional marks and letters.

Laryngeals and pharyngeals: $\,{}^{,}$ ^{c}

                                        h $\d{h}$

Stops (followed by their fricative variants):

	Voiced	Voiceless	Emphatic	
Labial	$b\,/\,\underline{b}$	$p\,/\,\bar{p}$	—	$(\dot{p})$
Dental	$d\,/\,\underline{d}$	$t\,/\,\underline{t}$	$\d{d}$	$\d{t}$
Velar	$g\,/\,\bar{g}$	$k\,/\,\underline{k}$		$\d{k}$

Fricatives:

Alveolar	z	s	$\d{z}$	$\d{s}$
		$\acute{s}$		
		$\check{s}$		
Interdental	$\underline{d}$	$\underline{t}$	$\d{\underline{d}}$	$\d{\underline{t}}$
Velar	$\gamma\,(\dot{g})$	$x\,(\underline{h})$		
Labial	(v)	f		

Sonorants:	$r,\,l,\,m,\,n$
Semivowels:	$w,\,y$

Affricates:	$\check{g}$	$\check{c}$

Vowels:

Short	a	e	i	o	u
Long	$\hat{a}\,(\bar{a})$	$\hat{e}$	$\hat{\imath}\,(\bar{\imath})$	$\hat{o}$	$\hat{u}\,(\bar{u})$
Extra-short	^{a}	^{e}		^{o}	

Abbreviations

Languages and dialects

Abkh.	Abkhaz	ELith.	Eastern Lithuanian
Aeol.	Aeolic (Greek)	Engl.	English
Akk., Akkad.	Akkadian	Est.	Estonian
Alb.	Albanian	Etr.	Etruscan
Arab.	Arabic	Falisc.	Faliscan
Aram.	Aramaic	Fi-U	Finno-Ugric
Arc.-Cypr.	Arcado-Cyprian	Finn.	Finnish
Arm.	Armenian	Finno-Perm.	Finno-Permian
Assyr.	Assyrian	Finno-Volg.	Finno-Volgaic
Att.	Attic	Finno-Volg.-Perm.	Finno-Volgaic-Permian
Av., Avest.	Avestan		
Bashk.	Bashkir	Fr.	French
Boeot.	Boeotian	Gaul.	Gaulish
Bret.	Breton	Geo., Georg.	Georgian
Buddh. Sogd.	Buddhist Sogdian	Ger.	German
Bulg.	Bulgarian	Gk.	Greek
Burm.	Burmese	Gmc.	Germanic
Celt.	Celtic	Goth.	Gothic
Ch.	Chechen	Hatt.	Hattic
Chin.	Chinese	Hebr.	Hebrew (Biblical)
ChSl.	Church Slavic	Hier. Luw.	Hieroglyphic Luwian
Chuv.	Chuvash		
Cl., Class.	Classical	Hitt.	Hittite
ClMong.	Classical Mongolian	Hom.	Homeric (Greek)
Copt.	Coptic	Hung.	Hungarian
Corn.	Cornish	Hurr.	Hurrian
Cret.	Cretan	I.	Ingush
Crim. Goth.	Crimean Gothic	Icel.	Icelandic
Cypr.	Cypriot	IE	Indo-European
Cz.	Czech	Illyr.	Illyrian
Dan.	Danish	Indo-Iran.	Indo-Iranian
dial.	dialectal, dialect	Ion.	Ionian (Greek)
Dor.	Doric	Ir.	Irish
Eg., Egypt.	Egyptian	Iran.	Iranian
Elam.	Elamite	Ital.	Italic

Jap.	Japanese	O	Old (followed by
Kabard.	Kabardian		language name)
Kartv.	Kartvelian	OBret.	Old Breton
Kashub.	Kashubian	OBurm.	Old Burmese
Kirgh.	Kirghiz	OChin.	Old Chinese
Komi-Zyr.	Komi-Zyrian (Zyrian)	OCorn.	Old Cornish
Kurd.	Kurdish	OCS	Old Church Slavic
L	Low(er) (followed by	OCz.	Old Czech
	language name)	OE	Old English
Lat.	Latin	OFal.	Old Faliscan
Latv.	Latvian	OFr.	Old French
LGer.	Low German	OFris.	Old Frisian
Ligur.	Ligurian	OGeorg.	Old Georgian
Lith.	Lithuanian	OHG	Old High German
Liv.	Livonian	OHitt.	Old Hittite
LSorb.	Lower Sorbian	OIcel.	Old Icelandic
Luw.	Luwian	OIr.	Old Irish
Lyc.	Lycian	OIran.	Old Iranian
Lyd.	Lydian	OLat.	Old Latin
M	Middle (followed by	OLG	Old Low German
	language name)	OLith.	Old Lithuanian
Maced.	Macedonian (of	OMaced.	Old Macedonian
	classical times)	ONorse	Old Norse
Manich. Sogd.	Manichean Sogdian	ONorw.	Old Norwegian
MBret.	Middle Breton	OPers.	Old Persian
MCorn.	Middle Cornish	OPhryg.	Old Phrygian
MDutch	Middle Dutch	OPol.	Old Polish
ME	Middle English	OPruss.	Old Prussian
Messap.	Messapic	ORuss.	Old Russian
MGk.	Middle Greek	OSax.	Old Saxon
MHG	Middle High German	Osc.	Oscan
Ming.	Mingrelian	Oss.	Ossetic
MIr.	Middle Irish	OSwed.	Old Swedish
MLG	Middle Low German	OTib.	Old Tibetan
mod.	modern	OTurk.	Old Turkic
Mong.	Mongolian	OUkr.	Old Ukrainian
Mordv.	Mordvin	OWelsh	Old Welsh
MParth.	Middle Parthian	Pal.	Palaic
MPers.	Middle Persian	Pamph.	Pamphylian
MWelsh	Middle Welsh	Parth.	Parthian
Myc.	Mycenean Greek	Pehl.	Pehlevi
Norw.	Norwegian	Perm.	Permian

Pers.	Persian	Toch.	Tocharian
Phoen.	Phoenician	Toch. A	Tocharian A
Phryg.	Phrygian	Toch. B	Tocharian B
PIE	Proto-Indo-European	Turco-Bulg.	Bulgar Turkic
Pol.	Polish	Turk.	Turkish
Polab.	Polabian	Tyrol.	Tyrolean
Prak.	Prakrit	Ugar., Ugarit.	Ugaritic
RChSl.	Russian Church Slavic	Ukr.	Ukrainian
Rum.	Rumanian	Umbr.	Umbrian
Russ.	Russian	Ur., Urart.	Urartean
S.Arab.	South Arabian	USorb.	Upper Sorbian
Sem.	Semitic	Uzb.	Uzbek
Serbo-Cr.	Serbo-Croatian	Ved.	Vedic
Skt.	Sanskrit	Venet.	Venetic
Slav.	Slavic	VLat.	Vulgar Latin
Sogd.	Sogdian	Volg.-Perm.	Volgaic-Permian
Sum.	Sumerian	Volsc.	Volscian
Swed.	Swedish	WGk.	West Greek
Syrac.	Syracusan	WOsset.	West Ossetic
Thess.	Thessalian	WSem.	West Semitic
Thrac.	Thracian	Yazg.	Yazgulami
Tkc.	Turkic		

Grammatical terms

A	active	imper.	imperative
abl.	ablative	imperf.	imperfect
acc.	accusative case	In	inactive
Ad	addressee	inf.	infinitive
adj.	adjective	instr.	instrumental case
aor.	aorist	intrans.	intransitive
C	consonant	loc.	locative case
caus.	causative	masc.	masculine
cond.	conditional	neut.	neuter
D	voiced stop	nom.	nominative case
D	designee	NP	noun phrase
dat.	dative	O	object
du.	dual	opt.	optative
fem.	feminine	p	preverb/adposition
gen.	genitive case	p.	person (e.g. 2p. =
H	laryngeal		second person)

pass.	passive	T	voiceless stop
perf.	perfect	trans.	transitive
pers.	personal	V	vowel
pl.	plural	V	predicate, verb
ppl.	participle	VM	version marker
pres.	present	voc.	vocative case
pret.	preterite	VP	verb phrase
R	sonorant, resonant	1	first person (e.g. 1sg. = first-person singular)
S	subject		
s.th.	'something' (in English glosses)	2	second person (e.g. 2pl. = second-person plural)
sg.	singular	3	third person (e.g. 3du. = third- person dual)
S_{Intr}	subject of intransitive verb		
S_{Tr}	subject of transitive verb		

Sources

Hittite Laws	Hittite Laws (Friedrich 1959)
ABoT	*Ankara Arkeoloji Müzesinde bulunan Boğazköy Tabletleri.* Istanbul, 1948
Aeschyl. *Prom.*	Aeschylus, *Prometheus*
Al.	*Der Vertrag des Muwatalliš mit Alakšanduš von Wiluša.* J. Friedrich, *Mitteilungen der Vorderasiatisch-Aegyptischen Gesellschaft* 34:1. Leipzig, 1930
AM	*Die Annalen des Muršiliš* (Goetze 1933)
Assur	Hieroglyphic Luwian inscription from Assur (Laroche 1960:XXX)
AV	*Atharvavedasaṁhitā* (Bloomfield 1899)
Bo	Unveröffentlichte Texte aus Boğazköy. 1. Bo (followed by text number) = Funde von 1906-1912, Istanbul. 2. Bo (followed by year number and text number) = neue Funde, beginnend mit Bo 68/1, Ankara
BoSt	*Boghazköi-Studien*, Leipzig 1916-1924
BoTU	*Die Boghazköi-Texte in Umschrift.* E. Forrer, Wissenschaftliche Veröffentlichungen der Deutschen Orientgesellschaft, Leipzig, 1922, 1926
C.A.	Cippus Abellanus (Buck 1905:126ff.)
Carchemish	Hieroglyphic Luwian inscription from Carchemish (Laroche 1960:XXV-XXVI)
Cic. *Sen.*	Cicero, *Post reditum in Senatu. Cicero: The Speeches.* The Loeb Classical Library, 48-99
Festus	S. P. Festus, *De verborum significatu quae supersunt*
H.	Haδōxt Nask (Avesta)
HAB	*Die hethitisch-akkadische Bilingue des Ḫattušili I* (Labarna II). (Sommer and Falkenstein 1938)
Hatt.	*Ḫattušiliš. Der Bericht über seine Thronbesteigung nebst Paralleltexten.* A. Goetze, *Mitteilungen der Vorderasiatisch-Aegyptischen Gesellschaft* 29:3. Leipzig, 1925
Herodotus, *Hist.*	Herodotus, *Historiae*
Hesychius	Hesychius Alexandrinus, *Lexicon*
Hipponax	*Les fragments du poète Hipponax. Études et Commentaires*, 43. Ed. O. Masson, Paris 1962
HT	*Hittite Texts in the Cuneiform Character from Tablets in the British Museum.* London, 1920

IBoT	*Istanbul Arkeoloji Müzelerinde bulunan Boğazköy Tablet-leri(nden Seçme Metinler) I-III*. Istanbul, 1944, 1947, 1954
Karatepe	*Phoenician-hieroglyphic bilingual from Karatepe* (Laroche 1960:XXV)
KBo	*Keilschrifttexte aus Boghazköi I-VI*. Wissenschaftliche Veröffentlichungen der Deutschen Orientgesellschaft, 30, 36. Leipzig 1916-1923; Berlin, 1954-
KUB	*Keilschrifturkunden aus Boghazköi*. Berlin, 1921-
Madd.	*Madduwattaš*. A. Goetze, *Mitteilungen der Vorderasiatisch-Aegyptischen Gesellschaft* 32:1. Leipzig, 1928
MS	*Muršilis Sprachlähmung* (Goetze and Pedersen 1934)
Pausanias	Pausanias, *Graeciae descriptio*
Pestgebete	*Die Pestgebete des Muršiliš*. A. Goetze, *Kleinasiatische Forschungen,* 161-251 (1929)
Plaut., *Truc*.	Plautus, *Truculentes*. The Loeb Classical Library, V
Pud.	Le voeu de Puduḫepa. E. Laroche, *Revue d'Assyriologie et d'Archéologie Orientale* 43:55-78 (1949)
PY	Pylos Tablets (Bennett 1955)
RV	*R̥gvedasaṁhitā* (Aufrecht 1955)
Supr.	*Suprasl'skaja rukopis'. Pamjatniki staroslavjanskogo jazyka* 2:1, ed. S. Sever'janov. St. Petersburg, 1904
T.B.	Tabula Bantina (Buck 1905:130ff.)
T.I.	Iguvine Tables (Poultney 1959)
Tel.	Telepinus myth (KUB XVII 10; KUB XXXIII 1-10); Order of King Telepinus (BoTU 23 A, B, C = KBo III 1 + KBo XII 5 + KBo III 68 + KBo XII 7)
Tun(n).	*The Hittite Ritual of Tunnawi*. A. Goetze, American Oriental Series 14. New Haven, 1938
UKN	Urartean cuneiform inscription (Melikišvili 1960)
Ull.	*The Song of Ullikummi* (Güterbock 1951-1952)
VBoT	*Verstreute Boghazköi-Texte*, hrsg. von A. Goetze, Marburg, 1930
Vd	Vidēvdāt (Avesta)
Y	Yasna (Avesta)
Yt	Yašt (Avesta)

Sans cesse l'ineptie absolue de la terminologie courante, la nécessité de la reforme, et de montrer pour cela quelle espèce d'objet est la langue en général, vient gâter mon plaisir historique, quoique je n'aie pas de plus cher voeu que de n'avoir pas à m'occuper de la langue en général.

