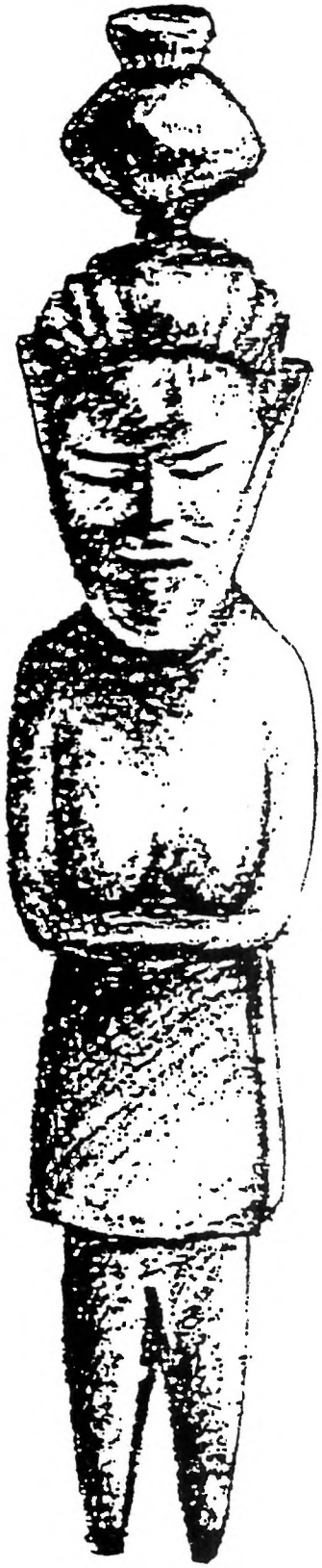
MOTHER TONGUE

JOURNAL OF THE ASSOCIATION FOR THE STUDY OF LANGUAGE IN PREHISTORY

Issue VI, 2000-2001



CONTENTS

Introduction to *Mother Tongue* VI

In Memoriam

3 Roger W. Wescott: *John D. Bengtson, Harold C. Fleming*
5 Cyrus H. Gordon: *Liny Srinivasan*
9 Joseph H. Greenberg: A Tribute and an Appraisal: *Harold C. Fleming*

Festschrift for Roger W. Wescott:
Paleolinguistics: The State of the Art and Science

29 Introduction: *John D. Bengtson*
31 Historical Overview: *Gyula Décsy*
35 Paleolinguistics in Denmark: *W. Wilfried Schuhmacher*
37 Paleolinguistics: Its Definition and Scope: *Irén Hegedűs*
41 Problems of Methodology in Long-range Comparison: *Peter A. Michalove*
45 Comparison and Reconstruction: Language and Mythology: *Michael Witzel*
63 Philosophical Differences and Cognitive Styles: *Sydney M. Lamb*
69 To What Extent is Paleolinguistics an Art and to What Extent a Science? What Constitutes Scientific Evidence in Linguistics?: *Paul Whitehouse*
75 What Constitutes Scientific Evidence in Paleolinguistics? *Winfred P. Lehmann*
79 On the Rules of Discourse for Paleolinguists in Relation to Notational and Transcription Systems: *Paul Whitehouse*
83 The Ever-green 'Beech'-argument in Nostratic Perspective: *Václav Blažek*

Austric Revisited

95 On the Origin of Affricates in Austric: *La Vaughn H. Hayes*
119 Comments on Hayes: *Paul Sidwell*
123 Response to Sidwell: *La Vaughn H. Hayes*

Book Reviews

129 *Sub-Gramnatical Survival,* by Mark R.V. Southern
Reviewed by Roger W. Wescott
131 *Indo-European and Its Closest Relatives,* by Joseph H. Greenberg
Reviewed by John D. Bengtson
137 *African Languages:An Introduction,* ed. by Bernd Heine & Derek Nurse
Reviewed by John D. Bengtson
144 *The Nostratic Macrofamily and Linguistic Palaeontology,* by Aharon Dolgopolsky
Reviewed by Vitaly Shevoroshkin
162 *Nostratic: Examining a Linguistic Macrofamily,* ed. by C. Renfrew & D. Nettle
Reviewed by Vitaly Shevoroshkin
182 *Numerals: Comparative-Etymological Analyses,* by Václav Blažek
Reviewed by John D. Bengtson
184 *Die Burushaski-Sprache von Hunza und Nager,* by Hermann Berger
Reviewed by John D. Bengtson

© 2001 by the Association for the Study of Language in Prehistory
ISSN: 1087-0326 *Mother Tongue* (Journal)

MOTHER TONGUE Issue VI, 2000-2001

Journal of the Association for the Study of Language in Prehistory

OFFICERS OF ASLIP

(As of Annual Meeting, April 21, 2001. Address appropriate correspondence to each.)

President: Michael Witzel witzel@fas.harvard.edu
Dept. of Sanskrit and Indian Studies
Harvard University
2 Divinity Avenue
Cambridge, MA 02138
U.S.A.

Vice-President: John D. Bengtson jdbengt@softhome.net
17 Hammond Street
Gloucester, MA 01930
U.S.A.

Secretary-Treasurer: Peter Norquest norquesp@U.Arizona.EDU
1632 Santa Rita Avenue
Tucson, AZ 85719
U.S.A.

Mother Tongue **Editor:** John D. Bengtson jdbengt@softhome.net
Associate Editor: Mary Ellen Lepionka mlepionk@ma.ultranet.com

BOARD OF DIRECTORS

Anne Windsor Beaman (Brookline, MA)
Ronald Christensen (Lincoln, MA)
Murray Denofsky (Somerville, MA)
Frederick Gamst (U. of Massachusetts)
Mary Ellen Lepionka (Gloucester, MA)
Jan Vansina (Madison, WI)
Allan R. Bomhard (Charleston, SC)
Gyula Décsy (U. of Indiana)
Harold C. Fleming (Boston University)
Kenneth Hale (M.I.T.)
Phillip Lieberman (Brown University)

COUNCIL OF FELLOWS

Raimo Anttila (University of Calif., L.A.)
Luigi Luca Cavalli-Sforza (Stanford University)
Ben Ohiomamhe Elugbe (University of Ibadan)
Vyacheslav V. Ivanov (Russian Acad. Of Sciences)
Winfred P. Lehmann (University of Texas)
Colin Renfrew (Cambridge University)
Vitaly Shevoroshkin (University of Michigan)
Ofer Bar-Yosef (Peabody Museum, Harvard)
Aharon Dolgopolsky (University of Haifa)
Dell Hymes (University of Virginia)
Sydney M. Lamb (Rice University)
Daniel F. McCall (Boston University)
Merritt Ruhlen (Stanford University)
Sergei A. Starostin (Russian State University)

ASSOCIATION FOR THE STUDY OF LANGUAGE IN PREHISTORY
Our Website: http://www.people.fas.harvard.edu/~witzel/aslip.html

Introduction to *Mother Tongue* VI

by John D. Bengtson

Since the publication of *Mother Tongue* V (1999) we have lost three of our most esteemed members, in order of death: Roger W. Wescott (ASLIP Vice-President), Cyrus H. Gordon (ASLIP Founding Member), and Joseph H. Greenberg (ASLIP Council Fellow). In honor of these scholars, we have devoted the first section of this issue to them *in memoriam.*[1] Tributes to them also appear in other sections, for example Gyula Décsy's article in the Wescott *Festschrift* section, and the book review of Greenberg's *Indo-European and Its Closest Relatives.*

In the Introduction to *Mother Tongue* V, Editor Roger W. Wescott briefly discussed sources of disagreements among linguists and other scholars and called for a substantive discussion of scientific linguistics. In honor of Roger, we solicited articles for the second section of this issue under the title "Festschrift for Roger W. Wescott: Paleolinguistics: The State of the Art and Science." We thank the nine writers (Václav Blažek, Gyula Décsy, Irén Hegedűs, Sydney M. Lamb, Winfred P. Lehmann, Peter A. Michalove, W. Wilfried Schuhmacher, Paul Whitehouse, and Michael Witzel) for their contributions.

The third section of this volume revisits the question of the proposed *Austric* (macro-)family of languages, which, at a minimum, unites the Austroasiatic and Austronesian families, and may include the Kadai and Miao-Yao families as well (and sometimes the "isolates" Nihali and/or Ainu). We first heard from La Vaughn Hayes in our first issue (1995: "More on the Austric Hypothesis and Austronesian's Inclusion"). In *Mother Tongue* II, III, IV, and V discussion of the Austric hypothesis continued. (See MT IV, pp. 2-3, for three alternative dendrograms of the Austric family.) In *Mother Tongue* VI Hayes resumes his elaboration of Austric phonology, with comments by the Australian scholar Paul Sidwell.

The last section of this issue contains seven book reviews. All the books reviewed relate to the themes of this issue: The review of Southern's *Sub-Grammatical Survival* is Roger Wescott's last contribution to *Mother Tongue.* The review of *Indo-European and Its Closest Relatives* discusses one of Joseph Greenberg's last two works, and the review of *African Languages* is intimately associated with Greenbergian questions. The two reviews by Vitaly Shevoroshkin (on the Cambridge Nostratic books) continue the discussion of paleolinguistic methods. Blažek's book on *Numerals* is an exemplar of current paleolinguistic research, and the review of Berger's *Burushaski-Sprache* touches on the discussion of a macro-family otherwise ignored in this issue: Dene-Caucasian. By the way, we are especially pleased by the active participation of three of our ASLIP Council Fellows in this issue, namely Sydney M. Lamb, Winfred P. Lehmann, and Vitaly Shevoroshkin.[2]

1. Readers are also referred to the obituary of Joseph H. Greenberg by Nicholas Wade in *The New York Times,* May 15, 2001, p. A21. It may also be accessed on the newspaper's website (www.nytimes.com), for a fee of $2.50, or in your library.
2. Incidentally, the special election of new Council Fellows resulted in the addition of three new Fellows: the archeologist Ofer Bar-Yosef (Harvard University, Peabody Museum) and the linguists Vyacheslav Ivanov (Russian Academy of Sciences) and Merritt Ruhlen (Stanford University).

Roger Williams Wescott
1925-2000

by John D. Bengtson

I was saddened when I learned that my fellow ASLIP Vice-President, Roger Williams Wescott, passed from this life on November 21, 2000, from a brain tumor. It was my privilege to have known him for a relatively brief time. We first knew each other through correspondence, after Roger graciously answered my letter in 1985 inquiring about his work with language origins and deep relationships among language families. We met in person in June 1988 at the Third International Meeting of the Language Origins Society (LOS) at Vanderbilt University in Nashville, Tennessee, and later that year (November) at the First International Interdisciplinary Symposium on Language and Prehistory at the University of Michigan in Ann Arbor. Later we got together at ASLIP meetings in Boston from 1996 through 2000. Whether in his letters (usually hand-written) or in person, I always found Roger one of the most erudite, stimulating, and congenial of scholars.

Roger's accomplishments and honors were many, and I will list only some of them here. He graduated *summa cum laude* from Princeton University in 1945, with a B.A. in English, and later with an M.A. in Oriental languages, and a Ph.D. in Linguistics. As a Rhodes Scholar at Oxford University, he earned the degree of M.Litt. in Anthropology. Following ethnolinguistic field work in Nigeria, he founded and directed the African Language Program at Michigan State University (1959). From 1966 through 1991, Roger was Professor of Anthropology and Linguistics at Drew University in Madison, New Jersey, where he founded and headed the Anthropology Department and the Linguistics Program. There he also served as Director of the Behavioral Science Program (1989-91). In 1988-89 he was the first holder of the Chair of Excellence in Humanities at the University of Tennessee.

Roger served as President of the Linguistic Association of Canada and the United States (LACUS, 1976-77); First Vice-President of the International Organization for the Unification of Terminological Neologisms (1988-96); President of the International Society for the Comparative Study of Civilizations (1992-95); member of the Board of Directors of ASLIP, and later Vice-President of ASLIP (1996-2000). He served as co-editor of *Futurics* and *Forum Linguisticum,* as well as *Mother Tongue.*

Roger was a pioneer in the fields of protolinguistics, paleolinguistics, and glossogonics (the study of language origins). In 1972 he organized the American Anthropological Association's first symposium on glossogonics, which resulted in the book *Language Origins* (1974), co-edited and co-authored by Roger with Gordon Hewes and William Stokoe. Roger was involved for many years with LOS (see above) and of course with ASLIP/*Mother Tongue* almost from the beginning.

Roger was free and often unorthodox in his thinking and activities. In 1980 he taught folklore and comparative religion aboard the S.S. Universe, a "floating college" sponsored by the University of Pittsburgh. He served as a forensic linguist in New Jersey state courts, hosted a cable television program ("Other Views"), and for two weeks in 1957 even appeared on the television quiz show, "The $64,000 Challenge."

ASLIP and the world are diminished without Roger's free and inquiring spirit.

Roger Williams Wescott

by Harold C. Fleming
Past ASLIP President
Gloucester, Massachusetts

Roger was named after no less a man than the founder of the state of Rhode Island, Roger Williams. Rhode Island was Roger Wescott's home state and, I suspect, there was a fair amount of Rhode Island heritage in his family's history. I regret that I never inquired sufficiently into that presumed rich past. As another New Englander with a healthy respect for the intricacies of such family histories, including the bitter wars with the local Algonkians, I could have learned a lot. Of course, it was my 'tribe' of intolerant Puritans who had driven Roger Williams out of Massachusetts Bay Colony in the first place. No apology is intended because our section of the 'tribe' had just moved to Connecticut (1635) when Reverend Williams left Massachusetts (1636).

Roger Wescott was one of the brightest scholars I've ever known and close to what the Germans call *Universalgenie,* a master of many fields or an expert in many different things. The word 'dilettante' is not quite right here because of its implications of amateur status (at least in French). In English the word does have some derogatory connotations too, but they are mild.

At root, in so far as he was specialized, Roger was an Africanist, a linguist, and a cultural anthropologist. He was not simply a historical linguist but also delved into theoretical matters. He was one of the early voices calling for serious inquiry into the origins of human language, as we have outlined before in *Mother Tongue.* It was Roger who gave us our label 'Long Ranger', for one who is willing to look into more distant relations than is customary. It was Roger, drawing on his knowledge of the history of science, who objected to the term 'emerging synthesis' that Renfrew had proposed for our common endeavors – the term was invented in the 1940s to refer to the new synthesis in biology. It was Roger who wrote a brief but highly persuasive comment on Trask's attack on Bengtson's Basque-Caucasic hypothesis.

Finally, perhaps what Roger wanted most to 'sell' to his colleagues was his phonological analysis of language prehistory and his general law of *apophony* or consonantal ablaut. In this he greatly resembled Carleton Hodge. As Gyula Décsy has said, this was another example of Voltaire's famous saying that "... the vowels count for nothing." But the moment for studying this was not at hand because the rest of us were more interested in the narrative, i.e., whether or which languages were related to each other. Roger left us very little narrative, very few specific hypotheses about genetic relationships, very little on taxonomy. The same was, of course, true of Hodge. Their attentions and great learning were directed at problems peculiar to themselves.

But, as we learned from conflicts in schedules and the like, Roger had a whole other commitment to the ISCSC (International Society for the Comparative Study of Civilizations), of which he was founder and for many years president, and the LOS (Language Origins Society). All in all, Roger was a tremendous scholar and we miss him!

In Memory of Professor Cyrus H. Gordon
A Great Twentieth-Century Humanist

by Liny Srinivasan

Professor Cyrus Herzl Gordon passed away peacefully on March 30, 2001 in his home in Brookline, Massachusetts, among his loved ones. A pioneer and leading twentieth-century scholar in the field of Near East Cultures and Ancient Languages, Professor Gordon has been an inspiration to his students, followers, and successors. In 1982, in his book *Forgotten Scripts*, Professor Gordon wrote, "Pioneers open fields and leave the refining process to less inspired but more meticulous successors. I shall endeavor to render justice to the refining process, but my sympathies are squarely with the pioneers, and against their destructive critics."

Professor Gordon's multi-faceted contributions (385 publications, including 35 books, are listed in his autobiography, *A Scholar's Odyssey*) crisscross most of the major fields of the humanities. It is impossible for a serious student or researcher in the arts; archaeology; linguistics; anthropology; religion; and the literature, history, and geography of ancient cultures not to interact with some of his brilliant discoveries, bold ideas, and global perspectives. It may be many years before we realize the total impact of his lifelong activities, which he often jokingly described as "breaking new grounds and opening new ways." I agree with L. H. Feldman that "to appreciate him adequately would require another scholar with the breadth of knowledge and versatility of a Professor Gordon" (*Biblical Archaeologist*, 1996).

Cyrus H.Gordon first gained international fame in the academic world for his Ugaritic studies. His World War II contributions toward breaking secret codes won him a variety of military honors. Recognition on the home front was delayed by his uncompromising attitude of adhering to facts, evidence, and inescapable conclusions, which put him into conflict with the prevailing conventions and general consensus among his close academic associates. Surprisingly, it is the same principle – "to seek the truth no matter where it comes from and wherever it might lead" – Gordon said was the foundation of his success. He was honored with five Festschrifts (1962, 1973, 1980, 1996, 1998), each consisting of papers on various aspects of his work. In 1975 the Royal Asiatic Society made him honorary fellow for his contributions in Ugaritic, Minoan, and Comparative Semitic. In 1995 Boston Hebrew College granted him the degree of *D. Litt.* He also was a member of the American Academy of Arts and Sciences and of the American Academy of Jewish Research.

The crowning event celebrating him was the special joint conference of the Association of the Professors of Hebrew, the Society of Biblical Literature, The American Schools of Oriental Research, and the American Academy of Religion. "Scholar of All Seasons - A Tribute to Cyrus H. Gordon" took place in Chicago in 1994. A year later the same Societies presented him with the March 1996 issue of *Biblical Archaeologist*, which was dedicated to him. That presentation took place in Philadelphia, where he was born,

and at the University of Pennsylvania, where he began his career as an instructor, field archaeologist and epigrapher. He was most touched by that award. Beautifully designed enlarged covers of the gift-issue decorated the top of his living room fireplace, and every time he talked about his former students, he looked at it with sparkling eyes. We knew that Professor Gordon's highest sense of achievement came from the success of his students. The March 1996 issue contains articles explaining and detailing most of his major research works, as well as results by former students who have become eminent scholars in their respective fields. The issue reflects Professor Gordon's philosophy focusing on "worldwide humanity as a unit." It also illuminates the versatility of his genius and his success in establishing through his students a variety of viable fields of expertise.

Shortly after the 1996 excitement, I showed Professor Gordon a Bengali newspaper (*Ananda Bazar*). It had published the Canaanite origin of Bengali *Desi* words as front page news, mentioning him as the world's foremost Orientalist, a linguist of international repute, and the publication of his joint paper in the *Mother Tongue* Journal.[1] His curious and intense look at the Bengali scripts reminded me of a picture I had seen of young Gordon examining Aramaic bowl inscriptions.

Beginning with the Ugaritic tablets, through the Ebla and Nuzi documents, to the comparative studies of Greek classics and the decipherment of Minoan scripts, Professor Gordon has broken down, one after another, the walls that kept biblical literature restricted to a limited geographic space and people. He not only restored biblical literature to its proper place in the environment of the ancient Near East civilizations, but showed their role as the pivot point of a highly dynamic cultural diffusion. In this context, I would like to reflect on his last great contribution, still unknown to most of his admirers. I have had the privilege of knowing him for seven years. With my background in the ancient Indian texts, I saw numerous points of similarities between the literature of the ancient Near East and that of India, particularly the *Ṛgveda*. Professor Gordon asked: "the *Ṛgveda* concerns cosmology rather than geography. Where in the *Ṛgveda* does the country *Ar-za-wa* occur?"

My answer was that the country *Arjika* is *Ar-za-wa*, just as the country *Saryana* reflects the Hittite *Sariyana*, Phoenician *Siryon* in the Anti-Lebanon region. The Upper sea and the Lower sea recall the familiar Akkadian names for the Mediterranean and the Persian Sea. The metaphor of four rivers of heaven echoes the biblical four rivers issuing out of heaven, while the people *Ibhyas* could parallel the Hivvites of the Bible. Professor Gordon agreed to examine my documentation along with the English sources. Neglecting his falling health, putting aside many impending chores, including book revisions, postponing his autobiography, and spending his precious time when he could work only

1. Liny Srinivasan and Cyrus Gordon, "Canaanite Vocabulary in Bengali and in Some Other IE Dialects of India," *Mother Tongue* I: 202-206 (1995).

for limited hours, he undertook another tremendous responsibility to investigate and find evidence for himself. After two years he wrote to me, "I am leaving the geographical names to you; Lebanon and Anti-Lebanon / Hermon are enough for me to mention."