The absolute inadequacy of current terminology, the need to reform it and show what kind of object language is in general, constantly spoils my enjoyment of history, although I have no deeper wish than to have no need to be concerned with language in general.

<div style="text-align: right;">

Ferdinand de Saussure
Letter to Antoine Meillet
January 4, 1894

</div>

Introduction

The linguistic system and the premises of diachronic linguistics

0.1. The linguistic system of signs

Language as a system of signs distinguishes two structural planes: expression and content. Consequently, every linguistic sign unites two aspects: that of the signifier (*signifiant*, signans) and that of the signified (*signifié*, signatum), corresponding to the two planes. The relation between signifier and signified in the signs of language is determined by their interconnections in the system, on both expression and content planes. The interconnections on the two levels show the complex character of Saussure's principle of the arbitrariness of the linguistic sign. The linguistic sign is arbitrary as regards what may be called its *vertical relations*; that is, the connection of signifier and signified in any individual linguistic sign is arbitrary, unmotivated. However, the *horizontal relations* among signs, i.e. associations among signifiers and among signifieds of related linguistic signs, display elements of motivation in the sense that the nature of relations on the expression plane (similarities and differences in the phonemic composition of related signs) is motivated to some extent by the nature of systemic relations between these signs on the content plane (Gamkrelidze 1972, 1974a), as shown in Figure 1:

Figure 1

Sign A   Sign B
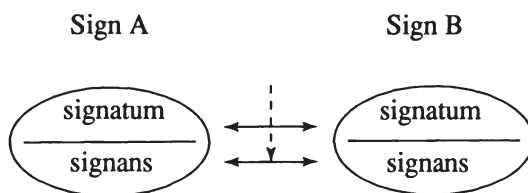
The expression plane is characterized by a complex internal structure. A linguistic sign is expressed (in spoken language) not by a single sound, but by a particular combination of sounds taken from the fixed inventory of sound units of that language. Hence the signifier of a linguistic sign has an internal hierarchical structure, which is defined as the *duality* of the sign. The distribution of

sound units among the signs of a language, the particular choice of sounds from the general inventory for the expression of a particular sign, is what constitutes the phonetic syntagmatics of language.

The fact that signs have internal structure makes it possible to create a potentially infinite set of signs using a limited set of elements, the phonemes of a language. Combining these signs according to particular structural schemes gives grammatically correct sentences in the language.[1]

A system whose signs actualize all theoretically possible combinations of elements in the system up to a particular length may be described as *saturated*. A system which does not exploit all possible combinations of elements for the expression of signs can be described as *unsaturated*, having a certain amount of redundancy. It can be regarded as a system with restrictions on possible combinations of its elements. The admissible combinations of elements in such a system define the structure of its sign expression.

Natural language is an unsaturated system, one which has a fairly high degree of redundancy, a fact which enhances its communicative effectiveness. The same fact makes possible diachronic changes in the sound structure of natural language. Changes of phonemes into other phonemes, splits or mergers of phonemes found at earlier stages, are made possible by redundancy in the linguistic system. In a hypothetical saturated system with maximal entropy, where all possible combinations of elements could yield signs, a change of one phoneme series into another would necessarily lead to a shift of all other series in the system, in order to prevent signs from falling together and avoid massive homophony and distortion of communication. In unsaturated systems with some redundancy, like those of natural languages, transformations of the phoneme system — mergers and splits — do not result in complete identity of signs, which is why these processes are possible. Hence linguistic systems are not frozen with regard to sound changes and diachronic shifts of phonemes.

0.2. The interpretation of formal and semantic similarities among signs of different languages, and the concept of related languages

When similarities of form and meaning — simultaneous similarities of both

---

1. The duality of the linguistic sign is evidently a feature characteristic only of human language and not typical of animal communication systems. Animal communication utilizes 'signs', or signals, of elementary structure, usually consisting of only a single element (Hockett 1958, Benveniste 1974, 1966a). Hence the number of signs in such a system is determined by the number of distinct elements. Since the possibility for producing such distinct elements is limited (by physical and physiological constraints), the number of possible signs is correspondingly restricted. The only possibility available to such systems is the syntagmatic combination of signs, the construction of 'sentences' made up of individual signs by combining them into longer sequences.

signified and signifier in signs — are observed between two or more languages, the question naturally arises what caused such similarities in the signs of distinct languages. Assuming that the arbitrariness of the sign is restricted (in the sense given above), formal and semantic coincidence of signs in distinct languages — phonetic similarity of two or more signs together with semantic similarity and/or identity — could be interpreted as accidental.

It is entirely possible that completely accidental combinatory factors would lead to the appearance in two or even more languages of words which were close in both sound and meaning. It is even possible to calculate with a certain degree of accuracy the probability of two or more identical or similar words of a certain length appearing in two or more languages (see Polivanov 1931:180-81, Greenberg 1957:35ff.). The plausibility of chance coincidence as an explanation for similarity will decrease as the number of languages exhibiting the coincident signs increases; and it will decrease even further as the number of coincident signs in those languages increases.

Given coincident signs and a sufficiently high number n of those signs (on the order of 20-30 or more), and a number k of languages (where $k \geq 2$), the probability of chance coincidence is virtually zero and a different hypothesis is required for explaining the similarities. (The probability of chance coincidence also decreases with increasing length of the resemblant words: the longer the words, the less the probability of their chance identity.) Another, more plausible, hypothesis for explaining such resemblances is historical contacts between the languages and borrowing from one to the other (or others) or borrowing into both from a third source.

But not all kinds of form-meaning correspondences can be interpreted as due to borrowing. There is a type of similarity between signs of different languages which is revealed in regular phonemic correspondences between resemblant signs. This type of similarity cannot generally be explained by borrowing from one language into the other. Resemblances of this type require correspondences between the languages, such that for every phoneme x of language A there is a corresponding phoneme y in a formally and semantically similar sign of language B, a phoneme z in a similar sign of language C, and so on. Where phoneme x of language A corresponds instead to phoneme y' of language B, this can be explained as positionally conditioned by the phonemic context. In such cases we can claim that phoneme x of language A corresponds to phonemes y, y' of language B in all positions. Such sound correspondences among languages are usually observed in words and morphemes referring to the basic, fundamental concepts of human activity and the human environment.

This latter type of resemblance cannot be satisfactorily explained either as due to chance coincidence or as due to borrowing (from one language into another or into both from a third source). Its only plausible explanation is a common origin for the linguistic systems in question, i.e. the descent of the

systems from a common source system, which has undergone transformations in different directions.

On this understanding, the phoneme correspondences established between attested languages can be viewed as the result of different transformations of original phoneme units. As a result of the breakup of a language L into related dialects, a phoneme X in language L yields phoneme x in one dialect, y in another, z in another, and so on; when the dialects are compared, these phonemes are seen to be in a particular correspondence to one another.

When phonemes are viewed as particular complexes of phonetic features, the transformation of an original phoneme X can be seen as the replacement of one or several features in the complex by other features, yielding various transformations of the original phonemic unit. After such transformations the regularly corresponding phonemes that arise in dialects A, B, C, ... will naturally be phonetically similar to one another; indeed, this is what establishes their historical similarity. In some cases the 'transformation' of the source phonemes consists in the preservation of all phonetic features, i.e. the entire complex. In such cases the corresponding phonemes of the dialects will display phonetic identity and reflect the feature complex of the source phoneme.[2]

Thus the 'similarity' of corresponding phonemes in related dialects can vary from complete identity to considerable difference of the phonetic features; the latter points to a change in one or several features of the original complex. In this view, the term 'similarity' or 'resemblance' must be understood as referring to the presence of regularly corresponding phonemes, and not as phonetic coincidence or difference in the signs of the languages regarded as related.

The regularly corresponding phoneme elements of related languages are generally phonetically similar to one another. Their feature complexes coincide or differ only in one or two features; it is this partial or full coincidence of features that is responsible for their phonetic similarity. It is precisely in such

---

2. Since the system of phonemic correspondences between related languages also includes elements which are identical in the two languages, there is always the possibility that two cognate forms will turn out to consist entirely of phonemes in identity correspondences, so that the forms in the two languages will be identical. In such cases there may be doubt as to whether the forms are native and cognate or rather represent loans from one language to the other or into both from a third source. Cf. e.g. Hitt. *yugan* 'yoke; season, year' and Skt. *yugám* 'yoke; season, cycle'; Hitt. *turiya-* 'harness' and Skt. *dhūr-* 'harness, gear'; Geo. *da* 'sister' and Mingr.-Laz *da* 'sister'; Geo. *ca* 'sky' and Mingr.-Laz *ca* 'sky'; Geo. *txa* 'goat' and Mingr.-Laz *txa* 'goat'; and others.

There are no strict formal criteria for unambiguously deciding this question. But the length in phonemes of the lexical items gives some grounds for determining the probability of one or the other interpretation: the longer the words, the more likely that they were borrowed, while if they consist of a small number of phonemes they are likely to have descended from a common source language and to be cognate. The probability that the words consist entirely of phonemes which happen to be identical in the daughter languages decreases as the length of the words increases. In this respect the Indo-European and Kartvelian words given just above, with their restricted phonemic length, can be regarded as lexical correspondences going back to a single protoform, and not as borrowings.

phonetically similar units that the historically observed resemblance between corresponding signs which testifies to their common provenience is most perceptible.

However, correspondence sets of related languages can also include phonemes which are quite distinct from each other, with only a small portion of shared phonetic features. The resemblance between such elements, and hence their regular correspondence, can be historically established only when we have forms in which these elements appear together with others which are minimally different and hence phonetically similar. In the absence of such forms it would be very difficult or even impossible to establish resemblance, i.e. regular correspondences, between signifiers in different languages which may go back to a common source.

Therefore in principle there may be languages which in actual fact descend from a common source, resulting from the breakup of a particular linguistic community whose membership is difficult to establish because there is no observable phonetic similarity between signs of potentially related linguistic systems. Such languages may be in what is called a *remote relationship*, the result of transformation to the point of overall replacement in the original differential features.[3]

On the other hand, assuming loss of a certain percentage of the original vocabulary over a given time interval, after a certain period of time related languages could lose all originally cognate words and affixes. In such a case it would be impossible to establish the relatedness of the languages.

The regular correspondences among phonemes in different languages which give evidence of linguistic relatedness are sometimes also observable in words known to be borrowed from one language into another. In that case, the regular correspondences can be observed when one language has two distinct subsystems S_1 and S_2, where subsystems are defined by their relation to another language (or languages). Ordinarily, such subsystems differ in the character of the words and morphemes they contain. Subsystem S_1, with phonemes reflecting common provenience, contains basic vocabulary, grammatical morphemes, and affixes, while subsystem S_2 consists primarily of cultural terms whose meanings change with cultural development. It is possible to establish a rough list of universal semantemes, found in all languages, which should primarily make up subsystem S_1 and are usually absent from subsystem S_2. In such instances the correspondences of subsystem S_1 are those pointing to common descent from a source

---

3. Establishing possible phonemic correspondences between systems which may ultimately be related but show great phonetic discrepancies amounts to establishing correspondences between groups of dissimilar phonetic units in signs with similar semantics, without regard to the phonetic resemblance or non-resemblance of those units. This can be done by sorting through a vast number of similar semantemes from a group of languages, a natural task for a computer. Establishing remote relationships in this way could become one of the research problems of computational linguistics.

system, while those of subsystem S_2 are due to regular phoneme corresponden-
ces which arise on borrowing from one language into the other.

In such cases we usually observe several correspondence subsystems, one for
each language, pointing to borrowing from those languages at different times,
whereas subsystem S_1 exhibits a single system of correspondences with all the
other languages, and each of them has the same types of phonemic correspon-
dences with the others. Subsystems of type S_2 generally contain fewer forms
than the basic system S_1, which marks S_2 as borrowed. However, in particular
languages the total set of subsystems of the S_2 type may contain more words
than the basic subsystem S_1, although each subsystem of the S_2 type usually
contains fewer elements than the S_1 set. The lexemes of S_2 may reflect bor-
rowing from different languages or from one language at different times.[4]

By identifying such lexical groups in a language and establishing their
phonemic correspondences to other languages we can separate the lexicon into a
number of subsystems, one of which is the basic vocabulary set which manifests
the primary phoneme correspondences with other languages and reflects their
common origin. The large portion of the lexicon that remains consists essen-
tially of words that entered the language later, after its removal from the com-
mon source language, as a result of contacts and interaction with other languages
throughout the course of its subsequent development.

In some cases when languages are in constant and intensive contact leading to
a lengthy period of bilingualism, there can be borrowings of a significant
portion of the lexicon of one language into the other, and regular correspon-
dences arise between words in sentences and morphemes within words. This
leads to structural convergence and structural resemblances between the lan-
guages, which can be characterized as *secondary kinship* or *acquired kinship*
(the 'allogenetic relations' between languages of Cereteli 1968). Allogenetic
relations between languages can obviously arise due to convergence in the
generative rules of languages in contact and the rise of a common unified
generative system reflecting the two original contacting systems. When this
kind of unification of generative systems occurs, there can be greater typologi-
cal similarity between the languages in contact than between either of them and
its genetic source.

In addition to allogenetic kinship, there also exist *linguistic areas*, groups of
unrelated languages or of related languages which separated from their source,
underwent a long period of separate development, and then were reunited by
contact in a common territory. Examples are the Indo-European languages
belonging to the Balkan linguistic area, the Lithuanian-Polish-Belorussian

---

4. Subsystems for loans from related and unrelated languages can be distinguished. The
subsystem of loans from a related language will itself contain a subsystem S_1 which reflects the
common heritage. The subsystem of loans from an unrelated language will lack an S_1 sub-
system.

frontier, etc. (Jakobson 1971a). Three kinds of linguistic areas can be distinguished: unions of related languages (e.g. the Scandinavian–Old English contacts), of unrelated languages (e.g. Uzbek and Tajik: see Polivanov 1968), and a mixed type involving both related and unrelated languages (e.g. the Armenian-Ossetic-Kartvelian interaction).

0.3. The common linguistic system and the means of reconstructing it. Typological verification of reconstructed models

When phonemic correspondences between languages are explained as due to their common descent from a source language, the source language must be reconstructed if the rise and transformations of the attested daughter systems are to be studied. Comparison of languages centered on establishing regular phonemic correspondences leads logically to the reconstruction of a linguistic model whose transformation in various directions has produced the attested daughter systems. Comparison of languages which does not have as its goal the reconstruction of a source system can be regarded only as a preliminary stage in the investigation of linguistic history (cf. Saussure 1915:299).

The history of related languages prior to their attestation in writing can be established only if all the diversity of the attested structures can be reduced to a common source model and the paths by which these systems arose and developed can be reconstructed, from the original stage to the attested ones. This approach to genetic comparisons and correspondences naturally raises the question of how the original system is to be reconstructed and by what linguistic methods (see Birnbaum 1977). It is assumed that the original linguistic model reconstructed by means of the appropriate linguistic method will be a first approximation to a linguistic system which actually existed in space and time and which broke up into attested daughter dialects. Comparison of these dialects makes it possible to postulate particular structural models which reflect the hypothetical ancestral linguistic system. Especially significant for such comparisons is the method of internal reconstruction, the reduction of alternating elements within one linguistic system to a single source.

Every level of language can be regarded as an aggregate of interconnected subsystems. On the lexical level, as mentioned above, semantic features can distinguish groups of words containing basic vocabulary for fundamental concepts common to all cultures: formally, these words are characterized by the fact that they exhibit formal correspondences to related languages; within the overall system such words are isolated to some extent and may display phonological and morphophonological features which distinguish them from the rest of the vocabulary. These features may be regarded as archaic, and they serve as the basis for internal reconstruction, that is, for reconstructing the period in the history of the language when the features in question were not an

anomaly but the norm and reflected productive processes which are diachroni-
cally explained as innovations. Delimiting an archaic subsystem of this sort is
one of the basic prerequisites for comparing the language with other languages.

When reconstructing the original linguistic models which must reflect the
system of the common source language, the linguist is faced with the
methodological question of the reality of reconstructions and the extent to which
they correspond to the actual language that existed in space and time and is taken
as the source for the group of related dialects. If the reconstructions are taken
to be real, this entails a number of methodological principles for comparative-
genetic linguistic research. The foremost of these is the close connection of
comparative-historical research with principles of linguistic typology and
universals (or frequentalia: see Serebrennikov 1974). In this respect genetic, or
comparative-historical, linguistics, the linguistics that establishes genetic rela-
tions among groups of languages and provides reconstructions of their source
models, in principle forms a single discipline with structural typology and
linguistic universals.

And in fact reconstructed protolanguages, if they claim to reflect actual
languages that existed in space and time, must be fully consistent with typologi-
cally inferred regularities of language, regularities established inductively or
deductively through comparison of many different linguistic structures. A
linguistic reconstruction which is in conflict with linguistic universals of course
cannot claim to offer reality in its reflection of a linguistic system that once
existed.[5] But even when reconstructions are consistent with synchronic lan-
guage universals, this cannot serve as sufficient evidence that they are real and
reflect an actual linguistic system proposed as a source for related languages. A
necessary condition for reality of reconstructions is that they must be consistent
with diachronic typological data, with schemas for the change of particular
linguistic structures over time, as established by the study of historical facts

---

5. Linguistic typology is not only a means for checking the plausibility of reconstructions,
but also often serves as grounds for positing plausible but unattested links in linguistic structure.
For example, in some cases it is possible to reconstruct only some of the distinctive features (for
instance, only the feature of syllabicity for vowel phonemes). But reconstructing the complete set
of distinctive features is impossible unless we go outside of comparative-historical linguistics and
bring in typological facts which make it possible to fill in some of the unreconstructible features.
The laryngeals, posited by Saussure on the basis of morphophonological patterning, can serve as
a good illustration of restricted reconstructibility of distinctive features of phonemes in the
absence of typological data. Strictly speaking, the claim that Saussure reconstructed only an
'algebraic system' is untrue. He reconstructed the most essential features of the laryngeals,
namely their syllabicity and nonsyllabicity (i.e. their sonant character) and aspects of their
influence on adjacent vowels. Subsequent research into the 'laryngeal theory' amounted to filling
out the set of features rather than establishing any qualitatively new constructs. The schematicity
of Saussure's conclusions can be explained by the incomplete set of differential features in his
reconstructed phonemes, which is often inevitable for exclusively internal reconstruction that
takes no account of typological facts. Further specification of the nature and phonemic status of
the laryngeals in the system became possible only with broader application of facts from
typological comparison of languages.

from individual languages.6

Thus reconstructions can be considered real if they are consistent with two basic typological criteria: they must agree with synchronic typological universals and they must agree with diachronic typological universals (general schemas for change and transformation of languages). These two criteria may be regarded as necessary and sufficient conditions for the reality of a reconstruction, which can then be seen as reflecting a parent linguistic system which once existed in space and time. Typological verification, synchronic and diachronic, of linguistic reconstructions thus becomes one of the basic prerequisites for positing linguistic source structures, and an indispensable one for testing their plausibility.

Formulation of regular rules for deriving attested daughter dialects from a posited source system may be regarded as a way of describing the rise and transformation of the dialects, from their original common state to their historical attestations. In contrast, reconstruction of a linguistic protosystem is achieved by comparing attested related linguistic systems and moving backwards in time from later to earlier linguistic stages and typologically verifying each system. This process continues until we reach a linguistic stage from which all the attested related systems can be derived by means of a set of typologically plausible regular transformations; these transformations are what determine the *diachronic derivability* of the system. They lead from the original system to the later stages which are the result of structural transformations of the original one.

In their explanatory power, diachronic transformations — which derive attested linguistic forms from theoretical constructs which can be regarded as chronologically earlier stages of these forms, their archetypes — can be compared with the transformations of generative grammar, which derive observed surface elements from theoretically posited underlying forms which define the deep structure of language.

A description of diachronic changes by means of transformational rules is essentially a regular sequence of discrete steps each of which reflects one of the synchronic stages in the development of a language. The less the chronological distance between such steps, the more precise and adequate the description of the language's development and the account of its successive changes from the original system.

---

6. A diachronic typology of possible transformations at various levels should be primarily based on attested material. Historically attested changes are the domain of historical grammar and therefore no comparative-historical or diachronic grammar can be constructed without taking into account the data arrived at by historical grammar.

Historical grammar is based primarily on material attested in documents. This material, which comes from earlier stages of language, has to be rendered in a special phonetic-phonological transcription, which raises technical questions about the relation of writing to the phonemic inventory of a language and about how phonemes and phoneme sequences are reflected in various writing systems. Therefore, study of the relation of writing to language is an indispensable prerequisite for constructing a theory of diachronic linguistics (cf. Hoenigswald 1960).

In this sense a rule deriving an attested stage from a reconstructed one can be broken down into a series of successive rules which derive the final result of the transformation from a series of regular steps that reflect possible stages in the development from the original linguistic system to the later attested one. For example, a phonemic shift of x to y, where x is the postulated stage and y the attested one, can be represented as a series of successive transformations:

$$
\begin{aligned}
x &\longrightarrow x' \\
x' &\longrightarrow x'' \\
&\cdots \\
x^{n-1} &\longrightarrow x^n \\
x^n &\longrightarrow y
\end{aligned}
$$

For example, a change of p to $\emptyset$ can be represented as a series of changes in a feature bundle, one feature at a time:

$$
\begin{aligned}
p &\longrightarrow p^h \\
p^h &\longrightarrow f \\
f &\longrightarrow h \\
h &\longrightarrow \emptyset
\end{aligned}
$$

Jakobson has noted the importance of making use of synchronic typology in linguistic reconstructions (1957a, 1963; cf. also Hjelmslev 1928), and this typological approach requires radical reconsideration of classical Indo-European comparative-historical grammar and a new interpretation of the linguistic correlations sought by that method, with an eye to the structural-typological verification and diachronic derivability of the system.

At the present stage in the development of linguistics, when one of the major concerns is structural typology and language universals, much of what is found in the traditional reconstructions of Indo-European can clearly be revised to bring the posited Indo-European protolanguage into agreement with the facts of typological linguistics. The adjustment of traditional Indo-European reconstructions to typologically plausible systems may entail a major reconsideration of the reconstructions.

Classical comparative-historical Indo-European linguistics was one-sided and restricted, since its reconstruction of Proto-Indo-European was based only on external comparison of the separate daughter systems, complemented in some theories by internal reconstructions based on relations within one system. No explicit attention was given to the linguistic plausibility of the model, in the sense of its typological consistency with possible linguistic structures. As a result, classical Indo-European linguistics postulated a source linguistic system which, since it contradicts synchronic typological facts, cannot be considered linguistically plausible.

0.4. Reconstruction of sound units as bundles or combinations of distinctive and phonetic features, and their hierarchical correlations in the system

When a phonemic inventory is reconstructed, the sound units posited for the language must be regarded as combinations of distinctive features which can be verified against linguistic typological facts. There are universal models for the combinability of features in a vertical sequence into simultaneous bundles. Some feature combinations are preferred over others, as is reflected in the high systemic and textual frequency of phonemes containing such combinations; other features have more restricted combinability, and this is reflected in the lower frequency of phonemes containing them. The latter include empty cells, or gaps in a paradigmatic system, which can be regarded as exemplifying difficult feature combinations.

Two basic types of feature combination can be distinguished in this connection: marked and unmarked. A marked combination is an unusual or rare one, as is shown by the low frequency and restricted distribution of a phoneme containing it. It may be entirely lacking in some linguistic systems, resulting in gaps in phonological patterns. A gap in a paradigmatic system can be regarded as a cell having the possibility of being filled by a phoneme with null frequency.

An unmarked combination is a normal and natural one, as is shown by the higher frequency and greater combinatory freedom of phonemes containing such combinations. Natural combinability of features is due to their articulatory and acoustic freedom to cooccur in a simultaneous (vertical) sequence. The result is a functionally stronger phoneme.

Thus the distinction of markedness and unmarkedness pertains not to the individual feature in combination with other features; rather, it is a property of the entire feature combination as a whole.[7] The feature values of marked (m)

7. An understanding of markedness which ascribes hierarchical dependency not to individual distinctive features but to feature combinations comes close to the traditional understanding of Prague School theory, where markedness and unmarkedness are defined in relation to unitary phonemes, which are viewed as bundles of distinctive features. There is a principled difference, however, in that our view interprets markedness not as the presence or absence of the feature in a phoneme, but rather as a hierarchical relation determining the degree of usualness, naturalness, and cross-linguistic frequency of particular feature combinations. Markedness in this sense is applicable to all types of phonological oppositions, not only privative but also gradual and equipollent (in terms of Trubetzkoy's classification). Hence the terms *marked* and *unmarked* depart from their original etymological meanings (*merkmalhaltig* 'having [a] feature', *merkmallos* 'featureless') to acquire a new sense of unusual vs. usual feature combinations.