His article "The Near East Background of the *Rigveda*"[2] gives stern notice to all Indologists that "The guidelines [he set forth] above are applicable not only to Scripture but to all ancient texts."[3] The *Ṛgveda,* abounding in horses, chariots and iron, and containing many references to the sea and sea-going vessels, can hardly be a local development on the plains of the Punjab.

Professor Gordon's last words regarding his works were, "I have done pretty much what I wanted to do." Indeed he finished the revisions of all of his major works and his autobiography, *A Scholar's Odyssey* (2000), for which he won an award from the Society of Biblical Literature. His legacy, rooted in facts, will no doubt play a signal role in bringing a new paradigm for twenty-first century humanity, particularly for the study of ancient languages and cultures.

2. In *Ancient Egyptian and Mediterranean Studies in Memory of William A. Ward* (Ed. by L. Lesko; Providence: Dept. of Egyptology, Brown University, 1998).
3. *Shofar* 12.4: 36 (1994).

JOSEPH HAROLD GREENBERG: A TRIBUTE AND AN APPRAISAL

by Harold C. Fleming
Gloucester, Massachusetts

I will eschew a standard obituary format because, for those interested in the biographical details of his life, there are, have been, and will be ample sources of these published. Above all we had Paul Newman's long biographical interview with Greenberg from *Current Anthropology*, paraphrased in *Mother Tongue*: The Newsletter (1991). More recently, Nicholas Wade had a long obituary in the *New York Times* (May 15, 2001) which did an excellent job.[1] And some aspects of biography will be included herein.

THE TRIBUTE

When an important man of science leaves us, we think of how much we benefitted from his work, how much he stimulated our work, and how things will be now that he is no longer working among us. Perhaps the most important first question is just how important was this scholar anyway? Apparently one would not be alone to say that Joe was a great scientist, easily one of the three or four most influential linguists of the 20th century and easily the top man in the genetic taxonomy of languages who ever lived.

His outstanding work on typology is not included here because I paid it little attention; yet it is clearly another kind of comparative method, one more familiar to anthropology and the other social sciences. So Greenberg excelled at two kinds of comparative strategies, the one genetic and historical or simply diachronic, and the other synchronic or achronic, closer to 'the physics model' or what most philosophers of science think of as the scientific method. One seeks to find the conditions under which certain kinds of phenomena occur and thus establish general laws for the occurence of these phenomena, and test them—ideally through experiments. Since being well trained in this comparative method, as used in kinship studies and ethnographic surveys, by G.P. Murdock, the difference between the synchronic methods of ethnology and the diachronic methods of genetic linguistics is very clear to me. However, this difference is often not understood by proponents of the two strategies, such as physicists and historical linguists. There has been confusion in archeology about this difference too.

One test or confirmation of the dual skills that Greenberg possessed is given by his election to the National Academy of Sciences, the highest scientific honor one can get short of the Nobel Prize in Prehistory, which has never been awarded. He was also elected to the American Academy of Arts and Sciences. Indeed, when once I solicited his résumé, his honors from various universities and scholarly groups took up as many pages as would suffice for an ordinary scholar's entire résumé!

---

[1] Except that Wade and his informant, Paul Newman, forgot the large contribution that was Greenberg's Indo-Pacific hypothesis. Since he had already (1954) commented in depth on southeast Asia, his regrets were possibly due to not returning and settling the Austric question. Conservative Australian and British linguists have largely rejected the Indo-Pacific hypothesis which argues for its probably being right. They acted the same way in Africa. [The reference to Paul Newman's interview is *Current Anthropology*, vol. 32, no. 4, August-October 1991, pp. 453-467. Excerpts in *Mother Tongue* (Newsletter) 15, December 1991. Ed.]

Two more anecdotal pieces of evidence for the high regard that so many scientists and scholars bestowed on him are offered here. Once Frank LeBar of Yale (Human Relations Area Files) discussed the classification of Miao and Yao of southeast Asia with Paul Benedict. One key point of their discussion was whether Greenberg had changed his mind or not on that subject (which he had), because anthropologists were completely dependent on Greenberg's opinion—such was their respect for his genetic hypotheses plus his great prestige. A second example came from my own department (BU). When once I told a colleague (Anthony Leeds) of my exciting discovery of some new aspect of Afrasian taxonomy, he replied: "That sounds good. Did you check it out with Greenberg yet?"

At least until 1987 Greenberg alone probably was regarded by more social scientists as the world authority on genetic classification than any other one, or two, or three scholars anywhere. After 1987 the furious opposition to his Amerind hypothesis by Americanist linguists and some Indo-Europeanists changed much of the public scholarly perception of his work. His African work remained virtually impregnable; it had been tested for almost 40 years and had held up. His Indo-Pacific hypothesis was increasingly ignored, while the Amerind effort became a battleground. While Ruhlen defended Amerind mightily, and *Mother Tongue* devoted much of its effort to Greenberg's defense, most Americanists turned away from Amerind to devote themselves full time to nitpicking. Greenberg's methods were scorned and he was thought of as a scholar whose best days were far behind him. Very great social pressure was exerted upon linguists to conform to the critiques of Greenberg; even his own students were frightened into silence. As Joe[2] told me himself several times, the Americanists, and increasingly 'the linguists', had embraced a new paradigm, albeit a mistaken one, and they disappointed[3] him a lot.

Yet, despite his advanced age (ages 75 to 85), he devoted the 1990s to work on the Eurasiatic hypothesis, finishing a first volume on comparative morphology a few years ago and then completing its matching lexicon just before he died. When doctors told him that he had uncurable cancer (in mid-winter) and would soon die, he shrugged and kept on working until he and Merritt Ruhlen had finished the lexicon. Nicholas Wade reports that Paul Newman asked Joe shortly before he died what his greatest regret was. Joe said it was his failure to finish up southeast Asia! Merritt and I saw him not long before he slipped into his final coma, making us the last scholars to talk with him. Yet even then his mind was clear and rational. Thus I mentioned that Gilyak had a word, roughly /irf/, meaning fox or jackal, and that it was found in various places in western Eurasia and north Africa. (This was in a context of discussing Karl Bouda's work on Gilyak). Joe said it was really something like /iRf/ because the *r*'s were like German or French *r*'s and that it was part of an etymology in the Eurasiatic lexicon. What a memory! What a scholar! And what a shame that his vast and unique knowledge of human languages had to leave

---

2. Greenberg has always been called 'Joe' among anthropologists. That is a tribute to his warm and modest behavior and the obvious affection with which he was regarded. We will stick with Joe from now on.

3. 'Disappoint' is used deliberately. Joe was extraordinarily rational, non-confrontational, and mild-mannered. As I told him just before he died, "I am the emotional one, you the rational; so I am going to tell you how I *feel*." Once, when Joe was being attacked by British linguists and Semiticists— as usual—, Dan McCall asked him why he didn't fight back. Joe replied that evidence would decide the matter in due course.

us, could not be electronically stored, and that such a great scientist had to die under a cloud of misguided criticism!

We can ask how anyone on their deathbed could be thinking about Eurasiatic or even Gilyak etymologies. The answer I propose is that this was what he was good at, this was what he loved, and indeed, this was his life. To his core Joe was an exemplary old-fashioned or traditional historical linguist. During our farewell visit (*Abscheidsfeier*), he said that he had begun thinking on his own about language when he was 12 years old. Nobody told him about it but he noticed phonetic patterns in English, his own and others, and puzzled it until getting a conclusion. Self-taught at twelve.

Who then trained Joe in linguistics? What school did he represent? A well kept secret perhaps, but the answer is—NOBODY. He was trained in cultural anthropology, did his field work in Nigeria on acculturation to Islam, and wrote his dissertation on that topic. He greatly admired Edward Sapir, whom he resembled cognitively, but he never studied with Sapir. He read Sapir and the great books on Indo-European and a great many works of 19th century German scholarship; from his youth he had read grammars for pleasure and remembered them. It helped that his mother spoke German, which thus was far easier for him than for the rest of us Americans for whom scientific German was a chore.

Three years ago at a conference in Baltimore one of our Russian colleagues in an excess of Neo-Grammarian zeal proclaimed that "Greenberg is an amateur!" (The same for Ehret and probably me, when I left the room). The accuser, Militariev, was the Russian who had first contacted me in Moscow in 1986, thus a co-founder of ASLIP was he. At least four times at that conference Militariev contrasted 'amateur' with 'professional'. Well, a professional was guided by Neo-Grammarian principles, while an amateur was a lower form of life. Despite my irritation at his arrogance, I later slowly realized that there was some truth lurking in his contemptuous remarks. In a sense Joe was an amateur, not having been trained like an apprentice by a master linguist and not having had his 'mistakes' (deviations) corrected repeatedly. This is not to say he never had a course in linguistics, never talked to a senior linguist, and never was told how linguists do their thing. He did have a little of that, but overwhelmingly he was self-taught—by reading voraciously and by thinking. Or, by using the common sense for which he was famous. And oddly enough the other two amateurs, Ehret and me, were very much the same; Chris was trained to be a historian but took a few courses in linguistics. I had two semesters with a trained linguist, Lounsbury, but that was all. (The course was required for all first year graduate students at Yale.) I was trained to be an ethnologist. Likewise some of our most productive or creative long rangers—Bomhard, Bengtson, Hayes, Whitehouse—are amateurs in the Greenbergian sense.

What was most astounding about the life work of Greenberg was not so much the ground he covered—which was immense—but the singularity[4] of his contribution. His ventures or hypotheses extended from 1948 (the first African articles in *Southwestern Journal of Anthropology [SWJA]*) to 2001 (the final lexicon of Eurasiatic); fifty-three years' worth of scientific creativity, i.e., hypothesis formation. In a moment we will list the noteworthy points where he extended our knowledge of linguistic prehistory in fruitful and reliable ways. For now, however, it is appropriate to ask: during those 53 years when Joe's cognitive fingers probed into prehistory, where were the professionals? What were they doing? What hypotheses about our

4. This notion is borrowed from Nicholas Wade's obituary wherein he referred to Joe as a "singular linguist."

common past did we get from them? Precious little, bloody little, damned little; what you call it depends on your dialect. Granted there were some active scholars, but we are not obliged to name each one because none of their contributions were both as extensive and as reliable as Joe's. There is always what linguists generally call the 'lunatic fringe' where individuals will propose daring hypotheses that usually fail to stand elementary testing or just get ignored. Mostly Europeans, their names include such as Mukarovsky, Bouda, and Pinnow, some of whose ventures are just now getting accepted.[5]

Two major exceptions to these conclusions about professionals exist. One is the work of Morris Swadesh in North America and the other the work of Illich-Svitych in Russia, with his colleagues Dolgopolsky and Dybo, and their students. Swadesh brought the concept of linguistic dating to fruition, although the resistance to his glottochronology was even fiercer than the attacks on Greenberg. Swadesh also attempted the ultimate taxonomy of the world, as Trombetti had done, but his efforts were not reliable and his mass of etymologies apparently never got published. Swadesh was an anthropologist but also a true professional linguist who did fine work on Amerind languages and some theoretical linguistic concepts (e.g., the phoneme). He and Joe worked together for a while in their youth but clearly were not good friends.

In Moscow the original daring work on the 'lunatic fringe' by Pederson on the relatives of Indo-European was checked, accepted and expanded by a small and highly creative group of young Russian professionals. They and their students produced Nostratic and later Dene-Caucasic, both bold and exciting ventures. They backed up their boldness by attempts at strict phonological controls, elaborate reconstructions, but massive compilations of data. More than anyone else, even including Greenberg, they broke the stranglehold of Indo-European exclusivity, the unacceptable notion that Indo-European had relatives, especially in the Mongoloid realms of the east. The resistance to this Russian work was much softer than that to Greenberg, but final acceptance has not yet arrived. Joe's work might help to push the matter over the threshold. Dare I say that the Indo-Europeanists seem more tolerant or more rational than their colleagues in the Americanist 'mainstream'? Sure, cultural anthropologists can say that sort of thing. Why not? 'Tis true.

Let us sum up Joe's singularity, why his contributions just dwarfed anyone else's and why he accomplished more than hundreds of American professional linguists *combined* in his 53 years of hard work.

AFRICA Between 1948 and 1963 he reviewed the literature on African languages and taxonomy, fought free of widespread European racial superiority assumptions, broke the bond between physical type and language genetics, and put some 1,500 languages into four large taxa where almost all have stayed ever since. Despite the belief among some woefully ill-informed American linguists that African languages are close to each other, like Bantu ones are, there are huge differences in phonology, morphology and vocabulary. In all of the phyla lexical retentions on a Swadesh list get down to 1%, for example; just in Afrasian (formerly Hamito-Semitic)

---

5. For example, Pinnow's old contention that Haida belongs to Na-Dene as a coordinate was accepted by Greenberg, but only this year by some Americanists, and not yet by the Russians.

between Berber and Omotic languages or Berber and South Cushitic languages we reach that low percentage. Or in Niger-Congo between West Atlantic (e.g., Peul or Fulani) and Kordofanian. Or between North Khoisan and South. And so forth.[6] Within each of major phyla (families) relationships often get quite remote. In some cases the remoteness leads to the relationships being questioned. For example, Songhai within Nilo-Saharan, Omotic within Afrasian, Hadza and/or Sandawe within Khoisan: each has been challenged—ultimately unsuccessfully.

SOUTH & SOUTHEAST ASIA In 1954 *Anthropology Today*[7] published a state of the art book of theory, involving the four sub-fields, with articles written by leading scientists in special fields. In the book Joe published a theoretical piece on new methods in historical linguistics. But he included in that a survey of some areas with unsettled questions. Having looked over the literature and much of the data, he ventured opinions that *faute de mieux* added up to a taxonomy of most of the world. In South Asia he agreed with traditional phyla such as Indic, Dravidian, Munda, and Tibeto-Burman but also stipulated that Nahali (Nehari) was distinct. He supported Paul Benedict's separation of Thai-Kadai from Sinitic and Pater Schmidt's creation of a large phylum called Austric. Joe missed Kusunda in the Himalayas, which was easy to do since it was buried in masses of Tibeto-Burman material in Grierson's *Linguistic Survey of India*. Almost everyone else missed it too.

THE SOUTHWEST PACIFIC & THE INDO-PACIFIC HYPOTHESIS While not venturing much outside of Malayo-Polynesian (Austronesian) and the mainland phyla, Greenberg was clearly stimulated by the problems of Melanesia and Papua, and later Australia and Tasmania. That area which we now suspect contained the first emigrants from the *Homo sapiens* homeland in Africa has linguistic and cultural diversity to match that of Africa or the New World. And it is OLD! Some evidence is found in the not-quite resolved archeological dates for Australia for 40,000 to 60,000 or more. But other and in some ways more interesting archeological dates come from insular Melanesia, where dates of 38,000 more or less are found. That settlement had to be sea-borne and most probably came from Papua, long before anyone would seriously propose that Austronesian sailors were involved.

Joe took twenty years to examine the hundreds of languages that physical anthropologists usually called the 'NAN' peoples, the non-Austronesian peoples of the western Pacific and Indian Ocean. Roughly the region from the Andaman Islands to Fiji, and from Tasmania to the Admiralty Islands, was the domain of his inquiry. Naturalists have observed of flora and fauna that northern climes have fewer species but larger populations while tropical climes have more species but smaller populations. That observation aptly portrays the human language situation, especially in Oceania, Africa and Latin America. Add to that the older biological conclusion that modern man is a tropical animal who has adapted culturally to northern climes.

By Ruhlen's count in his *A Guide to the World's Languages* (1991 edition) there are 731

---

6. Sergei Starostin has maintained that two languages having less that 5% on a Swadesh list should not be put in the same family. That is a serious confusion of mathematical probability thinking and the bases of linguistic classification which are not limited to Swadesh list vocabulary and include grammar, etc.

7. *Anthropology Today* was edited by A.L. Kroeber, at that time arguably the most influential anthropologist in the USA. Anyone studying for their comprehensive exams in anthropology felt obliged to read it.

cognate languages that are neither NAN nor Australian in this Oceanic realm. Joe proposed calling them Indo-Pacific, after his customary use of geographical terms to label linguistic taxa; he finished his classification in 1971.[8] Perhaps the biggest surprise of Indo-Pacific was its inclusion of Andamanese and Tasmanian, as far apart geographically as Berber of Morocco and !Kung of the Kalahari. A second surprise was separating Tasmanian from the Australian phylum a short distance away on the mainland, yet joining it to Papua a whole continent apart. A third point, although not so surprising, was Joe's refusal to link the Australian and Indo-Pacific phyla together. Had he done so he would have proposed the oldest linguistic taxon on earth, remembering those archeological dates above. Its African equivalent would be to link Afrasian and Khoisan or Nilo-Saharan with either of them.

It is noteworthy that Greenberg observed limits, i.e., he has never formally proposed a taxon older than his African foursome or Indo-Pacific. But he has suggested in a number of places that there were probably older taxa around, for example, Afrasian and Niger-Congo, Khoisan and Afrasian, Amerind and Eurasiatic, etc. He simply lacked the time and energy to try to establish them. And perhaps, considering the furor most of his proposals elicited, he was just tired of being yelled at!

Thirty years after Joe's Indo-Pacific proposal one cannot say that it has been accepted. No doubt some scholars, probably mostly anthropologists, quietly believe it is a viable hypothesis. No doubt some others consider it 'unproven' or foolish or the like. They are likely to be Australian or British, but they are usually quiet about it, not abusive. Mostly Indo-Pacific has been ignored, nearly to death. One major hope is offered by Paul Whitehouse (London) who is embarking on a grand review of the numerous new data on NAN languages plus the old etymologies Joe proposed. Since he has already convinced himself that Joe was right—that Indo-Pacific is viable—the future looks brighter for this somnolent piece of prehistory.

THE NEW WORLD AND THE AMERIND HYPOTHESIS In anthropological linguistics of mid-20th century social science a large part of scientific activity was focused on the Americas. With the work of Sapir, Kroeber, and indeed most of the Boasian school of anthropology being influential, and the ready availability of local informants, the work most people heard about concerned Native America. The only other major focus was on Indo-European; that existed primarily outside of anthropology and had a considerable tendency to play by its own rules. There were few departments of linguistics, and indeed perhaps only twenty of anthropology.[9]

---

8. Cf J.H.Greenberg, 1971, "The Indo-Pacific Hypothesis." *Current Trends in Linguistics*, volume 8. Ruhlen reports manuscripts written in 1958 and 1960 but unpublished.

9. In the post-World War II atmosphere, especially with massive governmental stipends to students or the GI Bill of Rights, the number of anthropology departments increased rapidly. By 1970, when this trend was aborted fairly abruptly, anthropology departments numbered slightly more than 100. Linguistics departments picked up steam later than anthropology but also were abruptly cut off by the early 1970s, which saw linguists increasingly seeking employment in anthropology departments. The new departments had produced too many PhDs and the glut contributed heavily to the aborted growth curves.

After finishing his training at Northwestern University in cultural anthropology under the dynamic Melville Herskovits, Joe went to Africa to do his field work in cultural anthropology (ethnology). When he returned and got employment at Columbia University, he was already familiar with the powerful Boasian milieu from his college days. Yale was not far away, where Edward Sapir had taught until 1939, and from which his student, Morris Swadesh, came to New York to teach at CCNY. Joe and Morris came into close contact and undoubtedly influenced each other. There are anecdotes that circulate among anthropologists about the seminal Greenberg-Swadesh interaction, but suffice it to say that the languages involved were native American. The conclusion has to be that Greenberg began work on Amerind before his African classification was finished.[10] Or he was working on Amerind languages before many of his Americanist critics.

By 1960 Joe had reduced the diversity of Central and South America to far fewer phyla or families than the prevailing picture of scores of independent families in Latin America. He gave another paper on classification in 1979 and another in 1981. Six years later he published his *Language in the Americas,* which gave his full hypothesis. Later on, some additions and changes to internal taxonomy were ventured by Merritt Ruhlen, but Joe's final effort on Amerind appears to have been in 1987.