Functionally strong, stable feature bundles (or phonemes), usually defined as unmarked in opposition to marked, functionally weak, unstable feature bundles or phonemes, can be renamed *dominant* bundles (contrasting with *recessive*). This reformulation in terms of paradigmatic dominance and recessiveness is desirable in view of the polysemy of the traditional terms *marked* and *unmarked*, which continue to be used in their original meanings (Gamkrelidze 1979). The terms *dominant* and *recessive* are taken from contemporary molecular biology, a field which, as is known, makes considerable use of linguistic terms for describing the structure of the

and unmarked (*u*) pertain not to the individual feature F, yielding [*mF*] and [*uF*], but to the whole feature bundle (Gamkrelidze 1975:23):

$$
m \begin{bmatrix} \alpha F_1 \\ \alpha F_2 \\ \vdots \\ \alpha F_k \end{bmatrix} \quad \text{or} \quad u \begin{bmatrix} \alpha F_1 \\ \alpha F_2 \\ \vdots \\ \alpha F_l \end{bmatrix}
$$

Such restrictions on combinability of features are based on properties of the human articulatory apparatus and on the psychological and physiological possibilities of speech perception. Taking these characteristics of speech into consideration, we can determine which feature combinations are articulatorily and acoustically optimal (phonologically, these will be the unmarked combinations) and which are non-optimal (phonologically marked): see Greenberg 1966, Postal 1968:170ff.

Hierarchical dependencies obtain between individual phonological units — feature bundles — and are revealed in dominance relations; this shows that there is a strict stratification of phonological values in the linguistic system.

We also find diachronic phonemic transformations in language, in accord with these universally significant correlations. A number of diachronic phonological changes which at first glance seem unconnected can be seen as interdependent and mutually conditioned, determined by the hierarchy of phonological values.[8]

A phoneme, which is a bundle of differential features, is realized in speech as a sound unit whose inventory of phonetic features is usually larger than the set of distinctive features it carries as a phoneme. The distinctive feature bundle becomes more complex in its phonetic realization, where it acquires additional phonetic properties. This is one of the differences in the expression plane between *language* and *speech* in the sense of Saussure and Trubetzkoy. In a strictly phonological transcription, which reflects only the phonologically relevant features, these additional phonetic properties should not be shown. However, their significance in the realization of phonemes in speech, as well as in sound changes, makes it imperative that they be taken into account in any linguistic description.

Two types of phonetic features must be distinguished: combinatorily conditioned ones, which can be deduced from the phonological features of the

---

genetic code (see Jacob 1977).

8. A universal hierarchy of phonological units implies, as noted above, the possibility that there will be phonemes with low or zero frequency (empty cells). Such systemic paradigmatic regularities should always be borne in mind, for both synchronic and diachronic description and in particular for reconstructions. An empty cell is not an anomaly from the point of view of markedness theory, and consequently the absence of external comparative data does not require that it be filled in internal reconstructions, as is often done in diachronic linguistics (for instances involving filling of *cases vides* see Martinet 1955, 1960).

context, and non-combinatory ones, which are imposed on the feature bundle in the phonetic realization of the phoneme. As an example of a non-combinatory phonetic feature, voiceless stops in many languages (Semitic, Kartvelian, and others) are realized as aspirated. Aspiration is a phonologically non-distinctive feature but plays an essential role as a phonetic feature in the realization of the voiceless stops.

Synchronically, the phonetic features are relevant not for the system of generative rules of the language, but for its realization in speech (see Pilch 1968:57ff.). Sound changes obviously involve interaction of phonological and phonetic features. Phonologically superfluous features can determine the direction in which phonemes change. This is why it is important to take account of phonetic features in describing diachronic phonological changes.

In diachronic as well as synchronic phonemic transcription, then, the phonetic features which are essential to realization and to determination of the direction of diachronic phonemic changes should be reflected insofar as possible. It is expedient to use a combined feature transcription which utilizes both phonemic differential features (set in Roman type) and the accompanying phonetic features (italicized). For instance, in languages where aspiration is a secondary phonetic feature, /p/ can be represented as the following feature complex:

$$
\left\{
\begin{array}{l}
[\mathit{aspirated}] \\
\left[
\begin{array}{l}
- \text{ syllabic} \\
+ \text{ nonsyllabic} \\
+ \text{ stop} \\
+ \text{ voiceless} \\
\quad \text{labial}
\end{array}
\right]
\end{array}
\right\}
$$

In place of this full transcription it is simpler and more convenient to make use of a shortened form, using a single symbol rather than a feature complex, with the additional phonetic feature as a superscript: p^h instead of the above feature complex. Similarly, it is simpler to express phonemic changes as changes in only the features that differ, without repeating the features that do not change. (The same obviously applies to the synchronic representation of similar phonemes, in which shared features need not be repeated.)

When we have full feature matrices where all phonemes are described exhaustively in terms of differential features, their diachronic changes and synchronic oppositions can be expressed in abbreviated form with a single symbol, and the changes of features can be derived from the phoneme matrices. This is partly similar to classical comparative practice, except that the latter did not differentiate between phonological changes of phonemes and phonetic

changes of individual sounds.9

On the other hand, where we have identical, symmetrical shifts of whole phoneme series, it may be more economical to rely on a transcription using changes of one or more features rather than writing an individual abbreviated symbol for each phoneme taking part in the change (see Halle 1962). For instance, the following shift of a series

$$/b^h/ \;\rightarrow\; /b/$$
$$/d^h/ \;\rightarrow\; /d/$$
$$/g^h/ \;\rightarrow\; /g/$$

can be more economically represented with a single feature rewrite rule:

$$[+ \text{aspiration}] \rightarrow [- \text{aspiration}] \; / \left[\begin{array}{c} \overline{\phantom{+\text{stop}}} \\ +\text{stop} \\ +\text{voice} \end{array} \right]$$

That is, the feature of aspiration with a plus value receives a minus value in the paradigmatic context of the features of stop and voicing.

Defining the phoneme as a feature bundle brings us to the question of what the distinctive features that represent the phoneme are. Phoneme representations in terms of primarily articulatory features have recently replaced the acoustic features that were more popular during the early rapid growth of spectrographic analysis. A strictly acoustic definition of distinctive features often led to an inadequate account of the articulatory properties of the phonemes. The importance of articulatory properties, not only for the formation of phonemes but also for their perception, is supported by a growing number of recent works (see Čistovič 1961, 1965, Liberman 1957, Ladefoged 1971, and others). Therefore it makes sense to select as distinctive features those articulatory features that define the essential properties of a phoneme, including acoustic features as necessary, provided the articulatory and acoustic features do not exclude each other. A feature inventory that does take account of articulation is preferable to an exclusively acoustic one.

Below, distinctive features will be described primarily in articulatory terms and will be based on features proposed in the literature, with some modifications (e.g. Jakobson, Fant, and Halle 1962, Jakobson 1971a, Chomsky and Halle 1968:93ff.).10

---

9. For reconstructed stages these phoneme matrices can be established by comparing typological data to the results of a distributional analysis of the phonemes in a synchronic slice of the reconstructed stage.

10. Instead of the features [± vocalic], [± consonantal], which are tautological when vocalic and consonantal phonemes are defined, [± syllabic], [∓ nonsyllabic] are used, based on the ability of segments to form syllable peaks, i.e. to function in a sound sequence as syllable-forming (central) or non-syllable-forming (marginal). Vowels will therefore be described as [+ syllabic, – nonsyllabic], consonants as [– syllabic, + nonsyllabic]; sonorants (sonants), which

0.5. Principles of semantic reconstruction

Reconstructing a language means reconstructing not only its phonemes and their correlations in a paradigmatic system, but also whole sequences of phonemes and combinations of them that yield morphemes, words, and phrases expressing particular grammatical or lexical meanings. It means reconstructing not only the expression plane but also the content plane, the meanings of forms and syntactic constructions.

While in formal reconstruction the investigator starts with a system of phoneme correspondences and postulates typologically verifiable source forms and archetypes, semantic reconstruction is complicated by a lack of formal criteria for positing source semantemes, i.e. for positing a content plane for the reconstructed protoforms — words and word combinations. The simplest case, as with formal reconstruction, is when formally corresponding words are close in meaning. Then the meaning common to those words — the set-theoretical intersection of their semantic differential features — can be reconstructed. When there are conversive relations, as with 'give' and 'take', 'buy' and 'sell', the reconstruction posits a semanteme in which the opposed features are neutralized. The conversive pair 'buy' and 'sell' can be reduced to a generalized term for exchange, comprising both buying and selling. This is analogous to the concept of neutralization in phonology and morphology. Thus what is reconstructed is not the concrete meaning of an individual word, but a situation in terms of which the meaning can be described (see Benveniste 1966a, 1974); compare contemporary works on linguistic semantics, where the meanings of words are described by means of meaning-preserving transformations of whole sentences.

A typology of semantic changes acquires particular significance for such reconstructions. Posited changes increase in probability as similar semantic changes are observed in the histories of individual languages or language groups.

0.6. The reconstructed linguistic system in space and time

A reconstructed linguistic model reflects a protolinguistic system which once existed, and the time frame within which it existed and changed must be reconstructed, as must the geographical aspects of its spread. If a protolanguage is regarded as a system which existed in space and time and had a history, then the dynamics of its evolution must be studied and account must be taken of its earliest scientifically recoverable stages and its history up to its breakup into

could be syllabic or not depending on context, will be defined as [+ syllabic, + nonsyllabic]: see Gamkrelidze 1975.

daughter dialects and their formation as independent linguistic units.

Many structural properties of Proto-Indo-European which are reconstructed in classical Indo-European studies as static schemas can be broken down into chronological stages. Features which are reconstructed for an undifferentiated source system often do not belong to the late period of its development and dialectal differentiation but reflect an earlier stage. This explains the frequent debates in Indo-European comparative grammar over linguistic structures that appear mutually exclusive; an example is the discussion of the number of laryngeals in Indo-European, where each of the several incompatible solutions has good evidence in its favor. In such instances the various solutions can be associated with different developmental stages of PIE, which permits us to regard many of the proposed structures as chronologically complementary and datable to different stages.

The motionless, static PIE scheme must be replaced with a chronologically dynamic system, one which, like any attested language, had its history and evolutionary dynamics. That history presupposes both internal evolution of the system and areal associations with other systems, reflected in contacts and interference. In this respect we can speak of linguistic borrowings into PIE from other languages and from PIE into other languages which were in contact with it.

Like any real language, the system of a reconstructed language must be understood to have existed as a set of interconnected dialects. Its breakup into historically attested cognate languages can be seen as the gradual differentiation and fragmentation of what were originally dialects of a common source linguistic system. Within the reconstructed system, therefore, individual subsystems can exhibit differences in phonological and morphological features which reflect dialect differentiation of the reconstructed language. In a model reconstructed for a particular chronological period, dialect differentiation takes the form of variant, or doublet, forms; these can reflect areal, dialectal oppositions within the common system at a late stage. They can in turn be explained chronologically in terms of the history of the common language (see Stepanov 1979). On this view, the dialect differentiation and breakup can be seen as a doubling of original structures which subsequently develop in different directions.

Any element x of the original system subsequently yields elements a and b in two related languages. This can be schematically represented in the form of a tree (an oriented graph):

Figure 2
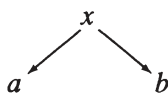
This is the result of a split of the original element into two dialect forms and the subsequent transformation of each of the dialect forms into the resultant attested units:

Figure 3

$$x \implies x \quad | \quad x$$
$$\downarrow \qquad \downarrow$$
$$a \qquad b$$
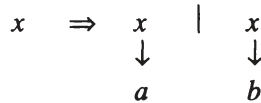
This is illustrative of the development not only of individual elements of the system, but also of the entire system.

The genealogical tree which represents the descent of individual languages from a protolanguage in traditional comparative linguistics is in essence a model reflecting the final results of these transformations of the original system. It actually reflects relations among units already formed, and indicates only the direction of evolution from the original system. In this respect the genealogical tree is not in contradiction to the wave theory, which models the rise and spread of innovations in language but does not show the final results of differentiation into separate dialects. Therefore there is no reason to oppose these two schemas and judge their relative merits and shortcomings. Each is incomplete, since they only reflect different aspects of the rise and development of linguistic structures from a source. They should therefore be united as complementary schemas for diachronic change of languages.

Insofar as comparative-historical, or diachronic, linguistics relies on a concept of the source linguistic system as a union of distinct dialects which existed in space and time, it is linked to the theory of formal relations between languages — structural typology and language universals. It also has some affinity in theory and methods to the theory of linguistic relations in space — linguistic geography, areal linguistics, language contacts.[11]

---

11. Diachronic linguistics has interesting links with structural dialectology in the sense of Weinreich 1954. Weinreich's *diasystem* is precisely the result of reconstructing a common system for dialects. On this view, comparative dialectology coincides with comparative-historical linguistics in its approach to the tasks of diachronic reconstruction. The diasystem of structural dialectology is more a diachronic and/or metalinguistic concept than a synchronic one. Therefore it cannot be regarded as a real communicative system for speakers of the dialects. What makes it possible for speakers of different dialects to understand each other is not mastery of a diasystem but knowledge of the rules for dialect switching — rules requiring that correspondences between the dialects be viewed as existing on several levels. However speakers of another dialect may assess these correspondences, they consider them to be distortions vis-a-vis their own dialect. It is knowledge of these correspondences that constitutes the rules for dialect switching.

To some extent this criterion can serve as grounds for distinguishing dialects from languages. Speakers of related languages, unlike speakers of related dialects, evidently do not know the correspondences between the languages, and hence they have no rules for language shifting. Ordinarily they regard the languages as unconnected independent systems.

0.7. The original territory of the common language and the migratory routes of speakers of its dialects. The problem of identifying linguistically reconstructed cultures with archeologically reconstructed ones

A group of related languages is formed when an original linguistic system disintegrates due to disruption of contacts between speakers of individual dialects; the languages are spread to their historically attested territories by migrations of the speakers. This means that the original range of the common source linguistic system lay in a particular area, an area more compact than the range of the daughter languages, from which the out-migrations originated. The size of the territory and the gradual dissolution of the community both depend on the culture, the geography, and the ecological conditions of the tribes that spoke the common language. Thus identifying the original range of a reconstructed language and identifying the migratory routes of the tribes speaking its dialects are the historical and geographic side of the specifically linguistic problem of dissolution of linguistic unity.

That the original territory was more compact than the range of the daughter languages is confirmed by typological facts which show historically attested spreading of related languages and settlement of larger territories by their speakers. Within the larger range we can distinguish a more compact area settled by speakers of archaic dialects which are linguistically closest to the original system (see Bartoli 1925, Sapir 1958, Ivanov 1958, Schlerath 1981).

The Indo-European languages, presently spread over a broad range in Eurasia, must have originated in a more compact original Proto-Indo-European territory. Identification of that territory is the essence of the traditional problem of the Indo-European proto-homeland. It will make it possible to reconstruct a picture of the migrations of the ancient Indo-Europeans, the tribes speaking Indo-European dialects, to their historical territories, and thus to establish the dynamics of areal associations among the Indo-European daughter languages.

Determining the homeland requires, first of all, that we establish an absolute chronology, at least an approximate one, for the migrations. This can be done by demonstrating affinities between the linguistically reconstructed culture of the proto-speakers and particular archeological cultures. This leads to the question of how one correlates linguistic data with archeological and historical cultural data. Specifically, it raises the question of whether a particular linguistic community can be correlated with a particular archeological culture. Presumably a linguistic system can be correlated with a culture if the culture yields the same realia as the language does: the realia are identical in their elementary composition and the interconnections among the elements are identical.

Fulfilling this task requires working out a typology of archeological cultures, with implicational relations between individual units of material culture.[12] If the implicational rules for a culture coincide with those for the culture reconstructed from linguistic data, then the identification of the two cultures becomes more plausible. Certain elements and features of the cultures may fail to coincide, but the essential requirement is that whole complexes coincide and there are no incompatible traits.

Since the reconstructions — both linguistic and historical — are necessarily incomplete, failure of individual components to coincide is no obstacle to their historical identification. With limited and incomplete reconstructions, the size of the one complex (A, reconstructed linguistically) can be either greater or less than that of the other (B, established archeologically). We can also have intersection of the complexes, with a larger or smaller portion of either the archeological culture B′ or the linguistic culture A′ outside the common part AB, as shown in Figure 4. It is essential that the non-coinciding parts A′ and B′ not contain features which are incompatible in view of implications internal to the two complexes (as for instance if an element of A entailed that the entire reconstructed system A contain some feature which is incompatible with a feature which either is reconstructed for part B or is implied by other elements in B).

Figure 4

A′ ——— (A B) ——— B′
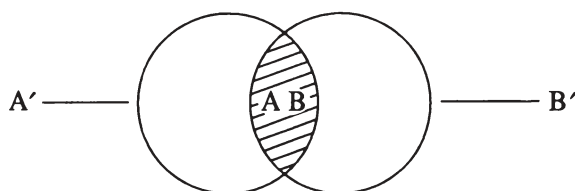
When the question of the original territory of the common language and the identification of its linguistically reconstructed culture with an archeological culture is posed in this way, the task of the linguist is to provide a systematic semantic analysis of all reconstructed words and phrases which point to individual features of material and intellectual culture that can typologically

12. Similarly, if a typology of archeological cultures is to be established there must be agreement on a unified description and inventory of cultural artifacts (see e.g. Gardin 1965, 1983, Kameneckij et al. 1975). A unified inventory is a necessary precondition for typological comparison of different cultures.

distinguish this culture from others. That analysis requires that the reconstructed words and phrases be linked to denotata and that the cultural-ecological and historical-geographical characteristics of those denotata be determined.

This section has proposed principles for analyzing language and the phenomena of its speakers' intellectual and material life which are connected with and reflected in language. These principles will be used below to set forth the system of the Proto-Indo-European language (Part I) and the Proto-Indo-European culture (Part II), their interconnection, and their typological connections with languages and cultures of historically adjacent areas.

Part One

The Structure of Proto-Indo-European

Section One

**The Phonological
System and
Morphophonology of
Proto-Indo-European**

Si le seul moyen de reconstruire est de comparer, réciproquement la com-
paraison n' a pas d' autre but que d' être une reconstruction. Sous peine d' être
stériles, les correspondances constatées entre plusieurs formes doivent être
placées dans la perspective du temps et aboutir au rétablissement d' une forme
unique.

The sole means of reconstructing is by comparing, and the only aim of
comparison is a reconstruction. Our procedure is sterile unless we view the
relations of several forms from the perspective of time and succeed in re-
establishing a single form.

> Ferdinand de Saussure, *Cours de linguistique générale*
> (transl. Wade Baskin)

Typological verification raises the probability of reconstructed phonemic and
morphological patterns, and permits changing the reconstruction from a mere
numerical catalogue into a more realistic portrayal of the linguistic system.

> Roman Jakobson, *Typological studies and their*
> *contribution to historical comparative linguistics*

Chapter One

The three Indo-European stop series: Paradigmatics and syntagmatics

1.1. The three stop series in Indo-European and the problem of the defective labial inventory

1.1.1. The traditional system of Proto-Indo-European stops

For late Proto-Indo-European, the stop system is traditionally reconstructed with three manners of articulation, or *series*,[1] and four points of articulation, or *rows*.[2] The series comprise phonemes which are homogeneous but heterorganic, and the rows are homorganic but heterogeneous (Martinet 1955:III.8). The three stop series are traditionally characterized as voiced (*mediae*), voiced aspirate (*mediae aspiratae*), and voiceless (*tenues*), and the four points of articulation as labial, dental, velar, and labiovelar (see Table 1).

Table 1
The traditional system of Indo-European stops

I	II	III
(b)	*bh*	*p*
d	*dh*	*t*
g	*gh*	*k*
gʷ	*gʷh*	*kʷ*

1. This refers to the most widespread view of PIE consonantism, as set forth, for example, in Lehmann 1952. Earlier works reconstructed a fourth series of voiceless aspirates, but most investigators beginning with Saussure (1892) have rejected the fourth series as a late development (see Pedersen 1951).

2. The full system of Indo-European stop points of articulation will be discussed below in a separate section dealing with the problem of whether Indo-European had palatovelar and labiovelar stops as well as plain velars. The question of how many phonemically distinct dorsal points of articulation there were is an essential one in reconstructing the full system of Indo-European stops and establishing the transformations it went through in the formation of the daughter languages, but it is irrelevant to determining the phonological nature of the three manners of articulation. The features that distinguish the three stop series are not directly relevant to the phonology of the dorsal phonemes. For an investigation of the three stop series, it is sufficient to have information for only three sets — labial, dental, and velar — without regard to the further subdivision of the velar set. Therefore, in the phonological transcription used below (up to the section of Chapter 2 dealing with the number of dorsal sets) a single symbol will designate palatovelar and plain velar consonants indifferently (the traditional transcription is g vs. $\hat{g}$, g^h vs. $\hat{g}^h$, k vs. $\hat{k}$).

In the traditionally reconstructed system there is an asymmetry in the labial series. It was pointed out by Pedersen (1951) that the voiced bilabial *b is absent, while there are many examples with *g and *d. Pedersen rejects as dubious all the cognate sets usually adduced in support of PIE *b. Specifically, he shows that of three forms usually regarded as cognate, Skt. *bálam* 'strength', Gk. *beltī́ōn* 'better', OCS *bolĭjĭ* 'more, to a greater extent', only the Sanskrit and Slavic forms can actually be cognate, and only if Skt. *bálam* is not a Dravidian borrowing (Burrow [1955]1976:358, 360, 374). There is an almost complete absence of forms which uncontroversially go back to *b (cf. Hamp 1954:40).

Non-initial PIE *b has so far been found (Szemerényi 1970:51, §7.1.2) only in two forms. One is Goth. *diups* 'deep', Lith. *dubùs* 'deep, hollow', OCS *dŭbrĭ* 'gorge, ravine, canyon, glen', Gaul. *Dubno-rīx* 'King of the world', OIr. *domun* 'world'. This correspondence set is attested in only one Indo-European dialect group, a western one, and the only evidence for its manner of articulation comes from Germanic. It could be that the voiceless *p of Gothic and other Germanic languages is the result of positional devoicing of *b* from *bh. It is significant that the very similar Gk. *buthós* 'depth, gulf, abyss' points to a doublet set **budh- / *bhudh-**, probably with metathesis of the original voiced consonants: cf. Gk. *puthmḗn* 'depth' (Chantraine 1968:I.201). These forms do not provide indisputable evidence for reconstructing a non-initial PIE *b (see Trubačev 1978:5.175).

The other possible cognate set showing *b is Icel., Norw. *slapa* 'limp, slack, drooping', OCS *slabŭ* 'weak', Lith. *slobstù, slōbti* 'become weak', usually compared to Lat. *labor, lapsus* 'slip, slide' (Pokorny 1959:655). Here again we have an areally restricted correspondence, pointing to a late origin for these forms, which are therefore not Proto-Indo-European.

However, even if we admit a PIE *b in the forms just examined, we cannot fail to note the striking quantitative discrepancy between forms with PIE *b and those with *d and *g. Based on a count we made of Pokorny, *d and *g occur over 250 times each in Proto-Indo-European words.[3]

The claim that IE *b was a 'weak' phoneme has recently been argued against by Djahukian 1982. Djahukian does not mention Pedersen 1951, which deals precisely with this topic; and his understanding of the comparative-historical and

---

3. These statistics differ from those of Jucquois 1966:61 and 1971, where the frequency of *d* is 83 and that of *g* is 70 (while that of *b* is 31). The difference is evidently due to the fact that Jucquois's frequency counts of Pokorny, unlike ours, were based only on root morphemes (as the root is defined by Benveniste). The frequencies of phonemes based on Pokorny obviously do not accurately reflect Proto-Indo-European frequencies, since Pokorny includes not only Proto-Indo-European material but also recent forms that arose in the separate branches. Hence the frequency of *b* in Indo-European forms based on Pokorny is much higher than the actual number of ancient Indo-European forms containing *b*; the possible examples of *b* have been surveyed above. Our count of Pokorny yielded 78 instances of *b*, as against Jucquois's 31. But even for these figures, based on all forms (Proto-Indo-European and later) in Pokorny, the relative frequencies are revealing: the frequency of *d* and *g* is significantly greater than that of *b*.

typological issues involved is not entirely correct. The issue is not the complete absence of voiced stops from Proto-Indo-European, since the traditional voiced aspirates must have been voiced stops; rather, it is the phonological interpretation of the stop series traditionally reconstructed as plain voiced (and in which the labial member was either missing or extremely infrequent).

Erhart 1984 also rejects the voicelessness of Series I on the grounds that the so-called voiced-voiceless opposition is very old in the Indo-European stops. He fails to note that the proposed reinterpretation of the traditional voiced stops as voiceless in no way removes the voicing opposition from the Indo-European stop system but merely redistributes it, transforming the system into one where Series II represents voicing and the other two are voiceless.

1.1.2. Pedersen's reinterpretation of the traditional system

Pedersen established the absence or near-absence of *b in Proto-Indo-European and on that basis first raised the possibility of reinterpreting the traditional voiced stop series as unvoiced (or voiceless). A language which lost *b* while preserving *d* and *g* seemed improbable to him, while there are many examples of languages which lose *p* while preserving *t* and *k*. On this basis, Pedersen proposed a reinterpretation of the traditional reconstruction that was daring for its time: the voicing values for the traditional Series I and III were reversed, so that Series III was now voiced and I voiceless, while Series II was seen as possibly voiceless aspirated (see Table 2).

Table 2

Traditional system				Pedersen's system		
I	II	III		I	II	III
—	b^h	p		—	p^h	b
d	d^h	t		t	t^h	d
g	g^h	k		k	k^h	g

Pedersen regarded the consonant system he had posited as early Indo-European or pre-PIE (*Vorindoeuropäisch*). It changed into the traditionally reconstructed system of Proto-Indo-European (*Gemeinindoeuropäisch*). As a typo-

logical example he adduced the development of voiced and voiceless consonants in the eastern and western dialects of Armenian (Pedersen 1951).