Why did it take nearly 40 years to fully classify the 583 Amerind languages into one phylum while the much greater number of African languages were classified much more rapidly into four phyla? Actually, the full and final African classification took 15 years, waiting in the last ten years for the deeper linkage of Niger-Congo to Kordofanian and the gathering together of several independent families into the Nilo-Saharan phylum or super-phylum. There seem to be five primary reasons for the greater amounts of time required for Amerind.

1) In Africa the 'transitivity principle' was easier to apply, because there were broad stretches of closely related languages adjacent to areas with more distantly related languages in turn adjacent to other related languages. To take the extreme example of Bantu we find hundreds of closely related languages spread over an area as big as the USA west of the Mississippi. That joined to a so-called Semi-Bantu or Bantoid in a much smaller area; that in turn to other groups in southern Nigeria; and so forth. The basic principle is a matter of logic. If A is related to B and B is related to C, then A is related to C. That principle fit most of Niger-Congo, Afrasian, and southern Khoisan. It did not do so well in Nilo-Saharan where languages such as Songhai, Saharan, Fur, Kunama, and Nyangeya were not only physically distant from each other but also not at all close linguistically. On the other hand it is not common in the Americas to find phenomena like Bantu or Arabic with their wide distributions.

2) In Africa the ground had been prepared by lumpers; in America by splitters. Not a few 19th century and early 20th century scholars, such as Koelle, Johnston, Cohen, Westerman, and Meinhof, *et al.,* had been inclined to assemble great globs of data or to make sweeping classifications based on a few typological traits. In a sense the job was to correct their errors, account for the ones they missed, and put it all together. For most of the African lumpers their gross error

---

10. I am indebted to Daniel McCall for conversations over many years about those famous Boasian days in New York. Dan was also one of Greenberg's first graduate students at Columbia. I would surmise that Greenberg's interest in African linguistics was triggered by his field experience in Nigeria where he found that Hausa of Chadic was not clearly grouped with many languages that were obviously related to it. Thus began the emphasis on Chadic that dominated his chapters on Afrasian (his Afroasiatic) in his first articles in *SWJA* in 1948.

rate was not so excessive. Granted, they made serious mistakes, but most of what they linked together was usually genetically true. Most of Afrasian was already laid out, Bantu and Khoisan were already in the literature, much of Niger-Congo under the name of Sudanic was in place, and parts of Nilo-Saharan. In more modern times Africa produced a generation of hyper-splitters of British origin whose distrust of hypotheses of relationship could match the amazing splitters of South America. Had Africa been left to their tender mercies it would resemble Latin America with scores of independent phyla. The received literature in the New World was much like it is now: genetic groups are small and numerous. However, Sapir and some others had gone far to modify that condition. But controversially.

3) In the New World many anthropologists found 'culture areas' in which much of the culture was widely shared among neighboring peoples. But Amerind language diversity is extraordinary in that it is frequently the case that few languages have close relatives in their own area. It is as if France, for example, contained French, German, Swedish, Russian, Hindi, and Armenian. There must have been a history or prehistory of mobility that produced such local diversity. Take the strange case of Algic with California branches separated from midwestern branches and separated from eastern branches, each of whom found themselves with seemingly unrelated neighbors, like Iroquois in the east or Penutian in the west. Or follow the distribution of the various large branches of Central or Southern Amerind in the Amazon basin. They resist analysis into homelands or geographical foci. It all looks more like a scattergram or scatter shot or a work of modern high-tone art. Actually, a major modern city like New York would be something like this if each of the entering ethnic groups had kept their language and lived in their own communities or 'tribes'. Africa is much like this in some areas but also has vast areas with little diversity. Only the Arctic in North America is anything like most of northern Africa.

4) Scholarly work and/or sources were more numerous and older in the New World than in Africa. Despite the great antiquity of two northern branches of Afrasian (Semitic and Egyptian) and some lesser antiquity for Ethiopic and Arabic of Islam, most African languages were described in the 19th and 20th centuries. A few on the western and southern coasts were contacted by Portuguese and Dutch explorers in the 17th century, but the records are not very full. On the other hand a large part of the Americas was described as early as the 16th century by Portuguese, Spanish, French, and English explorers and colonists. More to the point, anthropology and linguistics were growing up in the 19th century in Europe and North America. Just as the literature of Indo-European is a much more vast enterprise than any other of its ilk, so the Americanist literature was much larger and more sophisticated than that of Africa, which basically consisted of a few Europeans writing about African languages. So there is probably more data and grammatical analysis to read per language and a lot more reconstruction per group of languages.[11]

5) Greenberg's critics were more numerous and better organized in the Americas than they had been in Africa. While Greenberg did have severe critics among Semiticists, such as Wolf

---

11. Greenberg was criticised by numerous Americanist scholars for 'mistakes' (usually phonetic imprecision or erroneous morphological segmentations in grammar) or failures to use 'modern' reconstructions, i.e., their own work.

Leslau, most of his opponents were European. While British linguists were overwhelmingly hostile until recent times, the best European linguists were in France, Italy, and Germany; they were much less hostile and many were converted early on. In the Americas, on the other hand, Joe's critics were given years in which to decide what to do about his classifications (see above for 1960). They were a much more compact group and represented the victory of Indo-European thinking over the old Americanist ways. They were in effect organized by a series of introductory texts in linguistics that stressed methods, rigor, precision, and something like a Neo-Grammarian position. Moreover, since the Chomskyite revolution had swept American linguistics after 1957, many historical linguists probably felt threatened by the changes proposed by (1st) Swadesh, (2nd) Chomsky, and (3rd) Greenberg. Swadesh was practically run out of the country because of glottochronology, although many say the reason was his being a Communist. Chomsky himself told me years ago that he was appalled by the hostility he received from linguists. When interviewed by Nicholas Wade after Greenberg's death, L.L. Cavalli-Sforza (of biogenetics fame) is quoted as saying that Joe's critics were cruel, probably because they were jealous of his successes.

As a final note, Greenberg's Americanist critics were successful in one thing. They apparently convinced their colleagues in the rest of linguistics that Joe's work was under par and mistaken and in fact anathema. So today most of American linguistics is opposed to the Amerind hypothesis and the methods by which it was created. All of these developments were apparent to me in the late 1980s. I tried to warn Joe indirectly via an opinion in *The Atlantic Monthly,* but their editor cut the warning out for reasons of space.

The Amerind hypothesis, *qua* hypothesis, was a sweeping vision of the entire New World, since it was grounded in the notion that there were two other phyla present. By proposing that Eskimoan and Na-Dene were independent of Amerind, he contributed greatly to prehistory. With its representatives stretching all the way south to Cape Horn and eastward all the way to Laborador (Beothuk), Amerind was the obvious choice for first human occupancy of North and South America. And the great internal diversity of that taxon argued separately for a considerable antiquity of Amerind in the New World. Greenberg had decided independently that the age of Amerind in the New World was to be correlated with the archeological dates of first human entry. In association with Christy Turner (archeology) and Stephen Zegura (physical anthropology), he agreed to 12,000 BP as the likely date for that entry. Since more recent archeological research has increasingly challenged that date, the so-called Clovis horizon, Joe has not changed it. Although Ruhlen continues vigorously to defend the date, on Joe's behalf, I think it is a basic error on their part and their conclusion is being undone by current archeology.

EURASIA AND EURASIATIC While Africa is huge, the continent of Asia is even bigger. Combined with Europe, it becomes Eurasia, the largest of all the great land masses on earth. From a geographical standpoint most of Europe is a large peninsula of western Eurasia, with Arabia and India the same to the south. On the southeast the Malay Peninsula almost joins the insular world of Austronesia or Sundaland, which was cut off only when the Ice Age ended.

We have considered the southern parts of Eurasia above. The northern and western parts remain to be considered—roughly Europe, the Middle East and Siberia. The focus is on Europe because it was the one place on earth that did not seem to need Professor Greenberg poking around and upsetting things. One large phylum, Indo-European, dominated those parts, albeit somewhat challenged by Altaic. That large phylum (we will call it I-E henceforward) has been the

database par excellence for modern linguistics throughout its development. As Ruhlen is fond of saying, much of linguistic thinking is 'Eurocentric'; so too has the classification and reconstruction of proto-I-E dominated the theory and practice of historical linguistics.

While there have been numerous attempts to find linguistic kin for I-E in various parts of the world, all such efforts have been fought off or simply lapsed through being ignored. Or had been until Illich-Svitych came on the scene in the 1960s in Moscow. We have mentioned the Muscovite efforts above. For now it is enough to say that I-E was put in a genetic group that included Uralic, Altaic, Japanese, Korean to the east and Kartvelian (South Caucasic), Dravidian or Elamo-Dravidian, and Afrasian to the south. Early on it became apparent that the new super-phylum, Nostratic, did not have an accurate internal taxonomy, that Afrasian stood partly aside as a coordinate sub-phylum, and that the relationships to I-E were not well-established. Was I-E a western Nostratic entity, like Kartvelian and Dravidian, or was it closer to Uralic and Altaic to the east?

Indeed it was time for Professor Greenberg to poke around in this matter. One of his first determinations was that we needed to find 'valid taxa', i.e., those genetic groups closest to each other, even if related to others outside of that group. Thus Semitic, for example, was most probably related to I-E but neither of them were in the same valid taxon; so Semito-I-E was not itself a valid taxon. But Afrasian was a valid taxon and Semitic belonged in it. I-E did not. But instead of looking only at I-E as everyone had been doing Joe followed his own custom of looking at an area to find the valid taxa in it. Instead of being Eurocentric, he looked at the whole range of north Eurasian languages. Unlike the Nostraticists, Russian and American, he chose not to restrict the inquiry to language groups that had been well reconstructed.[12] He must have asked himself—how did I ever do Africa without reconstructions?

The result of Joe's search for the valid taxa[13] was to find a taxon to which I-E and nine other groups belonged—before they related to any outside groups. Thus I-E was more closely related to any of them—for example Ainu or Aleut—than it was related to Kartvelian or Semitic. From a taxonomic standpoint it was a neat solution, because this Eurasiatic group formed a line across northern Eurasia and had some interesting properties.[14]

Nevertheless, Eurasiatic was a shock to European sensitivities. First, it was not closest to the old civilized peoples of the Near East. Second, it was closest to people who were physically

---

12. This is an important component in the debate between the Taxonomy First and the Reconstruction First schools of thought. As we will see below, the Russian position was an Indo-Europeanist's.

13. It might be clearer to call them 'natural' taxa instead. Thus Dutch, Swedish, Portuguese, Sicilian, and Greek form a natural taxon—I-E—but the first pair and the second pair form two more natural taxa within the larger one, while Greek is by itself. Essentially, the whole discussion about valid or natural taxa is a sub-grouping problem. It would not occur if the languages under discussion had not been related to each other in overall terms.

14. From a phonetic standpoint I-E was like the rest of Eurasiatic in lacking glottalized consonants, pharyngeals or the retroflex sounds (found in Dravidian), although the Indic branch of I-E had acquired the retroflexes. The contrast with heavily glottalized Kartvelian and Afrasian is striking. It is perhaps not an accident that some Muscovites pioneered the reconstruction of proto-I-E as a glottalizing language. Joe never accepted that.

Mongoloid, *i.e.*, Orientals such as Japanese and Mongols. This immediately made no sense in prehistory because there was no correlation between people of that appearance and I-E languages. But there was a high correlation between people of European or Caucasoid appearance and I-E languages. Clearly, either somebody had changed their language in ancient times or Greenberg was mistaken. But even if we went back to the old Nostratic, some explanation was needed for the disparity between western Nostratic and eastern in biological terms. Finally, recent DNA studies make it very clear that the phenotypes of Europe are genetically determined, not due to modern climatic factors, and that neighboring peoples to the south and southeast were their closest relatives, rather than the peoples east of the Urals.

Cases where people of different languages exchange genes are common in the world. Cases where a population has changed its language but not its biology or not most of its biology are less common in the world. Modern Egyptians speak Arabic but their biology is largely derived from their Egyptian-speaking ancestors. The Ainu have finally lost their language and incorporated many Japanese genes. The Hungarians and Turks kept their languages but were absorbed by local European populations. Those ethnic groups of New York City gradually become mostly English speakers. But the more common case is where populations exchange genes and words, each becoming different from what it was but usually recognizable in physical and linguistic terms. At the moment no one has proposed a good solution to the I-E problem outlined here.

Another surprise of Greenberg's Eurasiatic has been that in recent versions of it he has incorporated Etruscan, the great mystery of old Italy. Although its precise taxonomic position was not completely clear because of translation problems, Joe thought it either a separate branch of Eurasiatic as a whole or a sister language to I-E. His final taxonomy is presented below.:[15] Eurasiatic was most likely related to other proposed members of Nostratic, viz. Kartvelian, Dravidian, and Afrasian, but with the internal taxonomy of Nostratic unspecified. Also Elamitic (suggested by the Muscovites) and Sumerian (suggested by Bomhard) were likely members. Eurasiatic proper had this membership:

```
I. †ETRUSCAN: †Etruscan
II. INDO-EUROPEAN:
     A.  ANATOLIAN: †Hittite, †Hieroglyphic Hittite,
          †Cuneiform Hittite, †Palaic, †Lydian,
          †Luwian, †Lycian
     B. ARMENIAN: Armenian, †Phrygian
     B.  †TOCHARIAN: †Tocharian A (= Eastern),
          †Tocharian B (=Western)
     D. INDO-IRANIAN
          1    INDIC: †Sanskrit, †Vedic, †Rigveda,
               †Prakrit, †Old Indic
          2    IRANIAN: †Avestan, †Old Persian
     E. ALBANIAN: Albanian
```

15. This taxonomy is taken unchanged from pages 279-281 of Joseph H. Greenberg, 1999. *Indo-European and Its Closest Relatives*. Stanford University Press, Stanford, CA. [See the review in this issue. Ed.] Capital letters for major branches, varieties in parentheses, extinct languages shown with †. Most individual lanuages and dialects are not shown.

F. GREEK: Greek (†Attic, †Doric, †Homeric, †Mycenean, †Aeolic, †Delphic, †Elean)
G. ITALIC: †Oscan, †Umbric, †Latin, French
H. CELTIC: †Old Irish, Irish, Breton, Welsh
I. GERMANIC: †Gothic, †Old High German, †Old Norse, German, †Old Saxon, English, Frisian
J. BALTIC: †Old Prussian, Latvian, Lithuanian
K. SLAVIC: †Old Church Slavic, Russian, Polish, Czech, Serbo-Croatian

III. URALIC-YUKAGHIR:
A. YUKAGHIR: Yukaghir (Kolyma, Tundra, †Omok, †Chuvan)
B. URALIC:
1 SAMOYED
a. NORTH: Yurak (= Nenets), Enets (= Yenisei Ostyak), Tavgy.
b. SOUTH: Selkup (= Ostyak Samoyed) (Tāz, Ket (sic), Tym), †Kamassian
2 FINNO-UGRIC:
a UGRIC: Hungarian, Vogul, Ostyak
b FINNIC:
i PERMIAN: Komi-Zyrian, Udmurt (= Votyak)
ii VOLGAIC: Mordvin, Cheremis (= Mari)
iii NORTH FINNIC: Saami (=Lapp)(Kola), Finnish, Karelian, Veps, Votic, Estonian, Livonian

IV ALTAIC:
A. TURKIC
1 CHUVASH: Chuvash
2 COMMON TURKIC: †Old Turkish (= Uighur), Turkish (Osmanli), Crimean Turkish, Gagauz, Turkmen, Azerbaijani, Uighur, Uzbek, Bashkir, Karaim, Kumyk, Tatar, Baraba, Crimean Tatar, Nogai, Karakalpak, Kazakh, Kirghiz, Yakut, Khakas, Sagai, Altai, Teleut, Shor, Tuvin, Karagas
B. MONGOLIAN: †Classical Mongolian, Mongol, Moghol, Dagur, Monguor, Yellow Uighur, Baoan, Kalmyk, Buriat, Khalkha, Ordos
C. TUNGUSIC:
1 NORTHERN: Even (= Lamut), Nigidal, Evenki, Solon, Orochon

```
            2    SOUTHERN: †Manchu, †Ju-Chen, Nanai,
                 Gold, Ulch, Orok, Oroch, Udihe
 V.    KOREAN-JAPANESE-AINU:
       A  KOREAN: Korean
       B  JAPANESE-RYUKYUAN: Japanese, Ryukyuan
       C  †AINU: †Ainu (Hokkaido, Sakhalin, Kuriles)
 VI.   GILYAK: Gilyak (= Nivkh) (Sakhalin, Amur)
 VII   CHUKOTIAN:
       A  NORTHERN: Chukchi, Koryak (Chavchuven, Palana),
            Kerek, Aliutor
       B  SOUTHERN: Kamchadal (Ukä, Sedanka)
VIII   ESKIMO-ALEUT:
       A  ALEUT: Aleut (Bering, Unalaska, Atka)
       B  ESKIMO:
            1    SIRENIK: Sirenik
            2    YUPIK:
                 a    SIBERIAN: Chaplino, Naukan
                 b    ALUTIIQ:  Chugach, Kodiak
                 c    CENTRAL: Kuskokwim, Nunivak,
                      Ekogmiut
            3    INUIT (= INUPIAQ): Inuit (St. Lawrence
                 Island, Norton Sound, Kangianermiut,
                 Seward Peninsula, Imaklik, Sigluit,
                 Mackenzie, Hudson Bay, Southhampton
                 Island, Labrador, Greenlandic).
```

AN APPRAISAL

When a great controversial figure comes up for historical review, two things are usually apparent. First, she/he is not likely to be as bad as the critics maintain; second, she/he may not be as good as the apostles and friends say either. That meaning of the term appraisal will have to wait for the historians or wait for the field of linguistics to settle down a little. And becoming more tolerant would not be a bad idea either.

However, being too old to wait for the mills of the historians to finish their grinding, I state my opinion; it is already obvious. Joe Greenberg is like two other great scientists whose appraisals have been mostly finished—Charles Darwin and Alfred Wegener. Both had careers similar to Greenberg's. Remarkably creative hypotheses that were crucial to the growth of their respective scientific fields but associated with vociferous, sometimes savage, criticism, nearly to the point of anathema and banishment from the scientific community, or the civilized Christian community (Darwin). We can leave Darwin's case rest in the archives because it is so well known. For Alfred Wegener things are rather different because his theories were not a threat to the dominant religion of his time and because his story happened in the 20th century. He transformed geology, or the earth sciences if you prefer, by proposing the **theory of continental drift**, which nowadays can be heard on the evening news explaining things as different as the earthquake potential of California to the steady upward rise of Mount Everest year by year. Wegener was right; his hypothesis was correct. And almost all of his contemporaries who scorned or ignored him were wrong.

In my youth, when people who lived during the First World War were still numerous, there was a favorite saying: "Forty million Frenchmen can't be wrong!" It reminds me of one of the favorite sayings of contemporary linguists (e.g., Ives Goddard, *et al.*) that since 'mainstream' linguists disagree with Joe, he (Joe) must be wrong. The Wegener and Darwin cases suggest that the 'mainstream' can be wrong and surely is wrong quite often. Because, you see, there is no real scientific logic to either the 'mainstream' or the 'forty million Frenchmen' argument. The entire populations of Texas and Florida may believe that the 'lost continent of Atlantis' has been found in the Caribbean Sea. But those 34 million opinions are like the smoke in the air over Houston: a good rain will wash it away. What will determine the debates over competing hypotheses will be the data and analyses that accompany them and test them. For that is how science functions in the long run; temporary passions and biases slowly but surely lose out. Racist theories about human differences were in vogue, dominant, 'mainstream', a century ago. Yet they did not survive a hundred years of anthropological research that destroyed their credibility. True, millions of white people still believe in their innate superiority, but those opinions are considered simple bigotry. The mainstream moved away from them. Yes, sometimes the mainstream is right!

That being the case, we must move to the argument proper, instead of popularity contests. What are the empirical and theoretical issues between Greenberg and his Americanist critics? Before tackling them, however, we have to face a fact. We cannot resolve this dispute the way physics could; we cannot experiment and we cannot work it out mathematically.[16] But remember that Darwin's and Wegener's hypotheses were basically prehistorical problems, not quite like those of physics. Each had a key problem that could be addressed by controlled or focused observations. Darwin had two key general questions: (a) is there evidence of evolution, i.e., can any species change into another? and (b) what evidence is there from the past that some species have changed? Wegener's dual question could be put as: is there evidence that continents move or have moved in the past? Besides, the fact was that much of the evidence for either Darwin's or Wegener's theories was physical, biological, and sedimentary stuff. Furthermore each could find material evidence preserved from the past in the form of fossils and/or rock formations. Except for the late-occurring writing in a few areas, linguists were stuck with contemporary evidence of socio-psychological or cultural events.

Yet the three fields were not so different as one might think. They created evidence of the past by hypothesis. While paleontologists and archeologists regard their prehistoric data as solidly factual, we know that is not entirely true. One gazes at a slab of rock and concludes that trilobites lived here several millions of years ago. That is not a fact; that is a hypothesis. A colleague of mine once looked at a bone at a site in Kenya and called it a cow, thus exciting everyone. His 'fact' was later found to be a native antelope, much to his chagrin. A linguist looks at French *chien* and Italian *cane* (among other words) and concludes that they come from the same ancestral word, something like ***kian* or ***kan*, for 'dog'. He too has created a prehistoric fact by hypothesis. His 'facts' are called 'reconstructions', but they are also based on preliminary

16. Thanks to Murray Gell-Mann for pointing out this difference between physics and historical linguistics. He made this point during a conference at the Santa Fé Institute in December, 1997. Contemplating the squabbling among linguists, he said: "We can't have this kind of problem for very long in physics because someone will make an experiment and settle the matter."

hypotheses—that *chien* and *cane* have a common ancestor—which usually are called etymologies or related forms.

Now we have come to an important difference between Greenberg and his critics (both American and Russian). Joe and his critics both start with basic facts, i.e., the words, phrases, and sentences recorded for each language and written down in interpretable symbols. In other words what most of us call the 'data'. When a linguist establishes her data base in several languages, she then begins to compare the sets of data with each other. Straight away, however, differences in approach occur. Some Russian linguists will set aside or disregard data from a language if its ancestor has not been 'reconstructed', believing that the quality of the facts is more important than their existence. Many American linguists, while not disregarding unreconstructed data, will still regard that kind of data with suspicion. Both share a belief that basic facts are not as reliable as reconstructions. Many contemporary theoretical linguists regard the basic facts as unreliable because they are 'realized' versions of the true 'underlying' facts. The attitude is amazingly similar to Indic religion in its belief in *maya* or sensory data as 'illusion'. To them the truth must be found behind the surface data.

As mentioned above, Greenberg was a traditional or old fashioned historical linguist. He took the basic data from every language, whatever the condition of its recording, and compared it with the others in the region he was working on. When possible he searched for old sources (or those in different scholarly languages) in order to get more basic facts on any particular language. His famous method of 'mass comparison', lately called 'multilateral comparison', was grounded in a dislike of 'binarism', comparisons involving only two languages and not the whole available set. To Joe it was more important to confront all the phenomena than settle for a refined pair.

So we have the basis of the first set of criticisms. Greenberg used poor data and overlooked some of the finest reconstructions in existence, said Americanists. Greenberg did not use phonetic precision, said some English critics, many of whom were pre-phonemic in their thinking and often ignorant of standard I-E procedures. Joe basically shrugged, partly because many etymologies had been established long before the modern high-quality analyses had been made. Much of this criticism was grounded in the I-E dominance of most linguistic departments. Stemming ultimately from German high standards in culture, eventuating in the Neo-Grammarian demand for exact correspondences without exception and culminating in American theoretical schools adopting that demand, linguistics became obsessed with the demand for 'rigor', precision and nearly mathematical exactitude.

The next criticism found his critics putting carts before horses. It was that Joe proposed etymologies (cognations) that were not grounded in precise reconstructions. That is, one has to reconstruct the ancestral forms (words, grammemes) before proposing the relationships. But in fact one cannot have any reconstructions before one has established etymologies. First, one must propose that *chien* and *cane* have a common ancestor; then one may propose sound correspondences (like French *ch* often corresponding to Italian *c*), and then one can reconstruct the ancestral **kan-*. Eminent theorists of I-E reconstruction technique, such as Hoenigswald, missed this point and for a good reason. They were used to having the etymologies in hand because of I-E reconstructions. Had most of them worked on Amerind or African languages in the field they would have realized the obvious: no established etymologies, hence no reconstructions, were available. One had to make them up out of raw data!

Furthermore, one could not make accurate reconstructions until one established priorities of relationship among languages being examined. If we agree that French *chien* and Italian *cane*

are cognate, but also join German *Hund* to that etymology (cognate set), then the reconstruction will change. But if we realize that French and Italian are part of a special set in which German is not involved, then we can still get **kan-* for French-Italian and something else for German-French-Italian. That is, first we find the Romance level and then the I-E level.

All this is based on the historical flow chart that is sub-classification. Those in the same sub-group have shared historical experience peculiar to them. French and Italian were descended from Latin, the dominant part of Italic which came from PIE (proto-I-E). Latin had altered the PIE word for 'dog' into *canis*; that was the source of **kan-*. German on the other hand, along with English and others, had a different history. Their ancestral word was **hund*, itself from original PIE **kwon*. At root each of these groups came from the same language but their individual descent lines made a difference in reconstruction. At root one can say that those who do not have an accurate internal taxonomy for a family are not likely to get as accurate reconstructions as they could with good taxonomy.

Such is the basis for the 'Taxonomy First' argument, as opposed to the 'Reconstruction First' group. The argument is in many ways a flat-footed difference between Indo-Europeanists and more practical minded scholars from the realm of unwritten languages in Africa, Oceania, and the New World. Again one can ask what portion of the languages of the world have written ancestors—as many do in Europe—against which to check reconstructions? The presence of old written languages has been a major factor in the evolution of I-E ways of doing things, as it has also in Semitic. But I-E and Semitic number maybe 200 languages out of a world total of 5000 to 6000 depending on whose count one takes. That is to say, 3% or 4% of the world's languages should establish the methods, the strategies and tactics, of historical linguistics? Why?

Another bone of contention between Greenberg and his critics has been the time frame of possible classification and reconstruction. It would appear that some Americanists have made up out of whole cloth a cut-off time of 6000 years, plus or minus a millennium or two for different 'theorists'. The reason given for this cut-off is that after such a length of time the evidence of relationship would have disappeared or become insufficient for accurate work. As far as I can tell, the first Americanist to propose this theory was Terrence Kaufman.[17] But ultimately the trail of this theory goes to Winfred Lehmann at Texas and then to its apparent source in an article that M.L. Bender wrote in 1976. That article, based on glottochronology or lexicostatistical tables, calculated the amount of vocabulary two languages would share after so many millennia. Bender's undergraduate degree was in mathematics (Dartmouth) and he was appalled by the low retention after 10,000 years (1%) or even 6000 years (7%). How could anything substantial be left to work with when over 90% of the vocabulary had been lost? He asked. So he and others such as Lehmann generalized the lexicostatistical conclusion to mean that most of the evidence of relationship was gone after 10,000 years or sooner.

It was a tremendous mistake in reasoning that was quickly pointed out to Bender[18] before

---

17. Unfortunately, I cannot find the original source. My knowledge is actually based on a personal communication from Kaufman. Pittsburgh, 1991.

18. Explicitly, Paul Black (Yale PhD 1975) and I reacted strongly and with dismay to Bender's article. Black wrote a nine page critique and I sent a shorter one. Bender's conceptual errors were blatantly obvious.

he published the article. Nevertheless the article was published. We should have shouted it down but we were too busy. What were the mistaken assumptions?

1) The automatic presumption of a binaristic scene. Had he not done so he would have realized quickly, being a good mathematician, that each new language added to the comparison increased the number of common retentions left over. So if three, rather than two, were compared, then 3% would be left after 10,000 years instead of 1%. If four languages, then 6%; if five, then 9%; and so forth until twenty languages yield 62% after 10,000 years.[19]

2) The assumption that, since the Swadesh list of 100 or 200 words was the most conservative vocabulary in any language, the loss in the rest of vocabulary would be even greater. That is probably true. But he forgot that the general lexicon is far more numerous than the Swadesh lexicon. Just suppose that general lexicon is 'lost' twice as fast as basic vocabulary. So after 10,000 years two languages would have only one word retained in common, or 1% of 100 words. Therefore there would only be 0.5% of general vocabulary left. Yes, but general vocabulary probably consists of 5,000 or 10,000 words. So 0.5% of that would be between 25 words and 50 words after 10,000 years. That is still binary, between two languages only. More languages would yield more; for example, three languages for 10,000 years would yield 75 words. But twenty languages at 5,000 years would yield 1,550 words.[20]

3) The misconception of what 'loss' or 'retention' meant in the Swadesh list. He forgot that words are 'on' the list when they are the dominant form in a language. But they may still be in the language and not even far away. There are many examples of this, as between English and German, for example. Nowadays English *dog* and German *Hund* are on the list, so the native English form was 'lost'. But it is still there as *hound* for hunting or sporting dogs. The same for *bird* in English whose older *fowl* is still in the language and cognate with German *Vogel*. Sometimes, of course, a word remains in the language, although it is lost until reconstruction or at least good etymologies have been made. Thus German *klein* for 'small' is cognate with English *clean*, while English *small* is cognate with German *schmal* 'narrow'. This leads to the well-known rule that as etymologies, sound correspondences, and reconstructions increase in number they make possible the discovery of more 'lost' words.

4) Most meaningful of all was the misunderstanding of what genetic classification consists of, or at a minimum what Greenberg did when he classified languages together. As is well known, there is a streak of extreme preference for grammatical evidence among Semiticists and Indo-Europeanists. Of course, not all share these extreme predilections in those fields, but their

---

19. These percentages are taken from Table A.2. "Recoverable Vocabulary Based on a Homogeneous Replacement Rate." In Joseph H. Greenberg, 1987. *Language in the Americas*. Appendix A, 341. Stanford University Press, Stanford, CA. Allowing for the Joos function or the 'dregs effect' yields even higher percentages. The reader is warned, however, that a few errors exist in the tables at the higher ranges (years) due to simple clerical mistakes.

20. See *Mother Tongue* (Newsletter) 24 for a longer discussion of this point, including the retention possible with 40 or 80 languages.

21. Only Chinese and languages like it prevented this procedure, as Greenberg repeatedly acknowledged, because there was a lot of syntax but not much morphology. Only recently with George van Driem's work on Sino-Tibetan pronouns have we broken out of that bind.

representation is numerous. Perhaps my friend, Robert Hetzron, was a prime example among Semiticists. Anecdotally, we hear that the inclusion of Celtic within I-E was held up for a long time because Celtic lacked 'crucial' morphological evidence.

Against that background we must realize that some scholars under the influence of Swadesh developed a strong preference for the lexicon, i.e., for words rather than grammemes. How else to account for Bender, Kaufman and others who used only lexical information to establish the alleged time frame or time limit of 6000 to 10,000 years? Yet all one has to do is look carefully, or to scan rapidly, Joe's African or other classifications to see that he *always* began with grammatical evidence in setting up his genetic groups.[21]

The principle can be stated quite clearly. Two or more languages can be classified together when the investigator finds enough etymologies involving basic vocabulary, general vocabulary, and grammatical morphemes (grammemes) to convince her that these languages had a common ancestor. Arguments from syntax, phonology, racial similarity and typology have turned out to be unreliable and misleading, so Joe didn't use them. Even such a striking thing as the presence of click phonemes does not necessarily lead to valid taxa, as the vivid case of South African Bantu languages can testify.

Even if the Swadesh retentions get very low, it does not follow that no evidence of relationship is left. That remains an empirical question, not one to be decided by lexicostatistical theorizing.

From this Greenbergian view it has been curious that Paul Benedict was able to hold up the achievement of phylum *Austric* because it "only had morphological evidence." So any Semiticist would have said that was fine and dandy! What is obvious about Austric with so little vocabulary evidence (allegedly) is that it must be very old, comparable to the African phyla with their low percentages of lexical retention.[22] But L.V. Hayes *does* find lexical evidence.

While many minor objections to Greenberg's work no doubt exist, the major ones seem to have been addressed. One remaining question is: why such vehemence, such fury, in attacking a mild-mannered scholar trying to help science understand our common prehistory? Is Cavalli-Sforza right in stipulating 'jealousy' as a motive for the attackers? Was this the same syndrome as that displayed by the Algonkianists in halting Sapir's work? Or is this just normal science in a field that cannot do experiments? It seems clear that Greenberg represented a throwback to an older paradigm of historical linguistics, an older Americanist and 19th century Indo-European tradition that threatened the new high-tech, rigorous, theoretical paradigm trying to establish itself as linguistics.

Next year, and probably sooner, a conference will be held to address many of these Greenbergian questions. Let us hope that the historians of science, as well as philosophers of science, become interested in the topic. It is truly interesting.

* * *

---

22. When Sheila Embleton and the rest of us get linguistic dating back on its feet, we will probably find that Austric is closer to 20,000 years old; that just figures from the great age of Homo sapiens in southeast Asia and the very low lexical factor. L.V. Hayes' etymologies then will be very valuable in our work.

References:

Bibliographical items have been kept to a minimum. I thought it more valuable to outline the issues, knowing that the specific authors and writings were fairly well known, than to produce five more pages of dense citations. HCF.

Festschrift for Roger W. Wescott:

Paleolinguistics: The State of the Art and Science

Introduction

by John D. Bengtson

In the Introduction to *Mother Tongue* V (1999), editor Roger W. Wescott wrote as follows:

> "The disagreements among the twelve authors of the first nineteen selections in MT-V seem to me to spring primarily from the philosophical divergence between absolutists and relativists. The absolutists appear to regard some genetic connections between languages as indisputable and others as inconceivable. The relativists, by contrast, tend to regard all such connections as possible, but only some as probable. The relativists, moreover, seem to treat probable affiliations as differing in degree, some being more probable than others. ... The question of scientific evidence for postulated cognation depends, clearly, on the definition of science. For mathematicians, science consisits primarily of logic. For chemists, science consists primarily of experiment. For linguists, science is rarely either of these. But what is scientific linguistics? I hope that a future issue of *Mother Tongue* may be devoted, at least in part, to a discussion of this question."

Roger's hope is now a reality. The editors sent a general invitation to many scholars, asking for articles addressing the following questions:

- What can we learn from a historical overview of the field of paleolinguistics?
- How, if at all, do methods of long-range comparison and traditional comparative linguistics differ?
- What core philosophical differences (e.g., absolutism vs. relativism; monogenism vs. polygenism; lumpers vs. splitters) exist among linguists?
- To what extent is paleolinguistics an art and to what extent a science?
- What constitutes scientific evidence in paleolinguistics?
- What time depth is possible in paleoliguistics?
- What should be the rules of discourse and polemics for paleolinguists?

Nine writers submitted contributions. The articles are arranged below in a sequence that roughly corresponds to the seven questions posed above, though some of the contributions cover more than one of the questions. The final article by Professor Blažek is included as an example of scientific paleolinguistics focusing on lexical comparisons.

Historical Overview: Roger Wescott and Paleolinguistics

by Gyula Décsy
University of Indiana

People seldom think of the remarkable fact that the number of professional linguists today – even if decreasing compared with 1960s and 70s – is incredibly high both in the U.S. and worldwide. All 3,500 or so American universities and colleges have English Departments; each of them has at least 8 to 10 Anglists. If we add the Modern Language departments and the Linguistics departments of larger universities, we come easily to 60,000 professional linguists in the U.S. No other country comes close to this enormously high number. The total number of professional linguists worldwide may be around 100,000, and one can easily get the impression that 99 percent of them accept as the sole methodological background for their work the "Modern American Formalism" introduced in the 1960s at the Massachusetts Institute of Technology (MIT). By the turn of the 21st century that methodological framework expanded to an almost global ideology, which we refer to in this paper as the MIT school. It follows that there may be as few as 1000 linguists (600 or so in the U.S. and 400 in other countries) who are not followers of this school.

As Modern American Formalism grew in importance it became entrenched to the near exclusion of other approaches in linguistics. To this day linguists who do not follow it often are defamed and their work ignored. It is important to know how the 600 or so nonconformist scholars have survived and persisted in their fields during the hegemony of the MIT school, long supported by the huge professional organization of the United States, the Linguistic Society of America (LSA). To find an answer to this question, one must examine the life and work of Roger Wescott. A task for linguistic historiography is to find out if any substantially important work has been done in linguistics outside of, or even in opposition to, the ruling MIT school. The most important achievements include, among other things, language origins research.

1. Language Origins Research. This term was created by Roger Wescott and his friends (Stokoe, etc.) at the beginning of the 1970s. (It went to Europe belatedly, in 1985, with the founding of the Language Origins Society (LOS) in Cracow, Poland, led by J. Wind and B. Bichakjian from the Netherlands). The plural origins hints at the uncertainty and methodological pluralism that is so important to the discipline. It emphasizes that there is more than one possible approach to language origins (or, the origin of languages). The MIT approach,on the other hand, sees language only in its autonomous, statistical, descriptive, and synchronic manifestations. A diachronic approach to linguistic prehistory would seem a complementary necessity. However, such a trend would have run counter to the near-sacred tenets of the MIT school and the monolingual-monistic character of American language studies. Powerful ideologues in influential LSA positions barred the representatives of the heretical diachronic approach from work in financially secure positions supported by large state and private foundations. Their biases were shared for decades also by the big publishing houses in the English language area and, first of all, by the central organ of LSA, Language. During the last three decades almost all important publications on language origins research (LOR) has had to appear in *samizdat*-like journals and has remained mostly unnoticed by the central establishment of American Linguistics.

Roger Wescott's initiative concerning LOR successfully led to or encouraged the development of special affiliation with different methods (protolinguistics, preprotolinguistics); groups devoted to long range comparison, such as the Association for the Study of Language in Prehistory (ASLIP) and its journal *Mother Tongue*, as well as the Language Origins Society; and the development of Nostratics. These were private initiatives, and we know how much R. Wescott did for the success of these endeavors. His clearly articulated views, such as his hypothesis that "The preconditions of the human sound-sequential language appeared phylogenetically with the Cro-Magnon ca.35,000 years ago" became language universals.

2. Semiotics. Compared with LOR, this discipline (its main representatives today include Thomas A. Sebeok) did not pose an ideological threat to the MIT school. Semiotics is also structural and formalistic and synchronic, with a descriptive and statistical working method, even if its results are often diachronically applicable. Semiotics thus may be seen as a non-linguistic field that does not contradict the official descriptivism of the MIT school. For semiotics, the doors of its infra-American and international development were not closed.

3. Global Linguistics. The most successful representative of global linguistics is the Summer Institute of Linguistics (SIL), with its connection to the journal *Word.* The term "global linguistics," created in the 1980s, is not used generally, but no doubt this is one of the most promising linguistic fields with a bright future. It is practically irresponsible in light of the globalization of communication in the world today that leading linguists do not embrace languages of the world in their great diversity. All the flora and fauna of the remotest corners of the world are considered and classified by scientists, but followers of the MIT school and ideologues of the LSA deal mainly with English alone.

4. Traditional Comparativism. Comparativists continue the European-founded Indo-European Studies, Uralic and Altaic, Germanic, Slavic, Indo-Iranian, Baltic, etc. Their head count may be around 200 to 300 linguists. Many of them-especially members of the younger generation-try to apply the inapplicable formalism of the MIT school to traditional research subjects in order to receive recognition and financial aid or simply to be tolerated by the mainstream linguists.

The hegemony of the MIT school came about as a result of the enormous post-1945 gain of prestige of the United States worldwide. In the subsequent globalization of American technology and popular culture, everything American became fashionable: hamburger, Coca-Cola, jazz, rock and roll, Elvis Presley, TV, CDs, blue jeans, computers-all products of private enterprise based on incredible economic prosperity (a dream for other countries). Perhaps the world expected an impressive breakthrough from the United States in linguistics too, and these expectations generated the necessity to offer something completely new. Enter MIT and the transformative, generative, formalized, diagrammatic approach in linguistics.