Pedersen's procedure for transferring his reconstructed 'early Indo-European' system into the traditionally posited PIE system was dictated by the difficulty of deducing the attested consonant systems from the system he proposed. Although the traditionally reconstructed system is contradicted by synchronic typological data (as was first noted by Pedersen), it nonetheless has the property of diachronic deducibility: the historically attested Indo-European consonant systems can be derived from it easily and without contradiction, utilizing typologically verifiable sound changes (i.e. changes supported by the historical development of many attested languages).[4]

The change of Pedersen's pre-PIE system into the traditional PIE system leaves the traditional view of Indo-European consonantism essentially intact, since all the Indo-European languages are derived from the traditional system, the one with synchronic typological contradictions. Hence Pedersen nullifies his own reinterpretation, turning it into a mere appendage to the traditional consonant system. This is undoubtedly the reason why, except for a few isolated observations,[5] his line of inquiry has not been followed up in subsequent Indo-European comparative grammar.[6]

There have been recent attempts to ascribe the reinterpreted system to a pre-PIE stage which then subsequently changed into the traditional system from which the daughter languages can ultimately be described, e.g. Haider 1982. Haider corrects the typological weaknesses of the traditional system by proposing a 'pre-IE' consonantism consisting of preglottalized voiced, plain voiced, and plain voiceless consonants; these shifted in PIE to respectively plain voiced, voiced aspirate, and voiceless (aspirate) as a result of 'intensification' (*Stärkung*). It is not clear what kind of phonetic process is meant by 'intensification', which shifted the plain voiced series into highly marked voiced aspirates. Nor does positing preglottalized voiced stops in place of the traditional plain voiced stops remove the typological obstacles of the traditional system (absence or weak representation of *b and high frequency of the voiced velar; absence of roots with two voiced stops; etc.), since preglottalized voiced stops, in contrast to glottalized voiceless stops, have the very same markedness relations — dominant labial, recessive velar — as plain voiced stops (see Greenberg 1970:125).

---

4. This may have been one of the reasons why the traditional Indo-European consonant system was so long accepted as the source of the attested systems, despite the fact that on further inquiry it turns out to be contradictory from the viewpoint of synchronic typology.

5. See Martinet 1953b:70. Martinet proposes an original interpretation of the traditional voiced series as glottalized; see below. Pedersen's model is adopted by I. Melikišvili 1971:216, except that the traditional voiced stops are replaced not by voiceless (nonvoiced) stops but by aspirated (rather than glottalized) stops, since aspirates are more marked than plain voiceless stops.

6. Only very recently have alternative models begun to be proposed, based on considerations of simplicity in description (see Emonds 1972).

This approach (and others like it) represents an attempt to save the traditional stop system at any cost. But the traditional system is only a reconstructed model, posited on the strength of theoretically interpreted correlations among historically related languages, and not a historically attested system; hence it is a hypothesis like any other. Even a time-honored theoretical construct like the classic PIE consonant system can be replaced by an alternate model supported by the most recent findings, when justified by the overall development of comparative-historical and typological studies in contemporary linguistics.

1.2. A typological interpretation of the three Indo-European stop series

1.2.1. Inconsistency of the traditional Indo-European stop system with the facts of phonological typology

The traditional picture of Indo-European consonantism was drawn up at the dawn of Indo-European comparative grammar and largely coincided with the consonant systems of the languages with ancient written traditions: the classical languages, Greek and Latin, and especially Sanskrit. These languages had great prestige, and — as has often been the case in comparative Indo-European grammar — effectively determined the shape of the reconstructed system. Linguistic systems displaying differences from the prestige languages were explained as having changed from an original system essentially identical to those of the languages with ancient traditions. This is why Grimm, who following Rask established the correspondences between Germanic and the classical languages, regarded the Germanic system as due to a shift (*Lautverschiebung*) of the original Indo-European phonemes. This view, passed on from generation to generation of Indo-Europeanists, dominates the comparative grammar of Indo-European languages to this day.

It can easily be seen that this picture is not the result of principled linguistic analysis, but is rather the product of historical accident due to the prestige of languages with ancient literary history. The structures observed in these languages were projected for Proto-Indo-European antiquity, while the structures of the other Indo-European languages were viewed as transformations and restructurings of the Proto-Indo-European system.

An inconsistency in this approach arose fairly early in relation to the Indo-European vocalism: Sanskrit vocalism proved to be secondary compared to that of other Indo-European languages, hence not representative of the Proto-Indo-European situation. However, there was no doubt in the minds of most scholars

as to the Proto-Indo-European nature of the Sanskrit stop system (and, in part, Greek and Latin consonantism).

Although the traditional Indo-European stop system offers diachronic deducibility, it lacks consistency with the facts of synchronic typology. The inconsistency pertains most saliently to the absence (or extremely low frequency) of voiced labial *b in Series I. According to synchronic typology (see Greenberg 1966, 1970, Hamp 1970a, I. Melikišvili 1972, 1974, 1976, Campbell 1973), in systems with a voicing opposition in stops the marked (i.e. recessive) point of articulation in the voiced series is *g*, and *b* is unmarked (i.e. dominant); while in voiceless series *p* is recessive and *k* is dominant. These markedness relations determine the relative frequencies of phonemes and gaps in stop systems. The recessive member of the opposition generally displays lower frequency than the dominant one. In many systems that low frequency is equal to zero, which creates a gap in the phonemic system.[7]

Consequently we can expect the following types of stop systems (see I. Melikišvili 1972):

	(A)			(B)			(C)	
p	*t*	*k*	—	*t*	*k*	—	*t*	*k*
b	*d*	—	*b*	*d*	*g*	*b*	*d*	—

In such systems the recessive phonemes — the voiced velar *g* and the voiceless labial *p* — are absent or extremely rare.[8]

---

7. The impossibility of a gap in the voiced labial series and the probability of such a gap in the voiceless series are confirmation of Pedersen's opinion, expressed as a diachronic claim for the impossibility of loss of the voiced labial *b* and the frequency of loss of voiceless *p*. The significance of Pedersen's claim in the light of Greenberg's typological observations has already been noted in Hamp 1970a.

8. Lifu is sometimes adduced as an example of a systemic gap of *b* and a counterexample to the claim that the voiced labial *b* is unmarked in comparison to the velar *g* (Martinet 1955:III, 36, 1960:123-24, 137). The gap is established historically by Lenormand 1952, on the grounds that /b/ is found in loan words from European or neighboring Melanesian and Polynesian languages, while /d/ and /g/, as well as /p/, /t/, and /k/, are found in native words. However, this argument does not reflect the actual diachronic situation, since it does not take into consideration the origin of the Lifu voiced stops. According to Haudricourt (1971:380-84, 392-93, and correspondence tables 30 and 31), the contemporary Lifu voiced stops were seminasalized until recently: Lifu *ḍən* 'leaves' < ⁿ*ḍau*, Fijian ⁿ*ḍau* (ⁿ*ḍ* is a seminasalized cacuminal stop), Lifu *maḍa* 'blood', Fijian ⁿ*ḍaa*. Parallel to the reconstructed development of ⁿ*ḍ* > *ḍ* and ᵐ*b* > *b*, a development of **p > h > Ø* (-*w*- intervocalically) is proposed (cf. the development of -*p*- in Armenian). A detailed investigation of the correspondences of Lifu /d g/ to the consonants of related Eastern Austronesian languages would be interesting: Fijian lacks /p/ in native words (it is found only in loans, like Lifu /b/), while it has /t k/ and the full voiced series /b d g/; but the other Polynesian languages have only a voiceless series (except for Tongan, whose voiced series has been explained as due to Fijian influence): see Dempwolff 1920, Dyen 1965.

The mixed character of Lifu was pointed out by Gabelentz 1891:273, who established its relationship to Melanesian (Gabelentz 1873); it has also sometimes provided grounds for regarding Lifu as more closely related to the Papuan languages (Müller 1882:69*n*); for examples

Analogous dominance relations can be found among voiceless stop series. The most marked of the voiceless stops are the glottalized series,[9] the next most marked are the aspirates, and the plain voiceless stops are the least marked. The hierarchy for increasing markedness is: plain voiceless – voiceless aspirated – voiceless glottalized (Greenberg 1966).[10] Consequently, the most marked consonant among such series is the glottalized labial p', as is shown by the great rarity or complete absence of this phoneme in many languages having glottalized series (Greenberg 1970, Hamp 1970a), e.g. a number of North Caucasian langages, and many African and Amerindian languages, which lack /p'/ entirely.[11] Thus we have systems of types B′, C′, and A′, while types D and D′[12] are extremely rare exceptions (Gamkrelidze 1974:14-15):

(A′)				(B′)			(C′)	
p^h	t^h	k^h	—	t^h	k^h	—	t^h	k^h
—	t'	k'	—	t'	k'	—	t'	k'
b	d	g	b	d	g	p	t	k

(D)			(D′)		
—	t^h	k^h	—	t^h	k^h
p'	t'	k'	p'	t'	k'
b	d	g	p	t	k

---

of Lifu words which do not show the regular correspondences see Kahlo 1960:28-29.

It is important that as his illustration of a phonological gap Martinet uses a highly uncharacteristic example, that of Lifu, for which there is no stage at which the absence of b can be reconstructed. Yet precisely this example subsequently appears in textbooks as a typical illustration of a gap in a phonological system.

9. We use the term *glottalized* in the strict sense, equivalent to *ejective* (see Ladefoged 1971:16ff.).

10. Objections can be made to the claim of Swiggers 1980 that in the South Arabian language Harsusi glottalized t, s, $\check{s}$, and q (and also preglottalized d) are more frequent than voiceless aspirates. For one thing, it is methodologically incorrect to oppose the entire glottalized series, which includes fricatives, to voiceless aspirate stops. Aspiration is not phonologically relevant in the voiceless aspirate stops, hence there is an opposition of glottalization within the voiceless stops. But in such an opposition, the defective glottal series _ t' q' (with missing labial member) cannot possibly be more frequent than the unglottalized series p t k (although the individual glottalized member q' may be more frequent than unglottalized k: see Melikišvili 1974:101ff.).

11. Greenberg's universals concerning the distribution of glottalized consonants in systems have been largely confirmed by the extensive survey of Fordyce 1980.

12. A system of this type, with a gap for the aspirated labial (instead of the glottalized labial, which we could expect to find lacking), is found in one dialect of Galla (Andrzejewski 1957, Sasse 1973).

The next inconsistency between the traditional Indo-European stop system and synchronic typological data is the absence of a voiceless aspirated series despite the presence of a voiced aspirated series (noted by Jakobson 1957a). There are no languages attested with a voiced aspirate series but no voiceless aspirates.[13]   In this respect the traditional reconstructed system is in clear contradiction with the facts of synchronic typology.[14]

1.2.2. A reinterpretation of the Indo-European stop system. Glottalization as a natural feature for the defective stop series

The respects in which the traditionally reconstructed Indo-European stop system fails to conform to synchronic typological facts demand reconsideration of the system so as to bring it into alignment with what is known about typology. Any reinterpretation of the Indo-European stop system must be done with an eye to the diachronic deducibility of the system, which must make possible non-contradictory, typologically verifiable derivation of all attested daughter systems from the posited protosystem.   It is of particular importance to define the distinctive features of the three Indo-European stop series in a way that will cor-

---

13. Jakobson's proposed universal has recently been disputed by Blust (1973) on the evidence of the Austronesian language Kelabit. This language has three clear stop series, which Blust interprets as voiced /b d g/, voiceless /p t k/, and voiced aspirated /bʰ dʰ gʰ/. Unfortunately, the paper does not give a precise phonetic description of the sounds to support this phonemic analysis. But even the examples cited give reason to interpret the series not as voiced, voiceless, and aspirated, but rather as voiced /b d g/ (with combinatory variants [b], [d], [g] and [-bb-], [-dd-], [-gg-], the latter set after stressed *ə́*), voiceless /p t k/ (with combinatory variants [p], [t], [k] and [-pp-], [-tt-], [-kk-], the latter set after stressed *ə́*), and half-voiced /b d g/ (with combinatory variants [b], [d], [g] and [-bp-], [-dt-], [-gk-], the latter set after a stressed vowel). Blust himself notes that 'aspiration' (more precisely, the voiceless onset of the following vowel) is optional in the geminated variants of these phonemes. Here we evidently have to do not with phonological aspiration of a voiced stop series, but with phonological half-voicing and hence greater intensity compared to the plain voiced series. This is shown by the appearance of sequences -bp[ʰ]-, -dt[ʰ]-, -gk[ʰ]- after a stressed vowel, i.e. a sequence of lax voiced consonant followed by tense voiceless, e.g. in alternants such as *táp[ʰ]əŋ* 'chopping down of trees' beside *təbə́ŋŋən* 'chop down (imper.)'.
    This interpretation of the Kelabit consonants as half-voiced finds a close structural parallel in northern Chinese dialects which have half-voiced /b d g/, each with two variants: [b d g] and [-pb-], [-td-], [-kg-] (Polivanov 1928, 1968:68, 250-51). These differ from the Kelabit series only in the distribution of allophones and the order of segments in the second allophone type. In Chinese, the voiced variant appears in intervocalic position and the half-voiced one initially. The absence of aspiration in the half-voiced allophone of Chinese is due to the order of its segments, with the voiceless one preceding the voiced, whereas in Kelabit we have the opposite order, voiced plus voiceless, which naturally allows for aspiration in the form of a voiceless onset of the following vowel.
    14. On this basis, Szemerényi (1967:94ff., 1970, 1972:134) attempts to reconstruct a fourth series of voiceless aspirated consonants, thereby returning to the system posited earlier by Brugmann (see also Back 1979). This is unjustified from the standpoint of comparative analysis, since the voiceless aspirates of Indo-Iranian and other branches give clear evidence of combinatory origin.

respond to the facts of both synchronic and diachronic typology.[15]

We begin with a detailed survey of each series of the traditional system (System I) in relation to the others. The violation of typological facts noted by Jakobson, namely the lack of voiceless aspirates beside the voiced aspirates of the traditional system, disappears if we reinterpret Series III ('voiceless') as voiceless aspirated. Such an interpretation is consistent with the reflexes of this series in a number of historical Indo-European languages. Thus we have voiceless aspirates (Series III) beside the voiced aspirates of Series II, which brings the traditional system into full accord with synchronic typological facts and provides for natural derivation of the attested systems from it.