After the Sputnik-success of the Russians in 1960, the U.S. relied more on huge, centrally-led, state-supported organizations like NASA to achieve spectacular wonders. The LSA and the groups of young, fresh, and mostly fast-trained linguists saw their chance. The 10 to 20 thousand young linguists of the MIT school received millions in government grants and tenured positions

with LSA recommendations. American linguistics thus came to demand an even more single-minded focus on syntactic structure. (This recalls not only the monistic Weisberger movement in Nazi Germany but, at least organizationally, the political terrorism of the party-led Marrism in the USSR between 1932 and 1950, when Stalin finally forebade the nonsense and reestablished traditional linguistics as an official field of study in the Soviet Union.)

The defection of linguists from the MIT school began early. Robert Hall and Charles Hockett were the first in the 1960s. Later LACUS (Linguistic Association of Canada and the United States) was founded which, however, fought the devil with the devil's method by remaining formalistic in orientation. New schools arose from the MIT tradition that refined or applied transformative/generative studies, such as (micro)-sociolinguistics, applied linguistics, pragmatics, information systems technology, and cognitive studies or psycholinguistics. Modern American formalism itself concentrated more and more on the form of description and explanation of well-known linguistic phenomena, which, ironically, became more important than the basic substance of linguistic studies-languages. Add to this the fact that linguists in developing countries lacking linguistic traditions have been trained in American universities, and one can see how the influence of the MIT school has tended to overshadow historical and comparative linguistics globally.

Roger Wescott's human and scientific greatness can best be understood in the historical and intellectual context described here. Wescott's contributions rest in part on his understanding of the non-English world (his devoted wife is of Estonian ancestry) and bespeaks hope for change. He found the right place in a world of professional imbalance and in an imposing way founded the promising new discipline of Language Origins Research on the basis of evolutionary theory. That along with his publications and service as president of LACUS and vice-president of ASLIP has earned him a special place in the history of American and Global Language Studies.

[Editor's Note: Gyula Décsy has published in two parts (in cooperation with John R. Krueger) *The Linguistic Identity of Europe* (Bloomington, Indiana: Eurolingua, 2000). This work is being reviewed for the next edition of *MotherTongue*.]

Paleolinguistics in Denmark: From the Language of Paradise to Nostratic*

by W. Wilfried Schuhmacher
Gadstrup, Denmark

"In the beginning there was only one man, Adam, and there was only one tongue and one language, the Adamish or Hebrew. And in the same way as all men have come from one, so from one language all the other. So one of the learned man has said that we all speak one language, Hebrew, changed however" (Syv 1979 [1663]: 87; translation from the Danish mine).

In 1866 the Société de linguistique of Paris forbade discussions about the origin of language. They probably wanted to exclude such opinions as that put forward by the Dutchman van Helmont in 1667, postulating that God had spoken Hebrew. More "realistic," though also on biblical grounds, had been the Dane Peder Syv's view four years earlier.

Peder Pedersen Syv, who took the name of his birth parish, Syv ("seven [hills]"), making it famous, was born 22nd February 1631 on the Syv Farm near Roskilde, Denmark (about 3 miles from my home). He studied at the University of Copenhagen, and became in 1658 headmaster of the Latin School in Næstved. And it was here that he "among the many and troublesome school hindrances" wrote his "Considerations about the Cimbrian [Germanic] Language," which he published in 1663. The following year, he began a career as parson in Hellested on the peninsula of Stevns where he stayed until his death on 18th February 1702.

As for linguistics in the 16th and 17th century, fancy played a dominant role. At that time one was interested in finding the oldest language, that spoken in Paradise. Thus, Peder Syv's etymological "Considerations" must be called quite up-to-date. And his work, however fanciful it may be (that term might also be applied to parts of post-WWII linguistics - no names!), contains even a grain of wisdom as his "Hebrew origin of Germanic," *mutatis mutandis*, has found a modern recapitulation in Theo Vennemann's "Germania Semitica," postulating that the prehistoric Germania and its Paleo-Germanic language developed under colonial Semitic dominance. (Cf., e.g., Vennemann 1998 - where it is shown that German *Volk* "people", English *folk*, Proto-Germanic **fulk-a* "division of an army" reflect Hebrew *plg*, etc. "to divide"). And even in Denmark the Hebrew-Germanic connection has seen a re-activation by Herman Møller = Möller (1850-1923), son of a German parson and a Danish countess, who at school had learned Hebrew - and who continued his Semitic interest also after he had become professor of Germanic languages in the University of Copenhagen: Too ambitious, however, was his comparative approach of Indo-European and Semitic (Möller 1911).

It was in the first decades of the 19th century that the great period of Danish historical-comparative linguistics had begun with Rasmus Rask (1787-1832), to be followed by Karl Verner (1846-1896) and Vilhelm Thomsen (1842-1927), and with a great follow-up in the 20th century by Holger Pedersen (1867-1953), Herman Möller's most famous disciple.

Holger Pedersen, professor of comparative linguistics in Copenhagen (an expert on Celtic, Hittite, and Tocharian), may today be mainly known as the creator of the term "Nostratic" that has been dug up again in modern long-range comparison: In the early 20th century, he proposed that Indo-European was related to several families in a big family he called Nostratic - originally including Indo-European, Semitic, Finno-Ugric, Samoyed, Yukaghir, Altaic, and Eskimo-Aleut. According to the modern Nostratic hypothesis, Indo-European is related to Afro-Asiatic, Kartvelian, Uralic, Altaic, and Dravidian. [I.e., the standard Nostratic proposed by Illich-Svitych. Ed.] Holger Pedersen (d. 1953) was, of course, not able to participate in the paleolinguistic revival that began in the 1960's with Illich-Svitych, Dolgopolsky, et al.

Vitaly Shevoroshkin and Merritt Ruhlen, especially, have been working toward reconstructing the hypothetical mother tongue, Proto-World, putatively spoken somewhere around 50,000 and 150,000 years ago (See also Schuhmacher 1993). According to them, all of the world's languages are very likely members of a single language family – which leads us back therefore to Peder Syv's "one language in the beginning": We all speak one language, Proto-World, changed however... .

POSTSCRIPT: To conclude the history of comparative linguistics in Copenhagen (i.e., Denmark): In 1965, when Louis Hjelmslev, Holger Pedersen's successor, had died at the age of 65, there was nobody around to take over. The chair was given to Gunnar Bech, a Germanist, to be followed by Fredrik Otto Lindeman, a Norwegian laryngealist, to be followed by Jörgen Rischel, a phonetician – each of them attracted by the fine title but unable to attract students. More appropriate linguists, either at home or abroad, were never asked. One potential candidate joined the Danish intelligence service. At the end, the chair of Comparative Linguistics was removed, and a new institute termed Institute of General and Applied Linguistics was created; however, the best people had already gone to the Copenhagen Business School.

*The text is a short excerpt from my paper "The History of Danish Linguistics" presented in Beijing at Beijing Normal University on 3rd April 2000.

References:

Möller, Herman. 1911. *Vergleichendes indogermanisch-semitisches Wörterbuch.* Göttingen: Vandenhoek & Ruprecht.

Schuhmacher, W. Wilfried. 1993. "Lingua Neanderthalensis." *Dhumbadji!* 1.3: 24-27.

Syv, Peder P. 1979 [1663]. "Nogle betenkninger om det Cimbriske Sprog." In: *Danske Grammatikere* I, 75-272. Copenhagen: Det Danske Sprog- og Litteraturselskab & C.A. Reitzel.

Vennemann gen. Nierfeld, Theo. 1998. "Germania Semitica: *plog-/*pleg-, *furh-/*farh-, *folk-/*flokk-, *felh-/*folg-." In: *Deutsche Grammatik - Thema in Variationen* (Festschrift Hans-Werner Eroms), 245-261. Heidelberg: C. Winter.

Paleolinguistics: Its Definition and Scope

by Irén Hegedűs
University of Pécs, Hungary

The notion of paleolinguistics has been used in a somewhat inconsistent manner either in a narrower sense – to refer to linguistic reconstruction at greater than usual time depth, or in a wider sense – to language origins research. This alternate usage of the term *paleolinguistics* obscures the significant theoretical distinction between these fields of study and thus can create the misunderstanding that linguists involved in long-range linguistic comparison are in pursuit of reconstructing Proto-Human (for a summary of alternative approaches cf. Békési 2000: 1f). Therefore it is desirable that the notion of paleolinguistics is clarified and consistently applied, because it is highly relevant to draw a borderline between these two types of research. Elsewhere I have already suggested a periodization of historical linguistic layers studied by long-range comparison (Hegedűs 1997: 66f). In the following table I intend to show that only the proto-, the paleo- and the archeolinguistic layers can be regarded as subjects of comparative linguistic research, and presently penetration into the deeper stages is still limited because paleolinguistic studies are yet to be developed. Judging by the progress made in the past two decades (especially in the study of the Nostratic hypothesis) the prospects look encouraging. The dates given in the table are not absolute but have significance in relative chronology.

	Primary input data	Time range	Output
Protolinguistics	from extant languages and from written documents of extinct languages	5-12 kyrs BP, Neolithic and post-Neolithic cultures	reconstructed proto-languages of language families (e.g. PIE, PU, etc.)
Paleolinguistics	reconstructed protolanguages of well established language families (PIE, PU, etc.)	12-25 kyrs BP, pre-Neolithic and Late Mesolithic cultures	a protomacrolanguage (reconstructed proto-language of a macro-family, e.g.: PNostr., PSCauc., etc.)
Archeo-linguistics	reconstructed protomacrolanguages	25- ??? kyrs BP, Mesolithic and Paleolithic culture	protoarcheolanguage
Glossogenesis	results of paleoanthropology, cognitive studies, archeology and other related disciplines	coeval and preceding the emergence of *Homo sapiens sapiens*	not a reconstructed language

One of the main achievements of long-range linguistic comparison in the past two decades is that it has managed to open up paleolinguistic perspectives by undermining the main persisting dogma in historical comparative linguistic research: the dogma of the inapplicability of the traditional comparative method at greater than usual time-depth, i.e., the dogma of the impossibility of reconstructing linguistic stages prior to ca. 7,000 BP. This

belief in the inapplicability of the comparative method at a greater time-depth grew out of the myth that lexicostatistics has a strict temporal limitation because substituting 14 of the 100 items in the basic vocabulary list over a period of one thousand years leaves us with nothing left to compare after 7 millennia of historical linguistic evolution, so it is impossible to say anything about genetic relationship between languages prior to this point in language history. The idea that the further back in time we go in our attempt to reconstruct protolanguages the fewer data for reconstruction we have at our disposal is wrong because "the number of cognates will increase significantly if older forms of languages in question are compared. The increase in the number of cognates becomes even more significant if we compare the reconstructions of related languages" (Vovin 1999: 92). Although some linguists have been calling attention to the potentials of the revised version of lexicostatistics (called "etymological statistics" or "root glottochronology"; cf. Jaxontov 1984 and especially Starostin 1989: 18; 1999: 26) and although some scholars were questioning the above-mentioned dogma long ago, there are still many who insist on embracing the comfort of time-depth limitation.

At the same time, there seems to be growing flexibility in accepting the validity of the traditional comparative method at greater time-depth. Bernard Comrie expressed his opinion on this matter in the following way: "I do not literally mean that a certain number of years can be specified beyond which the comparative method is inapplicable" (Comrie 1999: 244). He believes that the applicability or non-applicability of the comparative method is not necessarily related to the question of time-depth: "I would certainly not exclude the possibility that it might be possible to establish regular sound correspondences within one pair of languages separated from their common ancestor by a time-depth of 10,000 years and yet impossible to do the same for another pair of languages separated from their common ancestor by 7500 years" (Comrie *ibid.*). But in the formulation of what he calls the crucial point, he falls back into the groove cut by the dogmatic approach: "[G]iven enough changes ... it is in principle the case that it will be impossible to establish regular correspondences between languages that do in fact descend from a single common ancestor" (Comrie 1999: 244). With a small reinterpretation this statement practically repeats the old argument against long-range linguistic comparison that given enough time the amount of comparable data for reconstruction at greater than usual time-depth simply diminishes to the vanishing point.

Thus the fallacy of diminishing returns (Bengtson 1992: 30f) seems to persist in current linguistic thinking. This is well illustrated by the following train of thought: "The greater relative time-depth means that a greater number of perturbing factors have had the opportunity to be at play" from which he concludes that "there may come a point at which we will be unable to distinguish between common ancestry on the one hand and borrowing or chance similarity on the other" (Comrie 1999: 247). It should not be forgotten, though, that the more languages supply evidence for the reconstruction of a proto-unit, the less we are endangered by chance phenomena and the higher the plausibility and reliability of a reconstructed proto-language becomes. This is an argument that Comrie also supports: "I agree with a point made repeatedly by wide-ranging comparativists, namely that the reconstruction of a proto-system becomes more likely to succeed the more branches of the proto-language we have available for comparison" (Comrie 1999: 244).

The crucial point, however, is that despite the enormous possibilities for linguistic change over several millennia, it is still possible to establish regular sound correspondences between languages that are known to have derived from a common ancestor, which gives a theoretical foundation for the supposition that once we can establish regular sound correspondences between languages whose genetic relationship in the distant past is (only) suspected, we should consider these regular sound correspondences as the first, initial proof

of – or at least strong evidence for – the plausibility of distant genetic relationship between the compared languages.

Another recurring issue in studying distant linguistic relationship is how to deal with the residue of linguistic comparison, i.e., how to explain attested forms that do not conform to the pattern of regular sound change. Such irregular cases usually can be accounted for in the framework of analogical change, but sometimes they may defy a plausible explanation. Some experts believe that the number of etyma stubbornly resisting a pure pedigree that would stand the test of the neogrammarian principle of regularity increases with time. But the number of examples such as that of the notorious English pronoun *she* (Comrie 1999: 244) is not significant and it is easy to see that this specific difficulty of the phonological derivation of *she* from Old English */se:o/* does not render invalid the reconstruction of the 3rd person pronoun for Proto-Germanic or for Proto-Indo-European. In a thought experiment Christopher Ehret showed clearly that "we should expect that at each stage of historical reconstruction farther back in time, slowly more and more of the same ancient root words should emerge out of our work" (Ehret 1999: 110). It follows from this that a reconstructed protolanguage necessarily displays more roots than the number of roots preserved in any of the descendant languages. This circumstance indicates not the inefficiency or inapplicability of the comparative method at greater time-depth but, on the contrary, it ensures that the genetic inheritance is recoverable despite the tendency of obscuring ancient features and elements in individual languages.

References

Békési, Gábor. 2000. Őstörténet-kutatás és paleolingvisztika. A nyelv és a beszéd eredetének aktuális őstörténeti problémái. [Prehistory studies and paleolinguistics. Current problems of prehistory in the origins of language and speech.] Miskolci Egyetem Bölcsészettudományi Kar. (Diss. MS).

Bengtson, John. 1992. "On the fallacy of 'diminishing returns' in long range lexical comparison." In *Reconstructing Languages and Cultures,* ed. by Vitaly Shevoroshkin, pp. 30-33.. Bochum: Brockmeyer.

Comrie, Bernard. 1999. "Nostratic language and culture: some methodological reflections". In Renfrew & Nettle (eds.), pp. 243-255.

Ehret, Christopher. 1999. "Nostratic - or proto-human?" In Renfrew & Nettle (eds.), pp. 93-112.

Hegedűs, Irén. 1997. "Principles for palaeolinguistic reconstruction." In *Archaeology and Language I: Theoretical and Methodological Orientations,* ed. by Roger Blench & Matthew Spriggs, pp. 65-73. London: Routledge.

Jaxontov, Sergei. 1984. "Glottoxronologia: trudnosti i perspektivy." In *Drevnejshaja jazykovaja situacija v Vostochnoj Azii. Tezisy i doklady konferencii,* Part 4, pp. 39-47. Moscow: Nauka. [English translation in Shevoroshkin & Sidwell (eds.), pp. 51-59.]

Renfrew, Colin, and Daniel Nettle (eds.). 1999. *Nostratic: Examining a Linguistic Macrofamily.* Cambridge: McDonald Institute of Archaeological Research.

Shevoroshkin, Vitaly, and Paul J. Sidwell (eds.) 1999. *Historical Linguistics and Lexicostatistics.* Melbourne: Association for the History of Language.

Starostin, Sergei. 1989. "Sravnitel'no-istoricheskoe jazykoznanie i leksikostatistika." In *Lingvisticheskaja rekonstrukcija i drevnejshaja istorija Vostoka I:* 3-38. Moscow: Nauka.

Starostin, Sergei. 1999. "Historical Linguistics and Lexicostatistics." In Shevoroshkin & Sidwell (eds.), pp. 3-50.
Vovin, Alexander. 1999. "Some notes on linguistic comparison." In Shevoroshkin & Sidwell (eds.), pp. 67-92.

Problems of Methodology in Long-Range Comparison

by Peter A. Michalove
University of Illinois at Urbana-Champaign

In *Mother Tongue* V (1999:1) the late Roger Wescott raises a number of stimulating questions about the methodology used in historical linguistics, and particularly in work on the possibility of wider genetic connections. In memory of Wescott, John Bengtson has asked for some thoughts on these questions, including the differences between long-range comparison and "traditional" historical linguistics. The answer, in a nutshell, is there is no difference; the goals and methods of work at greater time-depths should be no different from those of comparative work on more recent relationships.

If we define long-range comparison as the attempt to establish and clarify genetic linguistic relations beyond those that are already known and widely accepted, then long-range comparison is no different from the classificatory work of earlier generations. The goal of modern long-range comparison, like that of its predecessors, is to learn more about the historical development of the attested languages of the world and the demonstrable genetic relationships that may exist among them. The pioneers of Indo-European studies, such as Rask (1818) and Bopp (1816), the early Uralicists like Donner (1879) and Paasonen (1917), and the originators of comparative work in other widely accepted language families were engaged in the long-range comparison of their day.

As with previous comparative work, the methodology of present-day long-range comparison inevitably involves learning from the accomplishments and mistakes of our predecessors. Thus, any Indo-Europeanist or Uralicist today would undoubtedly express admiration for the great names just mentioned, but I doubt that any modern scholars would accept all of their ideas uncritically today. Similarly, while Nostraticists today can learn much from the work of Illich-Svitych (1971-84) and other previous scholars, it is clear that this work contains much flawed material that we cannot accept today. Therefore, I can only reiterate previous appeals (e.g. Manaster Ramer, Michalove, Baertsch, & Adams 1998) that we work to identify and strengthen the more promising elements in previous work, to discard comparisons that do not withstand constructive criticism, and to try to build something meaningful and coherent on the stronger elements; rather than accepting previous work uncritically, or rejecting the entire concept because of individual, even widespread faults.

Thus, the methodology of long-range comparison today is (or should be) identical to that of the best comparative work of previous generations. And precisely because of the accomplishments of previous scholars, comparative work today (both long-range and short-range) benefits from a much deeper understanding of the nature of language development, the processes of language convergence, and other recent work on language change. It also benefits from newer fieldwork on many of the lesser-known languages, although all comparative work suffers from the recent extinction of so many languages that undoubtedly would have offered important information on the prehistory of the languages we have attested.

In recent work (Michalove, Georg and Manaster Ramer 1998: 466-470), we proposed a gradation of comparative proposals to distinguish obvious and transparent

comparative work, which we called "rhadio-comparison", from the task of unraveling more oblique and less evident relationships, or "khalepo-comparison". This gradation may be exemplified from one extreme by the slight distinction between individual idiolects within the same speech-community; and continues through varying dialects of what speakers consider to be the same language. A further level of language diversity is represented by dialect continua, where there is mutual comprehensibility among neighboring speech forms, but not between the ends of the continuum. This relationship often appears when political or other borders intervene, attracting a variety of non-standard speech forms to one or another literary standard (as we find among many of the Slavic or Romance languages). This situation may result in an array of languages, some of which are not mutually comprehensible among native speakers, but whose relationship is obvious even to those with no background in linguistics. A further stage in this continuum occurs among families like Indo-European or Afroasiatic, that are not transparently related, but whose common origin can be seen by identifying common morphological structures, regular (and sometimes phonetically dissimilar) phonological correspondences, and (when we are lucky) the examination of early written records.

Beyond that, there are proposals of relationships such as Nostratic, that have achieved varying but not unanimous degrees of acceptance, often because they rely exclusively or primarily on lexical agreements, and often show little or no mutual explanatory power. We may simply label proposals at this end of the spectrum as, "currently debated". Such proposals, if they turn out to be justified, will simply have been more difficult to detect than more obvious ones.

It is important to note that the relative ease or difficulty we find among proposals along the continuum of rhadio- to khalepo-comparison is not synonymous with time-depth, or geographical distance. Language families that share diagnostic morphophonemic parallels, such as Indo-European, or that have a larger number of distinct attested daughter languages or branches, such as Sino-Tibetan, will be recoverable for a longer period of time. Language families without such advantages may ultimately still be recoverable, but will fall further toward the khalepo-comparison end of the continuum.

Similarly, geographical proximity can often help to maintain common, reconstructible features for longer periods of time. Yet geographical proximity can often be the source of confusion in this regard: much of the debate over Altaic deals not with the undisputed presence of large numbers of common elements among these languages, nor with the uncontroversial fact that they have shared massive borrowings over the centuries. Rather the debate is over which elements are borrowed, and what remains when we filter out material known to represent loans.

This model of levels of difficulty in identifying and reconstructing language relationships seems to be a more useful and realistic approach than a simple dichotomy between relationships that are recoverable and those that are not. The level of difficulty is not the same as a statement of degrees of actual relationship among languages, although it may often coincide with such levels of relatedness. Rather the continuum between rhadio- and khalepo-comparison provides a scale for evaluating proposals of relationship. Considering degrees of difficulty seems to be a more fruitful means of considering various proposals that have been put forth than supposing different methodologies for "traditional" versus "long-range" comparison. And surely it is more

meaningful than the simple division of scholars into neat camps of "lumpers" and "splitters".

A practical reflection of this gradation of difficulty is that broader proposals require a background in more languages. The more languages a historical linguist deals with, the more material there is to master. It is not enough to extract forms from a dictionary of a language one does not know, or to take others' reconstructed work uncritically and use it as input for deeper reconstructions. One must at the least have a working knowledge of the phonotactics, dialect geography, and known history of the languages involved. Thus, attempts to establish wider genetic connections are inherently more difficult because they require a greater-than-superficial knowledge of more languages. Much as I admire the wide-ranging background of many scholars who have produced fine work in long-range comparison, it is hard to imagine how any individual can handle it all. At some level, long-range comparison can only succeed as a collaborative effort, in which specialists regularly interact with each other and evaluate each others' findings.

Further, viewing such proposals on a continuum of difficulty shows that we cannot establish any cross-linguistic ceiling on the time-depth, even an approximate one, beyond which language families cannot be recovered, as writers like Nichols (1992: 184) or Hock (1985: 566) have suggested. Rather, different languages change different features at different rates. And within the history of a language, there are periods characterized by greater and lesser levels of change. Thus, several millennia of divergence may produce a set of languages that are unequivocally relatable (such as Indo-European or Uto-Aztecan); or it may yield a series of relationships that are much less obvious, and therefore difficult to exemplify with a familiar name.

Like the supposed ceiling on time-depths that can be recovered, another fallacy that applies to linguistic comparison at any level is the claim that relationships can be identified or refuted by purely quantitative means, or "shortcuts", as Georg (2000: 433-434) calls them. Under such methods, the cognates or other isoglosses among some set of languages are counted and found to fall over or under some threshold that the investigator expects to find among related languages. Again, this view suggests that all original relationships fall into two neat groupings, those that can be recovered, and those that cannot. The failure of such approaches as they have been used so far is twofold: first, they have not specified what constitutes a similarity sufficient to be called a cognate feature. One researcher's similarity may be another's trash.

But more importantly, they fail to capture more critical factors, such as shared aberrancies and mutual explanatory power. In this respect, one good morphophonemic alternation is worth any number of lexical comparisons.

Thus, a promising approach seems to be the recent effort to identify common sets of functions among morphological structures as in Michalove (to appear), which notes the use of the affix **-mA* as marker of the definite direct object and as a means of specification of time and space across a number of the Nostratic languages. The significant point here is not that we have a shared accusative marker, or a common specifier of time and space in these languages, but rather the agreement of these two functions across these language families, suggesting an original marker of specification.

This approach can be especially powerful when paired with common suppletive patterns as in Cavoto (2001), which, for example, sees this affix as part of the *m* ~ *k*

alternation characteristic of near demonstratives (including first-person pronouns) as well as interrogatives across the same set of languages. Whatever the original semantic significance of this alternation, the important point is that it is much more likely to reflect an original suppletive relationship than a series of independent, identical innovations in these languages.

Thus, the endeavor sometimes referred to as "long-range comparison" is simply the continuation of long-standing efforts in historical linguistics. Its goal is the same as it has always been: to use the information available to us to reach the best understanding we can of the relationships among the attested languages of the world.

References:

Bopp, Franz. 1816. *Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache*. Frankfurt am Main: Andreäsche Buchhandlung.

Cavoto, Fabrice. 2001. *Introduction à l'étude comparative des langues nostratiques: éléments de morphologie comparée et essais méthodologiques*. Thesis, Copenhagen University.

Donner, Otto, 1879: *Die gegenseitige Verwandtschaft der finnisch-ugrischen Sprachen*. Helsinki: Drückerei der Finnischen Litteratur-Gesellschaft.

Georg, Stefan. 2000. Review of Salmons and Joseph 1998. *Anthropological Linguistics* 42.3: 431-435.

Hock, Hans Henrich. 1986. *Principles of Historical Lingusitics*, first edition. Berlin: Mouton de Gruyter

Illich-Svitych, Vladislav M. 1971-1984. *Opyt sravnenija nostraticheskix jazykov*. Moscow: Nauka.

Michalove, Peter A. To appear. "The Nostratic 'Accusative' in *-mA: An Altaic Perspective."

Manaster Ramer, Peter A. Michalove, Karen S. Baertsch, and Karen L. Adams. 1998. "Exploring the Nostratic Hypothesis." In Salmons and Joseph 1988: 61-84.

Michalove, Peter A., Stefan Georg, and Alexis Manaster Ramer. 1998. "Current Issues in Linguistic Taxonomy." *Annual Review of Anthropology* 27: 451-472.

Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.

Paasonen, Heikki. 1917. "Beiträge zur finnisch-ugrisch-samojedischen Lautgeschichte." *Keleti Szemle. Revue orientale pour les études ouralo-altaïques*, 13-17.

Rask, Rasmus. 1818. *Undersögelse om det gamle Nordiske eller Islandske sprogs Oprindelse*. Copenhagen: Gyldendalske

Salmons, Joseph C. and Brian D. Joseph, eds. 1998. *Nostratic: Sifting the Evidence*. (= *Current Issues in Linguistic Theory*, vol. 142.) Amsterdam, Philadelphia: Benjamins.

Wescott, Roger. 1999. "Introduction." *Mother Tongue* V.

Comparison and Reconstruction: Language and Mythology

by Michael Witzel
Harvard University

I. LANGUAGE

§ 0. As some detractors of long-range comparison have long maintained, one can always find some 50 look-alikes in any two languages. There are indeed many inherent difficulties that one runs into when one simply juxtaposes any two vocabulary lists. However, a little history of the beginnings of comparative linguistics, especially Indo-European linguistics, helps in evaluating such general dismissive statements and puts them into perspective.

§ 1. A little history

If we take a look back at how Indo-European linguistics began, late in the eighteenth and early in the nineteenth centuries, it soon becomes obvious that the early linguists did not proceed very differently from the way many long-range comparativists are proceeding now. Similarities of certain words had been noticed ever since the Romans started to learn Greek, since Christian missionaries noted similarities of Indian words with Latin or their own languages, but even William Jones (1786) did not actually test the then current idea[1] of some sort of genetic relationship between certain Asian and European languages. When this finally was done by Franz Bopp (1816) and simultaneously Rasmus Rask (1818), almost 200 years ago, they proceeded not very differently from what we are doing today. At the risk of repeating a few well-known facts, let me recapture this, albeit in brief form.

§ 1.1 Look-alike words

When F. von Schlegel (1808), R. Rask (1814) and F. Bopp (1816) started out at the beginning of the nineteenth century, at first a list of surprising look-alikes was established, along with some grammatical features that indicated the underlying (internal) structure of the supposed parent language.[2]

Luckily, they came across words with such informative force as "Father (Heaven)," etc.: Sanskrit *dyáus pitā́(r)*, acc. *pitáram*, Greek *zeús patḗr, patéra;* Latin *iuppiter, patrem;* Germanic **tiu* / Engl. *tue-s-[day]* (+ *father*), -- thus now reconstructed as IE **di̯éu̯s ph₂tḗ(r),*ph₂térm̥.*

Early Indo-European linguists still saw only a *tendency* towards sound correspondences and did not and could not account for cases where these did not fit perfectly. This allowed a

---

[1] Cf. Mallory 1989, introduction; e.g., Parsons (1767), see Mallory 1989: 10 sq; or Lord Monboddo (1773/1774, vol. I; translated into German with an introduction by J.G. von Herder in 1785). Monboddo already then claimed that "Greek, Teutonic and Persian" must be "dialects of the same parent-language", and stressed similarity in sounds and in inflection which he regarded as most important for proof of relationship; his vol. IV of 1787 has a comparison of the Greek and Sanskrit *-mi* conjugation.

[2] See Mallory 1989: 12 sqq, Szemerényi 1970: 3 sqq. -- At that time, most people still thought that Sanskrit was the mother of all Indo-European languages, note even Schleicher's heavily Sanskritized PIE tale (1868, reprinted in Mallory, 1989: 17): *Avis, jasmin varna na a ast...* "A sheep which had no wool..."

great amount of leeway, e.g. early etymologies such as *kéntauros : gandharvá-*, where neither Greek *k-* nor *-t-* fit the Sanskrit *g-* and *-dh-*, apart from the divergent meanings "centaur" : "demigod (of procreation, music)," etc.[3]

In any such comparison, the meaning of the words compared then was and should be the same or must be semantically related, (such as English *dog* : German *Dogge* 'bloodhound', and, in this peculiar case, even the exact reverse: Engl. *hound* : German *Hund* 'dog').

§ 1.2 Similarities in grammar

Based on, and often coeval with the discovery of sound similarities as well as one-to-one sound correspondences between the languages involved, the more or less regularly corresponding structure of Indo-European grammar was noted (Bopp 1816/1833-37). In the case of "father," quoted above, the nominative case has no ending in any of languages, but the accusative has *-em/am/a*. This kind of correspondence was especially clear in the extensive verb system (Bopp). Building on such principles, by the late nineteenth century the structure and the vocabulary of Indo-European had been reconstructed (Brugmann 1886).

The general correctness, even a large degree of *reality*, of such reconstructions is obvious from such items as the grammatical pattern of the verb "to be" and a whole group of verbs that follow the same pattern; they stand out by their marked difference in the singular (*h1s-*) and the plural forms (*h1es-*): Sanskrit *ás-ti / s-ánti*, Latin *es-t / s-unt*, German *is-t / s-ind*, etc., with very isolated relics in modern languages.

This kind of regularity is, in fact, the proof of the pudding: (1) regular sound correspondences, and (2) the formal identity and structure of the grammar underlying the languages in question. Obviously, item (2) is more easily seen when item (1) has already been established as in many, though not all languages, sounds tend to diverge so much over time as to defy immediate recognition as look-alikes. Note the hackneyed correspondence of Latin *duo* / Skt. *dvā-u* and Armenian *erku*. A straightforward comparison of, say, Old Irish or Gothic grammar with Sanskrit is easier than one of modern Irish with Danish and Bengali (or Armenian!).

§ 1.3 *Ausnahmslosigkeit der Lautgesetze*: regularity of sound correspondences

It was only some 70 years later that the *Junggrammatiker* (Neo-Grammarians) postulated and proved that the same sounds or their close derivatives appear across the board in all Indo-European languages, in other words, that regular correspondences (*Lautgesetze*) existed. Any exceptions from such regular correspondences in sound and meaning are explained by dissimilation, analogies, or some other conditions that developed in one of the languages involved.[4]

As is well known, the Neo-Grammarians' strictly established reconstructions have been reconfirmed later on, for example that of the laryngeal (h_2) in **ph₂ter-* 'father', or **peh₂ur-* 'fire'. The Indo-European laryngeals (*h* with the varieties h_1, h_2, h_3) were pure reconstruction until, early in the 20th century, they were actually discovered written in the newly deciphered Hittite

---

[3] See Szemerényi (1996, first ed. 1970, in German), introduction.

[4] Such as the *-pp-* in Latin *Iuppiter*, which is taken, via the *littera* rule (long vowel + C > short vowel + CC), from the vocative form **dⁱeᵘ-pater* "oh Father Heaven".

records of c. 1600/1200 BCE. Some of the laryngeals were still found actually written (*paḫḫur* 'fire'). They would have remained hypothetical if we did not have that early attestation of a now lost language.

§ 1.4 Fine-tuning

Since the Neogrammarians, we have seen a lot of fine tuning that did not radically change the nature and main categories of Indo-European even when a new set of sounds (laryngeals), was introduced, or when, more recently, a glottalized version of the consonant system was proposed. The exact nature of the Proto-Indo-European verb system remains under discussion, though the main categories and forms are beyond doubt. Advances have been made in suprasegmental features such as the formulas of the PIE poetic speech, in syntax, etc. But, the main work was done already in the seminal period between 1816 (Bopp) and 1886 (Brugmann's *Grundriss*) .

I have mentioned all these well-known points simply because they are *instructive.* We move from similarities - anyhow that is how the human mind works most of the time - to figuring out the rules behind these similarities. Why not proceed with long-range comparisons in a similar fashion?

§ 2 Long-range comparison

The Indo-European scheme of things was well-known, by and large, by the end of the nineteenth century, as were those of several other families. Occasionally, scholars have tried to further compare individual language families with each other, such as Semitic and Indo-European (Møller 1909) or Uralic and Indo-European (Collinder 1934). Such efforts were habitually dismissed by Indo-European specialists or classified as "too early to try" or "trying the impossible: the time depth involved is too big,"[5] or they were classified, along with many amateurish efforts, as "omni-comparativist."

Inter-family comparison, thus, had been at an impasse for about 100 years. The last decades, however, have seen important advances in this sub-discipline of comparative linguistics.

To be evaluated and accepted by 'traditional' comparative linguistics, long-range comparison should proceed in a similar fashion as Indo-European linguists have done in the nineteenth century. However, the scholars in the long-range field, while busily engaged at various stages of this (same) type of game, take various, often strong positions with regard to their respective approaches (as will be visible in this volume) .

§ 2.1 Problems in long-range comparison

Employing the same methods means that new reconstructions could be made between other language families established during the past two centuries, i. e. by comparing the old, known families such as Semitic and its African relatives (Afro-Asiatic), Uralic/Finno-Ugrian, Altaic, Austronesian/Malayo-Polynesian, Sino-Tibetan, Bantu, Khoisan, etc. and in the Americas the whole slate of families ranging from Eskimo-Aleut, Uto-Aztecan, Maya, Quechua to Guarani, etc. Initially, such comparisons could proceed just as in the case of Indo-European in the nineteenth century, by simple look-alikes.

---

[5] Cf. Cowgill 1986: 13 sqq.

However, establishing family relationship in language groups without written records has been more difficult than in those that had the benefit of old, sometimes even 5000 year old, records, such as Afro-Asiatic and Indo-European. Both were lucky, as they have quite old inscriptions and old (sometimes oral) texts. Not so in the majority of the other language families. The situation is much more difficult in, say, sub-Saharan African and Amerindian languages. This is, of course, why it took a Joseph Greenberg to establish the African and American language families.

A similar difficulty is felt with languages that have or had few or no affixes, such as Sino-Tibetan. Therefore, individual methods thus must by necessity differ slightly in the individual subdisciplines of long-range comparison. But the underlying principle of regularity of sound changes over time and a common core of grammatical elements should be maintained and should remain undisputed.

§ 2.2 Look-alikes

With the establishment of the principles of comparison in the nineteenth century all earlier attempts at comparison have been rendered fruitless: they clearly appeared to be random listings of words with some similarity but little or no similarity in grammatical structure. As mentioned, it has been said that one can find 50 words in any two languages of the world that look somewhat similar and have a similar meaning. Certainly, such comparisons still are more than occasionally being made by amateurs (the "omni-comparativists") and even some linguists who randomly compare anything with anything, for example Hungarian with Maori, Aztec with Greek, or various West African languages with the Dravidian of South India (Homburger 1949, Sergent 1997).

The establishment of regular correspondences immediately opens the door to a whole range of comparisons that are not immediately obvious. It is well-known that superficial similarity is no proof at all: O.Irish *athir*, English *father*, French *père*, Armenian *hair*, Tocharian *pācar* have fairly little in common on the surface, yet, they are closely related; on the other hand, two similar-looking words such as English *heart* (< **k'erd*-) and Sanskrit *hṛd* (< **g'hṛd*-) 'heart' are not *directly* historically related.

In a first trial, however, mistaken look-alikes (Greek *theo*-, Aztec *teo*) can indeed not easily be excluded, especially so in languages without a long history of written texts. However, many haphazard comparisons easily can be avoided if the languages involved are attested early: the number of accidental equivalents is simply reduced (if we assume that language has not been around for millions of years but only for *Homo Sapiens sapiens*). Yet, even if such accidental look-alikes are at first admitted, mass comparison will sooner or later show proper correspondences and discard them.

If we test this procedure between any two well-known related languages, many of the "straightforward" look-alike comparisons are easily falsifiable, and such procedure thus cannot lend support to those who deny any value to long-range comparison. For example, taking an example from well-known families, the distantly related Nepali *dui* '2' ~ Italian *due* represent mere look-alikes, while the rest of the numbers do not agree: *ek / uno, tīn / tre, cār / quattro, pānc / cinque*, etc. On the other hand, closely related language have a preponderance of closely agreeing words and forms, which disposes of accidental external look-alikes immediately: i.e. Nepali *dui* '2' / Hindi *do*, while the rest of the cardinal numbers are the same in both languages : *ek, (do/dui), tīn, cār, pānc*, etc.

Proceeding in the same fashion with languages of unknown relationship will eventually tease out the more regular correspondences, which should be stated as such, and irregularities researched further. All of this certainly is old news; however, I mention it as we should not just stop at this stage of (mass) comparison and say, e.g., Nepali and Italian as well as Hindi and Latin, Sanskrit, etc. are *somehow* related.

§ 2.3 Lack of early attested words/forms

In the ensuing quest for earlier reconstructed forms, the older the forms available, the closer they will be to the original or ancestral form of the language family in question.[6] The value of historically attested forms thus is that they cut short much otherwise necessary family-internal reconstruction and help to decide comparisons of words much more easily.

Another problem of working with 'late' languages is that developments may have spun off in 'unexpected' directions such as IE *p-* > *zero* in Celtic, or that they may have become full circle, as visible in IE **treies* '3' : Skt. *trayas*, etc., which is represented in Proto-Germanic as *θ (ϑ)*, thus Engl. *three*, Isl. *ϑrír*, corresponding to Dutch *drie*, German *drei*, but has turned in Scandinavian > *tr-*. We are thus back at the Indo-European stage as far as **tr-* is concerned. If we did not have related testimony--as is indeed often the case in other families--of German, Dutch, English, Icelandic, we might think that IE **tr* > Germanic **tr*. Even then, the important feature in this case is the maintenance of a *regular relationship* between, e.g. IE **tr-* and Scand. *tr-*, even though several (otherwise invisible) stages have intervened.

The same or at least similar results as can be achieved, of course, by laborious, family-internal reconstructions, starting from modern dialects and languages.

§ 2.4 Lack of early grammatical forms

If we do not have well-constructed protoforms of a language family, it is more difficult to discern grammatical developments. Such comparison may yield some, though not all, grammatical forms of the proto-language.

For example, it has often been said that it would be impossible to reconstruct the Latin case endings from Romance, as these are derived from Vulgar Latin, not Classical (Caesar's and Cicero's) Latin. But, consider the Romanian postposited article, nom. sg. *-u-l*, gen./dat. pl. *-lor* < Lat. *ille* + gen. pl. *-orum*. Naive comparativists may take the suffix as having its origin in a substrate feature of the Balkan Sprachbund (cf. the same function of the Bulgarian postposited article in *-yat*), and they would not suspect the ultimately Latin origin of *-lor* from a (genitive) plural suffix.