The reinterpretation of Series III as voiceless aspirated brings us to the question of how that series was related to the other two, especially Series I. The reinterpretation of Series III is sufficient to require a new interpretation of Series I. Moreover, the internal properties of Series I themselves demand reinterpretation and accommodation to typological facts. Series I, traditionally reconstructed as voiced, must be reinterpreted as nonvoiced in view of its missing labial member: as was shown above, cross-linguistically the labial member is often defective (missing or of low textual and systemic frequency) and hence marked in voiceless stop series. Furthermore, the most marked of these series, as has been shown by Greenberg 1970, is the glottalized series, for which the absence or low frequency of p' is typical. This universally valid feature of ejectives bears directly on the phonological status of Series I with its defective labial member. It is also relevant that Series I shows a lower overall frequency than Series II and an even more disproportionately low frequency in comparison to Series III. According to counts by Jucquois (1966), the overall frequencies of phonemes of the three series in root morphemes are as follows:

Series I   6.2%
Series II   8.9%
Series III   17.7%

This is consistent with the absolute frequencies of occurrence of phonemes from the three series in Indo-European words, based on our counts of Pokorny.

These frequency correlations in themselves are sufficient to raise doubts about the traditional interpretation of the three series as respectively voiced, voiced aspirated, and voiceless. The voiced aspirated series, typologically

---

15. The discussion to follow deals only with the three points of articulation that are certain: labial, dental, and velar. As has been noted above, the full membership of the dorsal set, in Indo-European or in groups of dialects, has no bearing on this problem and will be given separate attention later when the full stop system of Indo-European is discussed. It should be borne in mind that our conclusions about the manners of articulation within the velar set are also valid for the other dorsals grouped with the velars. Thus in our notation we use the cover symbols G, G^h, K for the velars together with the other dorsal phonemes.

marked in comparison to the plain voiced series, should not be the more frequent of the two according to the universal frequency relations of markedness. If Series I is interpreted as glottalized, the frequency properties of the three series are in full accord with typologically established frequency correlations of ejectives in relation to other voiceless series, in particular voiceless aspirates and plain voiceless consonants.

In interpreting Series III as voiceless aspirated, in accord with typological considerations and the reflexes in the daughter languages, we are forced to interpret Series I, with its defective labial, as glottalized and not, for instance, as plain voiceless or voiceless aspirated. The two latter interpretations are excluded in any case by the fact that within unvoiced stops a glottalized series is the most marked in comparison to aspirated and plain voiceless stops. The aspirated series is in turn more marked than the plain voiceless series. Thus if one labial member is missing among unvoiced stops, the gap must necessarily be in the most marked glottalized series, not in the less marked aspirates or the even less marked plain voiceless stops (cf. 1.2.1 above).

In summary, when Series III is reinterpreted as voiceless aspirate and frequency correlations are considered, the defective labial phoneme of Series I forces us to interpret Series I as a glottalized series. On this interpretation, the Indo-European stop system takes the following form:

Table 3

I	II	III
(p')	b^h	p^h
t'	d^h	t^h
K'	G^h	K^h

Defining Series II in this system as voiced aspirates brings us to the question of the phonetic reality of its consonants. Based on typological comparison with the modern Indo-Aryan languages (Catford 1964, 1977:106, Ladefoged 1967:9), the traditional voiced aspirates of Indo-European are generally seen as the type of stop characterized by murmured release, pronounced with breathy voice (Ladefoged 1971:12, Lass 1974, Hopper 1973, Butler 1974).

An essential phonological fact is that, in systems having phonemes with murmured release, this series is opposed to a voiceless aspirate series. Thus they can be defined phonologically as voiced (or non-voiceless) aspirates, phonetically realized as voiced or half-voiced (e.g., the Georgian voiced stops)

with accompanying aspiration. The aspiration may also be voiced, involving vibration of the vocal cords (or part of them, with the other part open). Such sounds are characterized by intensity of articulation. An example is the voiced aspirates of several contemporary Armenian dialects.[16]

The stop system we have posited for Indo-European is in complete accord with the synchronic typological facts: the defective glottalized labial, functionally the weakest member of the group of unvoiced consonants; the presence of voiceless aspirates as well as voiced aspirates; the relative frequencies of the various series, reflecting increasing markedness as we go from voiceless aspirates to voiced aspirates to glottalized consonants.[17]

1.2.3. A phonological characterization of the reconstructed Indo-European stop series

In the stop system we posit for Indo-European, aspiration is phonologically irrelevant, since series II and III are opposed not in aspiration but in voicing. Aspiration must be regarded as an incidental phonetic feature accompanying the phonemes of these series. In strictly phonemic terms Series I can be described as glottalized, II as voiced, and III as voiceless. However, phonetic aspiration is a very important feature of Series II and III, one which explains their subsequent development and eventual reflexes in the daughter languages. Such phonetic features play a major role in diachronic phonemic transformations, and it is essential that they be considered together with phonological features in the reconstruction of phonological systems.

Since aspiration is irrelevant to the phonemes of Series II and III, they can have allophonic realizations without aspiration. Aspirated and unaspirated phones both could appear as combinatory variants of Series II and III phonemes. Every Series II and III phoneme must have had both aspirated and unaspirated allophones depending on position in the word.[18] Table 4 shows the phonetic variants to be reconstructed for the phonemes shown above in Table 3.

---

16. For Armenian see Allen 1951; for aspirates in Gujarati see Fischer-Jørgensen 1968:88-89, Elizarenkova 1974:180-81; also Ternes 1973:20.

17. We proposed this stop system for Proto-Indo-European in 1972 (Gamkrelidze and Ivanov 1972, 1973; see also Gamkrelidze 1976, 1977). A similar system of Indo-European stops is reconstructed by Hopper (1973, 1977). The possibility of interpreting the defective stop series as glottalized had been noted in passing by Martinet in a work dealing with Semitic consonants (1953b:70). Glottalized stops are also posited by Širokov 1972, in Indo-European reconstructions which are not entirely clear from a typological point of view.

This conception of the Indo-European consonantism has begun to be called the 'glottalic theory' in Western works: Bomhard 1979. A similar view of the Indo-European voiced series has recently been advanced by Haudricourt 1975, although without reference to the preceding works.

18. Gamkrelidze 1976, 1977; see also Normier 1977.

Table 4

I	II	III
(p')	*bh/b*	*ph/p*
t'	*dh/d*	*th/t*
K'	*Gh/G*	*Kh/K*

This reconstructed Indo-European stop system is typologically plausible as regards the relations among the individual series and can be regarded as typologically realistic overall. Phonemic inventories of this type, where stops contrast in glottalization and voicing, are extremely frequent among the world's languages. An example where voiced aspirated and plain voiced sounds are allophones of one phoneme type comes from modern Armenian dialects, where aspirated and unaspirated voiced sounds are positional variants (see Allen 1951, Djahukian 1967b:78-81). The existence of such systems provides good typological confirmation for our reconstruction of the Indo-European stop system.

Table 5

Phonetic and phonological feature matrix for the Proto-Indo-European stops. (The matrix includes both phonological features and those phonetic features that were essential to the functioning and development of the Proto-Indo-European system; the symbols chosen are maintained throughout the book with optionality of aspiration left unexpressed in stops of the series II and III.)

	(p')	*t'*	*K'*	*bh*	*dh*	*Gh*	*ph*	*th*	*Kh*
Syllabicity	−	−	−	−	−	−	−	−	−
Stopness	+	+	+	+	+	+	+	+	+
Voicing	(−)	(−)	(−)	+	+	+	−	−	−
Glottalization	+	+	+	(−)	(−)	(−)	−	−	−
Labialization	+	(−)	(−)	+	(−)	(−)	+	(−)	(−)
Dentality	(−)	+	(−)	(−)	+	(−)	(−)	+	(−)
Velarity	(−)	(−)	+	(−)	(−)	+	(−)	(−)	+
Aspiration	(−)	(−)	(−)	±	±	±	±	±	±

1.3. Phonotactics and combinatory rules for the Indo-European phoneme series

1.3.1. The basic canonical forms of the Indo-European root

The allophones of the stops must have been positionally conditioned when phonemes were combined in the Indo-European word. The conditioning environment could have been either the adjacent phoneme or a more distant one, separated from the stop by an intervening vowel and possibly other phonemes in the same word. In forms with zero grade an otherwise distant conditioning phoneme could become adjacent.

The basic phonotactic rules for stop distribution within the word can be formulated as constraints and restrictions on the combinability of phonemes or individual features. A fundamental restriction imposed on the entire stop subsystem was a constraint against combining identical phonemes within the root. It can be formulated as Rule 1:

Rule 1. No two stop phonemes with identical feature values can cooccur in a root of the structure C_1VC_2-. (Roots of the shape T_1ET_2, where $T_1 = T_2$, are impossible.)[19, 20]

A fundamental constraint on phoneme combinability within the root was a prohibition on distant combination of two phonemes of Series I,[21] i.e. two glottalized consonants. Thus:

Rule 2. No two glottalized consonants (i.e. consonants of Series I) can cooccur in the same C_1VC_2- root. (Roots of the shape **t'ek'*-, i.e. traditional **deg*-, are impossible.)

This constraint against the occurrence of two glottalized consonants in the

---

19. Here and below, T stands for any stop, T^h for any voiceless (aspirated) stop, T' for any glottalized stop, and D^h for any voiced (aspirated) stop, with each symbol representing an entire series. Thus a form T^hET'- or $T'ET^h$- subsumes any combination of phonemes of the two series involved, except for identical phonemes, which are precluded by Rule 1.

20. There is reason to assume that Rule 1, prohibiting combinations of identical phonemes (see also Benveniste 1935:170-71, 1955:201-2), is an instance of a more general rule, Rule 1', which precludes combinations of two phonemes of the same point of articulation within a root of the structure CVC-. Evidently a principle of heterorganic phoneme combinations operated in the Indo-European root. This problem will be investigated in more detail when the Indo-European points of stop articulation are examined. At this point, what is of interest is the nature of the Indo-European manners of stop articulation, the series, and the relations among the series.

21. This principle was already clearly stated (in terms of the traditional voiced series) by Meillet 1937, 1938:191; see also Lehmann 1952, Ammer 1952, Szemerényi 1970, Jucquois 1966.

same morpheme finds wide typological parallels in languages having glottalized consonants. For instance, in native Kartvelian words no two non-identical glottalized consonants may cooccur. An analogous restriction is found in Shuswap, a Salishan language of British Columbia, where in roots of the form C_1VC_2- (also C_1RVC_2-, C_1VRC_2-) C_1 cannot be glottalized if C_2 is (Kuipers 1974:23). Similar constraints are observed in a number of other Amerindian languages, e.g. Yucatec (Mayan: Straight 1972:59) and Quechua (Carenko 1972:100, 1973:82, 1974a:19, 1974b:95, Rowe 1950:139-40, Orr and Longacre 1968:529-30), as well as in Hausa (Chadic), where two heterorganic glottalized consonants are never found within one word (Parsons 1970:280).[22]

Such constraints point to physiological and articulatory properties of the glottalized consonants, which tend not to cooccur in distant combinations. In general these constraints do not hold if the two consonants are homorganic; however, for Indo-European, homorganic distant combinations are precluded by Rule 1. Glottalization, like pharyngealization, aspiration, and several other features, can be described phonetically as a *dissimilating* feature, one which tends not to occur twice in a syntagmatic sequence (voicing, in contrast, is a non-dissimilating feature). This accounts for the cross-linguistic low frequency of words with two ejectives or two pharyngealized consonants, as well as for the tendency to avoid creating such sequences in inflected forms (see Ohala 1981:193-95).

Rule 2 is essential to the typological justification of Series I as glottalized (rather than, as traditionally, voiced). When Series I is interpreted as voiced, it is difficult to find typological parallels for a constraint against cooccurrence of voiced consonants in the root or for the impossibility of roots such as *ged-* and *deg-*. This constraint, noticed by Meillet (1937, 1938; see also Lehmann 1952, Jucquois 1966, 1971), has so far gone unexplained. When we reinterpret the traditional voiced stops as glottalized, the difficulty is removed and the Indo-European constraint can be reduced to a general typological regularity governing phonotactics.[23]

The combinability of glottalized consonants with those of Series III is unconstrained, and all theoretically possible combinations can occur. Thus:

---

22. The same constraint against combining two glottalized consonants (including glottal stop) also explains the origin of a tone in Lahu (of the Loloish subgroup of the Lolo-Burmese branch of Sino-Tibetan): Matisoff 1970. There is a phonetic constraint against two pharyngealized (emphatic) consonants occurring in the same root in Akkadian: in the Akkadian root, one of two Semitic heterorganic emphatic consonants undergoes dissimilation and appears as a simple (unmarked) consonant (Geers's Law): see Geers 1945.