Or note the incidental occurrence of English, Dutch and German "plural" forms: Engl. *ox: oxen; child: children*, with *-(r)en* as pl. suffix? Cf. further Dutch *os : ossen*, German *Ochs(e): Ochsen*, with *-n* as pl., or German *Kind : Kinder*, Dutch *kind : kinderen* with *-er(en)* as plural suffixes. Based on such evidence, one could suppose that we have older "IE" plurals in *-er* (as in Etruscan!) and *-en*. This is, of course, nowhere near the truth, as Skt. *ukṣan-*, pl. *ukṣān-as*, etc. show the origin of *oxen* in the common IE *n*-stem, cf. Engl. *name*, Skt. *nāman-*, or Latin *rex/regin-a* : Skt. *rājan-* 'king,' etc.

Furthermore, there is the still older variation seen in *-r / -n* stems in certain old words, such both *-n* and *-r* as in conflated form in Latin *iter, itin-er-is* 'path,' or better Skt. *ah-ar, ahn-as*

---

[6] Notwithstanding some archaic modern dialects.

'day'. This is best seen in the various forms of the Indo-European words for 'water' and 'fire': English *fire*, Greek *pūr*, Hittite *paḫḫur*, *and* Engl. *water*, *Hitt. watar/wetenaš*. But even the extremely "archaic" feature of *r/n* (as well as Caland's *-r/-i*) stems in Indo-European is a fairly 'recent' innovation from the point of view of Nostratic.[7]

In sum, obvious examples such as these show that incidental observation of modern grammatical irregularities can turn up archaisms that go back several thousand years. Only, we do not always see which exact function they might have had in the proto-language. Rumanian *-l-or* and Engl. *-en*, *-ren* were not original plural suffixes but had quite diverse origins. Such examples teach a valuable lesson: unless our reconstructions are supported by *extensive* materials outside the realm of such incidental archaic comparisons, we will reconstruct, by absolutizing and generalizing, a faulty proto-language.[8] Therefore, we have to deliberate seriously the various stages needed to reach a proper reconstruction, with the more limited materials at hand in less early attested languages, in comparison of various language families, and certainly in "Pan-Gaean".

§ 3. Procedural stages

§ 3.1 The safest way : reconstruction

As will have appeared from the above examples, chosen at random from well-known languages and families, the safest bet is in (mass-)comparing various languages to try to reconstruct their proto-forms and that of the family they belong to. This method has been used, in fact, in two or three of the recent approaches.

The recent effort of establishing a Nostratic super-family, by some Russian linguists (Illich-Svitych, Dolgopolsky, Dybo, Shevoroshkin, etc.) who followed up on an earlier proposal by Jespersen, is the prime example. The method used by these scholars is the classical 'Indo-European' one: comparing words and establishing the rules of sound changes between various languages (*Lautgesetze*). There still was little comparative grammar in the original work by Illich-Svitych. However, even a casual observer[9] can easily establish, between IE, Uralic, Altaic/Japanese, some relationships in sounds and grammar, such as the various representations of **m* and **n*, and then the corresponding representation of the accusative suffix in **-m*, or the possessive/genitive in **-n*, or the personal pronoun 'I' **m-*, with the oblique/possessive **mVn-*, or **nVm-* 'name'.

It is more than curious, then, that the Nostratic theory has not been accepted by a majority of the "traditional" comparative linguists. One "reason" often heard is that one cannot reconstruct beyond c. 6000 BCE. But who decreed that? And Why? If sound correspondences and grammatical features based on the usual *Lautgesetze* agree across the board in a perceived superfamily, then this *is* the proof of the pudding, just as in IE. (*patr-em : pitar-am; *me-no-, etc.*)

Another recent example is the reconstruction, by John Bengtson, of Macro-Caucasian (Bengtson 1991a,b) . He has established exactly the same feature for MC as highlighted above for

---

[7] Witzel 1992, appendix on the words for 'fire'.

[8] Cf. Szemerényi 1970: 310.

[9] As I did for my own benefit, when first taking serious note of Nostratic some 12 or 15 years ago, and with only a very spotty knowledge of Asiatic languages outside Indo-European and Japanese.

Indo-European and Nostratic: sound correspondences which also are reflected in the difference between nominative and oblique cases for the personal pronoun (Bengtson 2001). This again, is the proof of the pudding: *Lautgesetze* and Grammar.

As for the way to proceed, one could argue that we first have to reconstruct super-families such as Nostratic or Macro-Caucasian, and then proceed further to compare both with each other and with other families. One may argue that if we *lack* all early attested forms, such as for "Papuan" or Australian, we should either proceed with an internal reconstruction of one language or initially perceived language family or, better, of all Australian or Papuan languages before long-range comparison. As pointed out above, reconstruction of individual families and only then, establishment of super- or macro-families is the safest bet.[10]

The projected results in the cases just discussed have as their *pro* that they will be the safest attainable; as their *contra,* that the method used will be the most labor-intensive. The level of attainable "sophistication," compared to that of Indo-European and other well-established families is that of good post-1870 Indo-European reconstructions.

§ 3.2 Simultaneous internal and external comparison

Otherwise, we could proceed, on the one hand, with mass comparison of *all* languages[11] in the suspected range of a language family, or on the other hand, of comparison of one or more languages with a similarly defined target of comparison.

The late Joseph Greenberg, most prominently, has looked at the multitude of language families in Africa, and comparing the vocabulary of these languages across the board,[12] he established a two super-families: Nilo-Saharan and Niger-Congo. This classification has been by and large accepted by African specialists. His work on the Amerindian super-families for the Americas, instead of the accepted c. 150 language families, has not yet been welcomed or accepted to that extent, and that of an Indo-Pacific family simply has too few scholars interested and knowledgeable about the languages involved.

The problem with African and American languages, as has been indicated, is that we hardly have records older than a few hundred years. Therefore time depth, so important in the early reconstruction of Indo-European and (Hamito-)Semitic (Afro-Asiatic), is altogether lacking. The earlier a word (or grammatical form) is actually attested, the closer it is to the reconstructed Proto-family. While language change is not constant, early forms *do* make a difference.

In both cases, that of sub-Saharan Africa and that of the Americas, Greenberg thus had to rely, almost exclusively, on the present day, "surface" materials of these languages. Furthermore, he committed two "cardinal sins" in the eyes of 'traditional' linguists: he did not establish regular relationships between the sounds of *all* the various languages involved and he did not reconstruct much of a common grammar for them.

However, what he in fact did was not very different from what Bopp and others had set out to do for IE: to compare some words which looked *similar* in shape (sounds) and had the same or closely related meanings. Yet, even the first grammar by Bopp (1816) included sound

---

[10] As similar stance is taken by I.Peiros 1977.

[11] See the summaries by M. Ruhlen 1994: 13sqq.

[12] See his summary in MT 2, 1996.

relationships and a comparison of the underlying grammar and the *structure* of IE, while this is still rare in Greenberg's Amerindian.

Greenberg's method thus seems to agree with the stance taken at the beginnings of Indo-European studies, from 1816 to 1886: studying similarities, including some sound-to-sound correspondences, and some incidental grammatical features. His method thus cannot be put on the same level of sophistication as that of the Nostraticists or of Bengtson's Macro-Caucasian.

Other scholars, such as Merritt Ruhlen, have tried to establish more wide-ranging comparisons, beginning with a Basque-Chinese link, and Greenberg now establishes Eurasiatic links which include Indo-European, the Palaeo-Siberian languages, Japanese, and Inuit.

When one is faced with a large number of seemingly unrelated languages and language families, the choice of the method of mass (or 'multilateral') comparison can be readily understood, and it has been successful where the entrenched Sub-Saharan and Amerindian comparativists of the past century could not make any headway. In consequence, this is not to be understood as criticism. Greenberg et al. are not *wrong* per se, it is just that the actual relationship between the Amerindian subfamilies (long supposed to exist in splendid isolation) still has to be worked out and proven by rules of regular sound changes.

The next logical step, thus, would be to compare all these related languages with each other in depth, following the steps outlined above, -- something that has been *anathema* for so long among Americanists. It should lead, just as in Africa, to a *more structured* system of sub-families of Amerindian.

So far, the level of confidence in this approach is only that of the early nineteenth century IE, of Bopp/Rask, not that of the Neogrammarians. The pro is, of course, that so far *unrecognized relationships* are made visible. A lot of footwork will have to follow.

§ 3.3 Shortcut

Finally, still other researchers would like take a shortcut and establish Proto-World etymologies straight-away. It is precisely here, in the ensuing tale of the long-range, 'Pan-Gaean' comparativists, that some perspective and even criticism is appropriate. They have made efforts to link all major (twelve or so) language families and to reconstruct a one-world original language, that of the "African Eve" or the Y-chromosome-based "Adam."

Some scholars such as Blažek and Ruhlen compare some materials from all resulting hyper-families and arrive at an original language that could have been spoken by the African Eve, as postulated by the geneticists. The ultimate aim is to trace back the established language families to several levels of super- and hyper-families, or skipping this, to arrive directly at the *one* original language of the mitochondrial Eve.

The chance for accidental look-alikes is greatest in this type of procedure as it cuts short some 50,000 or 100,000 years of deviation and (circular?) developments of sounds.[13]

Obviously, these exciting efforts, based as they are, on one or more handfuls of Pan-Gaean etymologies, still lack the consistent and thorough methodology outlined above that has been successfully used in the 19th and 20th century reconstruction of Indo-European and of several other large language families.

---

[13] The ultimate result, i.e. the relationship between certain sounds, may be correct in some cases but not their actual proto-form; see above the case of Scandinavian *tr-* from IE *-tr-* which 'skips' the Proto-Germanic **ϑ* stage.

The level of certainty here is that of the earliest Indo-European observations from Jones to Bopp. Many cases, such as the *gandharvá-* = *kéntauros* equation (cited above), may be included in such comparisons. They may indicate some or even many valuable connections, but they are in need of support by the methods outlined earlier.

§ 3.4 Ideal Procedure

Recently we have heard that non-Africans left the African continent as recently as 50,000 BCE -- all 10,000 of them. The chance, then, of purely accidental look-alikes is somewhat reduced. Yet, even 50,000 years is a long time, compared to the c. 7000 years of Indo-European or c. 12,000 of Nostratic.

In the end, Proto-World or Pan-Gaean etymologies should follow the same procedures as established above: we would like to see a *set of rules* that cover the Pan-Gaean, "Proto-World" etymologies (of Blažek, Ruhlen, etc.); *and* in addition one would like to see, -- faint hope perhaps -- *some grammatical features* that can be extracted from of this type of comparison, similar to what Greenberg could detect for Amerindian languages by straightforward mass comparison.

I now leave language aside for the moment, and turn to an unexpected field, myth.

II

MYTHOLOGY

§ 4.1 Comparative mythology

In addition to archaeology, anthropology and genetics, all of which have already delivered important results for our undertaking of reconstructing and understanding the earliest forms of humans, their behavior and their speech, I propose to add the study of the earliest forms of mythology[14] as well.

Although myth has been studied for a long time, and comparatively so, for more than a hundred years, such comparisons have not yet yielded a cogent system of relationships.[15] However, all such interpretation are usually restricted to *one* myth, or variants of it. If similarities between myths in various cultures are noticed they are explained in various fashions, the two most current explanations being *diffusion* and *archetypes.*

---

[14] This section is based on the results of my still unpublished book, *Origins*, much of it written -- on renewed direct experience and contact with Japanese religion, myth and folklore -- in 1990, during a sabbatical at the Institute for Research in Humanities (*Jimbun Kagaku Kenkyujo*) of Kyoto University, and slightly expanded and revised later on, but not yet finalized. These results have been used (and tested!) in classes at Harvard 1990-2000, and in some public lectures (June 30, 1990 at Kyoto University, in March 1993 in a conference on mythology organized by Phyllis Granoff at McMaster University, and in Feb. 1998 in the century-old *Shop Club* at Harvard, where we "talk shop" about our own fields of study).

[15] There is a long list of interpretations of myth, from the antique and Renaissance stance (Vico) regarding them as allegorical or euhemeristic, from Max Müller's disguised nature myths to astral mythology, from ritual to Malinowski's social charter, from Freud's theories of repression to Jung's universal psychic archetypes, from myth as disguised history to Lévi-Strauss' binary, structural analysis supposedly reflecting the structure of the human mind.

Diffusion entails that the similarities in widely distributed myths are due to a gradual dispersion of such motifs from a known or reconstructed center.[16] In most cases, however, we can no longer follow the trail of the diffusion of such myths or complexes of myths. Shamanism and its myth of the shaman's death, recomposition of the body, ascent to the heavens, etc., is spread over a wide area, from Siberia to Nepal and from Lapland to the Americas, but we do not know how it spread and when, or whether it really was the predecessor of certain mythologies and religions now existing in Eurasia. The same holds for individual myths such as the Orpheus myth which is found in Greece, Japan, North America, etc.[17]

The other currently common theory that aims at explaining such similarities is based on Jung's psychology. According to his and other's (such as Joseph Campbell's) explanation, certain motifs, or composite parts of them (archetypes) are *universally* human.

The image of the mother deity, the Goddess, is supposed to be one such archetype. Since such archetypes are generally human, they can appear in dreams, visions and myths, and are supposed to re-emerge even in areas where they have traditionally not been prominent, such as in certain European societies.[18] If this were correct, we would expect that these motifs or archetypes would indeed turn up in *all* parts of the globe. This, however, is not the case.

As explanations, diffusion and archetypes – and incidentally, all others proposed so far[19] – fail to address the central, unnoticed problem: the comparability of *whole systems of myths*, or, to use a linguistic simile, the comparison of whole grammars, not just of a particular word, form, declension, conjugation or syntactical feature. It can indeed be noticed – though this has not been done so far – that whole mythologies, such as the Vedic Indian or Japanese, not only have similar contents (individual myths with similar motifs/archetypes) but these contents also are arranged in similar fashion. In fact, a fairly large number of these mythologies exhibit a *common story line.*

Thus, the currently fashionable explanations referring to general human universals cannot explain the extraordinary amount of similarities and congruities, whether they suppose binary structures of arrangement of mythological items (Lévi-Strauss), or psychic archetypes (Jung, Campbell), or of diffusion (Baumann, Stith Thompson); for these similarities and congruities are found in large areas of the world, but they are *not found* on all continents, nor are they *evenly distributed.*

§ 4.2 Laurasian Mythology

The new comparative approach proposed here has been developed over the past decade. The steps undertaken include first to look at the common (story line) features, then to take account of the whole extent and structure of the various local mythologies, and finally to reconstruct a coherent mythology for much of Eurasia, North Africa and the Americas. Its

---

[16] Well attested cases are, of course, that of Judaeo-Christian or of Buddhist myths, which have swept large parts of the globe well before the age of European discovery, travels and colonialization.

[17] The spread of such myths has been studied especially by Stith Thompson and his school.

[18] Conveniently forgetting about the pre-Protestant image and worship of Mary, mother of Christ, which is mythologically very complex: Mary as mother, immaculate virgin, ruler of the world, and as a sort of heavenly bride, -- all under the guise of a very important Christian saint.

[19] See W. Doty, *Mythography* (1986).

designation, *Laurasian Mythology*, is derived from the geographical term, Lauretania, in Canada, and that of Greater Asia.

This new approach and the steps taken are in fact similar to the well-tested methods of historical linguistics. First, a general reconstruction is made, based on a number of obvious similarities, of the complete mythological structure. This is characterized by a narrational scheme that encompasses, in succession, the ultimate of origins of the universe and the world, the subsequent generations of the gods, an age of semi-divine heroes, the emergence of humans, the origins of "royal" lineages. It frequently includes a violent end to our present world, sometimes with the hope for a new world emerging out of the ashes. Ultimately, the universe is seen as a living body, in analogy to the human one: it is born from primordial incest, grows, develops, comes of age, and has to undergo final decay and death.

In passing it might be mentioned that some of the mythological comparisons seem to overlap even with linguistics proper. For example, in the historically unconnected mythologies of old Japan (*Kojiki*) and earliest India (*Ṛgveda*), the male deity who opens the primordial cave is described or even named in the same semantic terms (though linguistically unrelated), i.e., 'arm-strong' (O.Japanese *Ta-jikara,* Vedic Skt. *tuvi-grābha, ugra-bāhu* [Indra]); in both mythologies the deities of fire are male, and those of water are female, etc.[20]

Second, once the main outline and geographical extent of Laurasian mythology have been established, we can proceed in several further ways:

(1) noting the "regional" (sub-)varieties, for example the Indo-European one or the Near Eastern myth-families;[21]

(2) exploring their historical development by setting up a family tree of such (sub-)groupings;[22]

(3) concentrating on individual mythologies and exploring how closely they represent the reconstructed Laurasian type, what they miss, and what can be reconstructed by internal comparison;

(4) further delineating the influence on individual mythologies from the surrounding areas and, by contrast, purely local developments.

Such internal reconstruction will fill the gap between, say, the reconstructed Near Eastern branch and the individual local mythology, e.g. that of the Sumerians or Hurrites.

Third, even though this is a large-scale project, possible only with the help of many specialists in the fields of individual cultures, the project cannot stop here. Instead, initial exploration carried out over the past few years has indicated that Laurasian mythology, though

---

[20] See Witzel, Vala and Iwato (forthc.) It must be stressed that Japanese myth (in its recorded form, of 712 CE, going back at least to the first half of the first millennium CE) has no direct or indirect connections with (Vedic) India (1500-500 BCE) before the introduction of Buddhism around 500 CE. -- Many other incidental, isolated, unexpected details and (poetic) motifs could be adduced, such as that of the Vedic Indian fire god, *Mātari-śvan* 'growing inside the mother', and the archaic Japanese fire god *Ho-musubi* 'growing (as) fire (inside his mother)', who at his birth burned his mother Izanami so severely that she died. Or, there are such isolated motifs as that of the Vedic deity Uṣas exposing her breasts as a greeting to her close friends (the poets), a feature also found in the unrelated Siberian Amur region, etc.

[21] Including their mutual interrelations and secondary influences upon each other.

[22] Always keeping in mind that secondary influences may have changed the picture, as in the case of the close cultural interaction of preclassical Ancient Greece / Anatolia / Syria-Palestine.

covering very large parts of the globe, is not the only one in existence, and that it is not isolated among the several other existing types.

§ 4.3 Gondwana mythologies

The mythologies of the Australians and of the Papuans, as well as that of most of sub-Saharan Africa, represent distinct types that are very different from the Laurasian one. It is significant that certain motifs are altogether missing in this *Gondwana* belt.[23] Typical examples are the lack of creation myths that tell the origin of the world or of the lack of flood myths, or of details such as the lack of female witches.[24] One may also add details such as a literary phenomenon, the preference for improvised magical spells and the disregard of the power of "true," well-formulated, and secretly transmitted magical poetry, so typical of much of Laurasia.

While Laurasian mythology can be described as being highly interested in origins, especially the origins of the universe and the succession of the various generations of the gods and that of four subsequent ages, the mythologies of Africa and Australia/New Guinea generally do not take notice of this question and generally confine themselves to describing the emergence of humankind in an already existing world.

Obviously, the Gondwana mythologies must be investigated carefully and their types, structure and development must be established to the extent possible in cultures that do not have a long-term written record. This problem is similar to that faced in long-range comparison when one must work with languages that have only recently been recorded. The very geographical isolation of some mythologies may help, though, as is the case in those of Australia and highland New Guinea.

We must study the relationship with and mutual influence, if any, of the Gondwana mythologies and Laurasian mythology. In West and East Africa, for example, certain northern (Sahel, north African) influences have long been observed by Africanists. In the present context they can be seen to have overlaid the older Gondwana patterns. Even then, all Afro-Australian mythologies are genuinely different both from each other and from the Laurasian type.

§ 4.4. First myths

The implications of the project do not stop here. Even initial exploration has brought out the surprising fact that certain individual motifs and myths occur across all four types ... Sub-Saharan African, Laurasian, Papuan, and Australian. As briefly described above, Laurasian myth is characterized by a coherent storyline, and so are *some* of the Afro-Australian ones, despite the lack of content on creation, origin of the gods, or the four ages. More importantly, what is significant about the few newly emerging, truly universal motifs is not just their world-wide spread, it is the fact that these universals also occur, but are *isolated* in, Laurasian myth. They often go against the grain of Laurasian myth, and are 'superfluous' variants of topics treated comprehensively and systematically elsewhere in Laurasian myth.