23. Glottalization of the Indo-European Series I also explains the fact (long noted) that the traditional 'voiced' consonants are almost never found in affixes, unlike the consonants of Series II and III. This would be difficult to explain on typological grounds if Series I were treated as voiced. Glottalized stops, on the other hand, are typologically characterized by restrictions on their distribution among morphological elements: Hopper 1973:157.

Rule 3. Glottalized consonants can cooccur with any member of Series III in any order. (The combinations T'ET^h and T^{h}ET' are possible.)

There is a striking absence of Indo-European roots combining consonants of Series II and III (in either order). There are no roots of the shape D^{h}ET^h, T^{h}ED^h (in traditional terms, **bhet-* and **tebh-*). Thus:

Rule 4. Non-glottalized consonants within the root must have identical values for the feature of voicing. (Only roots of the shape D^{h}ED^h and T^{h}ET^h are possible.)[24]

The absence of Indo-European roots of the type D^{h}ET^h and T^{h}ED^h could be explained diachronically, as the result of distant voicing assimilation: the earlier T^{h}ED^h and D^{h}ET^h would have yielded T^{h}ET^h or D^{h}ED^h, still within Proto-Indo-European. Thus the roots with homogeneous voicing may conceal some earlier roots with discrepant voicing (see Miller 1977a:32). This could explain the fact that the root types T^{h}ET^h and D^{h}ED^h are more frequent than the other admissible types: they include the reflexes of roots with heterogeneous voicing as well as those with originally homogeneous voicing.

These combinatory possibilities for stops in the Indo-European root can be reformulated in terms of distinctive features. For example, the constraint against sequences T'ET' can be stated as a constraint against combining two feature matrices with positive values for glottalization. The non-occurrence of the shapes D^{h}ET^h and T^{h}ED^h reflects the impossibility of combining two feature matrices with different values for voicing; the only distant combinations allowed have identical values for this feature.

The combinatory possibilities for individual phonemes in the Indo-European root can be surveyed after the full inventory of point-of-articulation types is given and the restrictions on their combinability are established (i.e. the inter-sections of the series and the local rows; see above and note 20 for Rules 1 and 1'). At this preliminary stage we have been working with the minimum number of local rows accepted by most investigators, which is sufficient for non-contradictory decisions about the nature of the series and the definition of their distinctive features. The information introduced so far concerning the number of rows and their nature is necessary and sufficient for judging the phonological nature of the Indo-European stop series.

---

24. This rule was first noted by Saussure (see Meillet 1912:60, Benveniste 1935:171*n*1, Szemerényi 1972:143*n*51). Based on this rule, Kuryłowicz (1973b:68, sec. 1.5, note 6) suggests that the phonemes of Series II and III were opposed not by two distinctive features (aspiration and voicing) but by 'one phonemic feature', a view which is close to ours. However, Kuryłowicz considers Series II to be neutral in voicing.

1.4. The distribution of allophones of the voiced and voiceless series

1.4.1. The distribution of aspirated allophones and their reflexes in Sanskrit and Greek. Grassmann's Law as the Indo-European rule determining allophones for Series II

A characteristic property of the Series II (voiced) and III (voiceless) stops is their phonetic feature of aspiration. Each phoneme of these series had two allophones, aspirated and unaspirated, depending on the phonetic environment.

The linguistic facts tell us with a fair degree of precision in what positions the phonemes of Series II and III appeared in their aspirated and unaspirated forms. The aspirated variant is the basic allophone, since it appears in most of the phonetically independent positions which can be established for Proto-Indo-European on comparative grounds. The unaspirated variant appears in particular phonemic contexts. Consequently the task of distributive analysis is to determine the positions in which the phonemes appeared in their unaspirated forms. The second, voiced series lends itself most precisely to this kind of analysis. In the daughter languages it has left quite clear traces which make it possible to reconstruct its distribution in Indo-European wordforms.

One of the major principles determining the behavior of Series II is the fact that when two Series II phonemes appear in a single stem they must always appear as different allophones — one aspirated, the other unaspirated. This is formulated in Rule 5:

Rule 5. In a stem with two Series II stops in distant combination, one of them is always aspirated and the other unaspirated.[25]

Thus any stem can contain only one unaspirated sound. If the initial consonant is unaspirated, the second is aspirated, and vice versa.

This distributional property of Series II consonants can be clearly seen and reconstructed on the basis of Indo-Iranian and Greek evidence. These languages reflect an allophonic rule for Series II whereby the unaspirated allophone is found initially and the aspirated one non-initially before a vowel or sonant:[26]

---

25. A typological parallel is the connection in Quechua between a rule against combinations of aspirated stops and one against combinations of glottalized stops (Carenko 1972:102).

26. The unaspirated *d* in Skt. *dvā́r-* 'door' beside Gk. *thúra*, Lat. *forēs* 'double door' (PIE **d*ʰwer- / ***d**ʰur-) is evidently due to a secondary, Indic loss of aspiration. In Sanskrit, initial *dh* is an unstable sound which tends to lose one of its features (usually stopness): cf. *hitá-* 'supplied; good' < **dhitá-*, *gr̥há-* < **gr̥dhá-* 'house', etc. (Wackernagel 1896:I, sec. 217ff.; Bloch 1934:64ff.). Independently of this tendency, loss of aspiration in the initial *dh of 'door' could have been facilitated by identification of this word with 'two, double', *dváu* (Pokorny 1959:279), cf. the meaning 'double door' of Lat. *forēs*.

Skt. *bāhúḥ* 'arm', Gk. *pêkhus* 'elbow' point to PIE [*baG^hu-s], with an initial unaspirated allophone and a non-initial aspirated allophone of PIE *b^h and *G^h (both underwent devoicing in Greek).

Skt. *badhnáti*, later *bandhati* 'ties, binds', *bándhuḥ* 'kinship, kinship by marriage, relative', Gk. *pentherós* 'father-in-law, wife's father' (from 'related by marriage'): PIE [*bend^h-].

Skt. *bahú-* 'dense, thick, numerous', Gk. *pakhús* 'fat, thick': PIE [*benG^h-].

Skt. *bódhati*, *bódhate* 'wakes up, wakens', Gk. *peúthomai, punthánomai* '(I) recognize, notice, stay awake': PIE [*beud^h-], [*bud^h-].

Skt. *budhnáḥ* 'soil', Gk. *puthmḗn* 'soil': PIE [*bud^h-].

Skt. *dáhati* 'burns', *nidāgá-ḥ* 'heat, summer', Gk. *téphra* 'ashes': PIE [*deG^h-].

Skt. *deh-* 'rub, smear, spread on, anoint', *dehī́* 'wall, dam, embankment', Gk. *teîkhos* 'wall': PIE [*deiG^h-].

These Sanskrit and Greek cognates provide good evidence for the distribution of aspirated and unaspirated allophones in Indo-European.

We find analogous distributional behavior of Series II phonemes in forms with reduplication of the initial consonant. Reduplicated forms of the Indo-European verb roots *d^hē- 'put, place', *G^hē- 'leave', *b^her- 'carry', and others yield particularly good evidence. Reduplicated forms are represented by Skt. *dádhā-mi*, Gk. *títhē-mi* '(I) put' from PIE [*di-d^hē-mi]; Skt. *jáhā-ti* 'leaves', Gk. *kíkhēmi* '(I) overtake', 2sg. *kíkheis*, from PIE [*gi-G^hē-t^hi]; Skt. *bí-bhar-ti* 'carries' from PIE [*bi-b^her-t^hi], cf. Gk. *es-piphránai* 'carry in' (Mayrhofer 1963:II.475).

The closeness, even identity, of the Sanskrit and Greek verb forms with initial reduplication, together with current assumptions about the unity of the Greek-Aryan verb system, permit us to see the pattern as an areal one within Indo-European, which also confirms the Indo-European source of these phoneme alternations. They cannot be explained as due to deaspiration and independent parallel development of unaspirated phonemes in the two branches (the usual understanding of Grassmann's Law in classical Indo-European linguistics: Grassmann 1863, 1863a), but must be seen as the reflex of a shared distributional rule that had areal status in Proto-Indo-European. The facts from the other Indo-European languages — Italic, Germanic, and others — are fully consistent with this treatment of these correspondences.

Thus Grassmann's 'Law of Deaspiration' acquires a completely new interpretation. It can be seen as the alternation of aspirated and unaspirated sounds at the allophonic level in Proto-Indo-European, not as independent deaspiration of aspirated phonemes in Sanskrit and Greek. The virtual identity of the Sanskrit and Greek deaspiration is due to their common origin in Indo-European.[27] It

---

27. For Grassmann's Law as Proto-Indo-European in nature cf. also Butler 1974:19.

follows that this process is to be reconstructed for Indo-European as an allophonic rule, which later turned into a phonemic alternation when the reflexes of the aspirated and unaspirated voiced allophones were phonologized in the daughter branches. After the reflexes of Series II and III were phonologized in Sanskrit and Greek, the former allophonic rule turned into a rule for alternation of aspirated and unaspirated phonemes within the stem. This explains the productive Greek morphophonological alternation of *trikhós* ~ *thríks* 'hair' (PIE [*$drig^h$-]), *takhús* ~ *thássōn* 'fast' (PIE [*$d\underline{n}g^h$-]),[28] and formations such as Skt. *vidátham* 'distribution' (from *$*vidh$-$atha$-, with positional deaspiration of *dh* to *d*: see Mayrhofer 1976:III.208).[29]

Thus the Indo-European protoforms for the words discussed above are to be reconstructed not as *$b^h\bar{a}G^h$-, *b^henG^h-, *b^heud^h-, *b^hud^h-, *d^heG^h-, *d^heiG^h-, etc., with subsequent independent deaspiration of the reflexes of the initial *b^h and *d^h in Indic and Greek, but rather as [*$b\bar{a}G^h$-], [*$benG^h$-], [*$beud^h$-], [*bud^h-], [*deG^h-], [*$deiG^h$-], etc., with positionally opposed aspirated and unaspirated allophones of Series II voiced phonemes. These forms are reflected regularly in Sanskrit as *bāh-*, *bah-*, *bodh-*, *budh-*, *dah-*, *deh-*, and in Greek as *pākh-*, *pakh-*, *peuth-*, *puth-*, *teph-*, *teikh-*, etc. (with the regular Greek devoicing of both allophones, aspirated and unaspirated, of Series II). Thus the process of 'deaspiration' by Grassmann's Law acquires a common explanation in Sanskrit and Greek.

In addition, Sanskrit also preserves traces of different reflexes of these same roots in a different phonemic context: before pause (i.e. in absolute final position), and before -*s*- and certain other obstruents (those of Series II but not those of Series III): see below on Bartholomae's Law. Here the roots appear in forms such as [*b^hud-]: nom. sg. *bhut*, suffixed form *bhotsyati*, instr. sg. *bhudbhis* (Whitney 1889: §§ 141, 153-55, Anderson 1970, Phelps and Brame

---

28. Therefore there is no justification for dating Grassmann's Law to the sixteenth to fifteenth centuries B.C., the time when palatalization was operating in Greek, on the basis of Greek alternations like *takhús* : *thássōn* (see Janko 1977).

Sporadic Greek forms with two aspirates, such as *thuphlós* beside the regular *tuphlós* 'blind', should be regarded not as counterevidence to the ancient allophonic principle for two aspirates (contrary to Thumb and Scherer 1959:266; see also Miller 1977b), but rather as reflecting the rule assigning aspiration to one of two stops in a sequence and a subsequent generalization of aspiration over all segments of the sequence (see Hoenigswald 1965a). Another relevant factor is the later aspiration assimilation of stops (/C...C^h/ > /C^h...C^h/ or /C^h...C/ > /C^h...C^h/), which yields forms with two aspirates as found in Old Attic and Old Cretan: *Théthis* instead of *Thétis*, *Phánphaios* instead of *Pámphaios*; cf. also forms such as *thuthén* 'sacrificial' from *tuthén* (see Dressler 1975:65).

29. On the other hand, there is insufficient justification for deriving Skt. *kumbhá-* 'pot, mug', Avest. *xumbō* 'pot' from Indo-Iranian *$*khumbha$- (with two aspirates), since the correspondence of Skt. *k* to Iranian *x* is regular (Mayrhofer 1956:I.234). The source is rather to be reconstructed as [*$kumb^ha$-], with one aspirate in non-initial position, following the allophonic law of Indo-Iranian. The correspondence of Skt. *k* to Iranian *x* could have arisen as the result of later spirantization of the initial stop in Iranian, as is particularly characteristic of consonant clusters.