Frequently, these variants are not part of the "official" local story line but occur as isolated myths, generally in form of folk tales or *märchen*. For example, the origin of humans from trees or from tree trunks is not at all normal or common in Laurasian myth. Yet, it occurs

---

[23] The name is again taken from geography, where it includes, however, not only Africa and Australia but also India, Madagascar and South America.

[24] Cf. for Europe and N. Asia, Carlo Ginzburg (1991).

in Icelandic and Japanese myth: in Iceland as a minor part of the main story line (*Askr/Embla 'ash/elm'*), but in Japan only as folktale (*Kaguyahime*) which is not part of the 'official' mythology in the Kojiki/Nihonshoki. The motif otherwise is found in Austronesian Taiwan, in those parts of Central Africa not influenced by Laurasian traits, and commonly in the isolated Australia, which has been cut off from Asia for at least some 6000 years, with initial Australian immigration at c. 40,000 B.C., or earlier. What we have in examples like tree origin tales in Laurasian cultures are fragmentary *remnants* of a tradition that *precedes* the individual four types of mythology briefly described above.

It thus appears that Laurasian mythology may be an offshoot of the older Gondwana type, underlying the Sub-Saharan African and Papua/Australian mythologies. Based on these three or four types, an earlier version of a *Pan-Gaean* type might be reconstructed.

In short, Laurasian mythology is the *first novel*, and the Pan-Gaean motifs are the *oldest tales* of humankind. At least, they are the oldest ones that actually can be discovered, barring any new insight about Neanderthal speech and ritual.

And this is their fascination. The Laurasian (and Gondwana) project will take us back beyond all written literature, and beyond most cultural data encapsulated in individual languages or reconstructed for the various language families. It will enable us to take a glimpse of the human condition as experienced by our most distant ancestors, both before and after they moved out of Africa.

The new theory clearly is worth being tested by a larger group of specialists of various cultures, literatures, languages and religions. Such close cooperation will, at any rate, lead to many new insights, whether the initial theory of a mono-local origin of all human myths and the three or four subdivisions proposed here can be maintained or not.

As such, Laurasian and Gondwana Comparative Mythology forms a fourth approach in our endeavor,[25] next to that of language comparison, genetic study and archaeology.

To return, finally, to language proper.

III. Technical Matters

§ 5 How to proceed?

Ideally, one would assemble leading linguists (and textual scholars dealing with myth), preferably in one location, such as an ASLIP institute, and proceed with comparisons. However, no longer by using hand-written notebooks (as the late Joseph Greenberg still did so successfully), but with the help of large electronic storage capacities as well as with (super)computers and programs that allow us to compare a large array of data and to search for *patterns*. That is, patterns both in sound relationships and as in the meanings of words and their derivatives. By this process scholars will be able to achieve, in a much shorter period and more comprehensively, results that formerly took decades. And, they will be able to carry out constant counterchecks, test possible alternatives, apply new methods, and evaluate differing proposals--all 'at the push of a button.'

---

[25] In passing, I mention that there are other fields that I have explored in this context, such as gestures, preference for certain musical scales or rhythms, choice and combination of colors, all of which represent regional features that are in need of detailed comparative study in conjunction with language and myth.

The rather laborious and tedious search for correspondences among the major language families could be greatly assisted by computers; yet such tools have hardly made their entry in the field.[26] The combination of well-tested methodologies and modern means could lead to clear and convincing results (including the rejection of previous proposals). If pursued consistently, the results of such comparisons will make it clear whether we have to reckon with the expected one-world Ur-language ("Pan-Gaean") or with several areas of origin. This approach also will be of immediate interest for the further development of the myth theory, outlined above, and for mutual comparisons of language and myth with genetic distributions and with palaeontological and archaeological records.

§ 5.1 Programs

We can, of course, work like the late Joseph Greenberg, with two dozen notebooks, intelligence, patience and a good memory. But, things are easier now with electronic aid. First, we would have to evaluate the use of recently developed, more "intelligent" text storage and retrieval systems that go beyond simple databases and search engines.[27] They can be used by any up-to-date PC or Mac. Examples include XML, the versatile new version of HTML. As an additional benefit, this encoding will remain stable as a text entry and tagging system regardless of innovations to come.[28] One of the earliest, if not the first, examples of such work in the humanities has been that of a member of ASLIP, the Indologist J.R. Gardner in 1998sqq.[29]

Other programs include the extremely versatile *Structured Information Manager* (SIM) software of RMIT/Australia, which allows multiple links between highly structured data, comparisons and searches.[30] The obvious drawback is, of course, the price of the program.[31]

Another important development is the use of "mapping" programs which allows one to link, somewhat similarly to the SIM approach, various items whether previously tagged or not, both with each other and with data outside the database. This will be especially important for the

---

[26] Excepting, of course, such well-designed and useful search engines such as the etymological lists of Sergei Starostin, Moscow, at http://starling.rinet.ru/babel.htm and Alexander Lubotsky, Leiden, at http://iiasnt.leidenuniv.nl/ied/index2.html.

[27] However, note again Starostin's etymological dictionaries of Altaic (including Korean and Japanese), North Caucasian, Semitic, Dravidian, Yeniseian, Sino-Tibetan, Chukchi-Kamchatkan; and Lubotsky's Indo-European dictionary.

[28] Cf. http://iiasnt.leidenuniv.nl/ied/sgml-xml.html.

[29] See Gardner 1998, *IJTS* 4-1, 2000,*EJVS* 7-1, 2000.

[30] SIM is an Object-oriented Database System which includes native support for XML, MARC, Z39.50, Dublin Core and related structured information storage standards. After the release of the ACE 3.0 scripting language, it can handle, with Unicode support, both the quantity and heterogeneous nature of non-Roman-script resources for full querying in native scripts as well as complex abstracted data. J.R. Gardner has set up a SIM program at ATLA/Emory University, with the XML capabilities of Oracle 8i, as part of structuring the storage of a 50 gigabyte cross-disciplinary resource. Working with XSL/T, widely differing information resources could be transformed for synthesis into the Oracle structure, allowing access to the unbounded hierarchies of highly structured information being stored and processed in the project.

[31] It costs, before academic discount, some $ 50,000. Consequently, my efforts to get it for Harvard have not worked out.

vital aspect of the various meanings of words to be compared (see below § 5.3). An initial trial of mapping procedures for linguistic procedures is in progress now.[32]

This trial is undertaken (in collaboration between Gardner and me), at a free facility provided to us by a commercial firm. We will keep members informed about the progress. Once the parameters will have been set up well, we expect to 'plug in' any data that are provided by members.

In order to be able proceed with such ambitious and costly proceedings, I suggest cooperation, first among ASLIP members, in establishing 'raw' word lists, with meanings attached, of important languages such as those being compiled by Paul Whitehouse, Sergei Starostin, L.V. Hayes and others.

§ 5.2 Correspondences

Once more materials and more electronic word lists have become available, it will become easier for researchers to use the new facilities to establish correspondences by quick searches, while simultaneously taking into account the semantic range of the words compared.

This procedure will be of great importance for non-look-alikes which more often than not nevertheless represent typical underlying correspondences, such 'unlikely' relationships as Latin *duo* = Armenian *erku*. Once we have enough electronic data, we can simply look for such 'unexpected' cases, i.e. IE **dṷō(u)* and Arm. *erku*, just by looking for patterns, without any preconceived correlations in mind, or we can do so to confirm a suspicion about a possible relationship.

Another beneficial outcome of setting up such databases and programming is, naturally, that long-range comparativists will be able to quickly test hypotheses across a broad range of individual cases and across many languages.

§ 5.3 Semantics

More important perhaps is the effect that such databases and programs will have on testing the *semantic side* of comparative questions. From the beginnings of Indo-European studies it has been realized that certain words from two or more languages make perfect pairs phonetically but that their meanings agree only to some extent, or not at all. Since the *meaning* of individual words is the least formalized aspect of language, this provides a considerable challenge to comparativists.Even then, by using theories such as the noeme theory of E. Koschmieder and K. Hoffmann[33] or the *Wortfeld* theory of J. Trier and his modern successors,[34] will help to arrive eventually at some structure that can be utilized in our effort. Ultimately, we will have to construct a picture of reality as reflected as in as many ancient and modern languages as possible. If this sounds utopian, it is worthwhile to point out that computer specialists are busy

---

[32] We start with the discussions held on the MT list (http://204.156.22.2/cgi-bin/demogate/mothertongue/lwgate/ MOTHERTONGUE/) which will soon move to a new location (see the ASLIP home page: http://www.people.fas.harvard.edu/~witzel/aslip.html).

[33] See K. Hoffmann 1967, 1975-67, Koschmieder 1929, whose approach ultimately goes back to Husserl's phenomenology.

[34] See the detailed discussion of the recently expanding use of Trier's method for European languages other than German, in A. Hintze (2000), introduction.

with precisely that (*comme d'habitude*, just for English!), that is by constructing something they call 'ontologies'.

Based on such words lists and semantic classifications, the search for cognates will become more readily possible. Searches would include levels of decreasing probability for semantic 'hits', for example, decreasing from *'dog' : 'dog'*, to *'dog' : 'hound'*, to a less likely correspondence of *'dog' : 'wolf'*, and the almost excludable *'dog' : 'cow/horse'* (which is possible, say, in some languages of the Arctic). The establishment of semantically linked lists of words will of course allow the reverse as well, that is, to look for words that are only semantically related in various languages but then are found to include some look-alikes or even deeper relationships.

Needless to say, the same process also can be used--with the same or with even more benefit--to compare the virtually endless versions of myths across the globe.[35]

§ 5.4. Comparison with other sciences

Finally, just as ASLIP members have done in the past, help can also be obtained (and "mapped" by the procedures outlined above) from archaeology, population genetics, comparative textual studies and early myth. This requires a wide range of international cooperation. We need openness, a willingness to share--not sitting on one's data--a desire to exchange materials and ideas, and a willingness for debate on all sides, including that of the 'traditional' comparativists.

References

Bengtson, J.D. 1991a. "Notes on Sino-Caucasian." In: V. Shevoroshkin (ed.). *Dene-Sino-Caucasian Languages*, 67-129. Bochum: Brockmeyer.

---------- 1991b. "Some Macro-Caucasian Etymologies." In: V. Shevoroshkin (ed.). *Dene-Sino-Caucasian Languages*, 130-141. Bochum: Brockmeyer.

---------- 2001. "Genetic and Cultural Linguistic Links Between Burushaski and the Caucasian Languages and Basque." Paper at the 3rd Harvard Round Table on Ethnogenesis of South and Central Asia. (see: http://www.fas.harvard.edu/~sanskrit/RoundTableSchedule.html)

Bopp, F. 1816. *Über das Conjugationssystem der Sanskritsprache : in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache. Herausgegeben und mit Vorerinnerungen begleitet von K. J. Windischmann.* Frankfurt am Main : Andreaischen

---------- 1833-1837. *Vergleichende Grammatik des Sanskrit, Zend, Griechischen, Lateinischen, Litthauischen, Gothischen und Deutschen.* Berlin : F. Dümmler.

Brugmann, K. 1886-93. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen.* Strassburg: Trübner.

Collinder, B. 1934. *Indo-uralisches sprachgut; die urverwandtschaft zwischen der indoeuropäischen und der uralischen (finnischugrisch-samojedischen) sprachfamilie.* Uppsala: Almqvist & Wiksells.

Cowgill, W. 1986. *Einleitung.* In: J. Kurylowicz, M. Mayrhofer (*eds.) Indogermanische Grammatik. Band I. 1. Halbband,* 11-71. Heidelberg: Winter

Doty, W. G. 1986. *Mythography : the study of myths and rituals.* University, Ala.: University of Alabama Press.

---

[35] See the index of Stith Thompson (1932).

Gardner, J.R. 1998. *The Developing Terminology for the Self in Vedic India.* Ph.D. Thesis, Iowa University.

---------- 2000a. "Magically Storming the Gates of Buddhahood: Extensible Text Technology (XML/XSLT) as a Simulacrum for Research." *IJTS* vol. 4, Special Issue, June 2000. (http://www.asiatica.org/publications/ijts/back.asp)

---------- 2000b. "Virtual Vidhaana: E-Textnology (XML and XSL/T) and Vedic Research." *EJVS* 6-1, Nov. 2000. (http://nautilus.shore.net/~india/ejvs/issues.html)

Ginzburg, C. 1991. *Ecstasies : deciphering the Witches' Sabbath.* New York, N.Y. : Pantheon.

Greenberg, J.H. 1996. "The genesis of Multilateral Comparison." *Mother Tongue* 2: 145-148.

Hintze, A. 2000. *"Lohn" im Indoiranischen : eine semantische Studie des Rigveda und Avesta.* Wiesbaden : L. Reichert.

Hoffmann, K. 1967. *Der Injunktiv im Veda*, Heidelberg: Winter.

---------- 1975, 1976, 1992. *Aufsätze zur Indoiranistik.* Ed. J. Narten. 2 vols. Wiesbaden 1975, 1976; vol. 3. Ed. S. Glauch, R. Plath, S. Ziegler. 1992. Wiesbaden: Steiner.

Homburger, L. 1950. "Exposé sur les faits linguistiques communs aux langes négro-africaines et au dravidien." *Bulletin de la société linguistique de Paris* 45: xix-xx.

---------- 1959. "De quelque noms d'animaux en sindo-africain." *Bulletin de la société linguistique de Paris* 54: xxiv-xxv.

Koschmieder, E. 1929. *Zeitbezug und Sprache, ein Beitrag zur Aspekt- und Tempusfrage*, Leipzig / Berlin: B.G. Teubner.

Mallory, J.P. 1989. *In Search of the Indo-Europeans. Language, Archaeology and Myth.* New York, N.Y. : Thames and Hudson.

Möller, H. 1909. *Indoeuropaeisk-semitisk sammenlignende glossarium.* Kjøbenhavn: Universitetsbogtrykkeriet J. H. Schultz.

---------- 1911. *Vergleichendes indogermanisch-semitisches Wörterbuch.* Göttingen : Vandenhoeck & Ruprecht.

Monboddo, J. Burnett, Lord. 1773. *Of the origin and progress of language.* Vol. I. Edinburgh, Printed for A. Kincaid & W. Creech, etc., etc.

---------- 1774 -1809. *Of the origin and progress of language.* 6 vols. Edinburgh, J. Balfour; London, T. Cadell [etc.].

Peiros, I. 1997. "Macro-Families: Can a Mistake Be Detected?" In: *Indo-European, Nostratic, and beyond: Festschrift for Vitalij V. Shevoroshkin*, ed. by I. Hegedűs, P. A. Michalove and A. M. Ramer. (Journal of Indo-Europan Studies, Monograph 22.) Washington, DC: Institute for the Study of Man

Rask, R. 1818. [1814] *Undersøgelse om det gamle Nordiske eller Islandske Sprogs Oprindelse.* Kjøbenhavn : Gyldendal [written in 1814!].

Ruhlen, M. 1994. *On the Origin of Languages. Studies in Linguistic Taxonomy.* Stanford: Stanford University Press

Schlegel, F. von. 1808. *Über die Sprache und Weisheit der Indier. Ein Beitrag zur Begründung der Alterthumskunde von Friedrich Schlegel. Nebst metrischen Uebersetzungen indischer Gedichte.* Heidelberg: Mohr und Zimmer.

Sergent, B. 1997. *La genèse de l'Inde.* Paris: Payot.

Thompson, S. 1932-1935. *Motif-Index of Folk Literature. A Classification of Narrative Elements in Folk-Tales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jest-Books, and Local Legends.* Bloomington.

Szemerényi, O. 1970. *Einführung in die vergleichende Sprachwissenschaft.* Darmstadt: Wissenschaftliche Buchgesellschaft. = *Introduction to Indo-European linguistics.* [4th ed.] Oxford : Clarendon Press / New York : Oxford University Press 1996

Trier, J. 1931. *Der deutsche Wortschatz im Sinnbezirk des Verstandes. Die Geschichte eines sprachlichen Feldes.* Heidelberg: C. Winter.

---------- 1973. *Aufsätze und Vorträge zur Wortfeldtheorie.* Hg. von A. van der Lee und O. Reichmann. The Hague: Mouton.

Witzel, M. 1992. "Meaningful ritual. Structure, development and interpretation of the Tantric Agnihotra ritual of Nepal." In: *Ritual, State and History in South Asia. Essays in honour of J.C. Heesterman*, ed. A.W. van den Hoek, D.H.A. Kolff, M.S.Oort, 774-827. Leiden.

---------- Forthcoming. "Vala and Iwato. The myth of the hidden sun in India, Japan -- and beyond."

Philosophical Differences and Cognitive Styles

by Sydney Lamb
Rice University

One of the questions raised by Roger Wescott concerns "core philosophical differences". This topic, of great interest to me, can be illuminated by considering the notion of *COGNITIVE STYLE*. Why do different people have philosophical differences? One might suppose that they are the result of different paths taken in a logical process of reasoning. But in most cases, one's basic philosophical views probably come first and the reasoning used to support them, if engaged in at all, comes later in recognition of a need to justify. That is another way of saying that there is something more basic than a reasoning process, something so basic that it leads one to favor certain views and beliefs over others, even certain reasoning processes over others. For this more basic something we can use the term *COGNITIVE STYLE*. The concept of cognitive style has been discussed in recent years in the fields of psychology and education, and a fair amount of literature has been built up (e.g. Claxton & Murrell 1987, Sternberg 1997, Sternberg & Zhang 2001). Summing up their extensive survey of literature on the implications of learning style for educational practices, Claxton and Murrell identify two "fundamental orientations" that are repeatedly identified with a variety of terminologies: "Splitters, field independents, serialists, and abstract analytical learners are more in the objectivist mode of knowing, and lumpers, field sensitives, holists, and concrete learners are more in the relational mode" (1987: 75).

As a very basic feature of cognition, one's cognitive style will manifest itself in many aspects of thinking. Among investigators of linguistic prehistory, we may conjecture that one's cognitive style plays an essential role in one's philosophical approach. To put it simply, we may say that 'lumpers' and 'splitters' operate in different cognitive styles. As cognitive style manifests itself broadly in thinking patterns we can expect to find correlations between views in one area and those in another. Consider the three contrasts mentioned by Wescott to illustrate "core philosophical differences": "absolutism vs. relativism; monogenism vs. polygenism; lumpers vs. splitters". We may conjecture that absolutism goes with 'splittism' and that relativism correlates with 'lumpism'.

As for the difference between monogenism vs. polygenism, it all depends on what meaning is given to these terms. For commonly held opposing views, relating to the origin of language, I would propose an entirely different argument, as I believe a realistic assessment of the possibilities leads us to reject both views. Both are based on a mistaken notion that language developed either as a single step or a series of steps that together comprised a short time span. But if it was otherwise – if the development took place as a series of many steps over thousands or millions of years – then how could it make sense to apply either a concept of monogenisis or one of polygenesis to that entire series of perhaps dozens of separate steps, which surely occurred in several or many different places? Rather, the issue should be considered separately for each of these steps.

On the other hand, we can adopt the more limited view of monogenesis, which holds that all of the languages *now spoken in the world* stem from some common source. That common source would have been a language or perhaps a simpler language-like

system representing the end of a long line of steps of originating, a system likely to have been one of many in use at the time it was spoken. This would be a concept that makes sense.

But let us return to the cognitive styles of splitters and lumpers. To me one of the striking features of 'splitter-think' is its absolutism with regard to probability. It is commonly held by splitters that one cannot consider two languages to be related unless that possibility has been demonstrated beyond a reasonable doubt, by means of regular phonological correspondences. This view has been discussed at length by others, including Joseph Greenberg, who has pointed out that extreme forms of it that have sometimes been proposed would even prevent different dialects of English from being considered related, since there are irregularities in the correspondences. But we needn't and shouldn't suppose that all splitters are such extremists. Even the less extreme splitters seem unable or unwilling to let probabilities enter into their assessments.

Some years ago I proposed that we should distinguish four kinds of linguistic classification: 'probable relationship', 'probable truth', 'established relationship', and 'established truth' (Lamb 1959). Of particular interest here is the contrast between the 'probable truth' type of relationship, which one might suppose all scientifically inclined linguists would want to achieve, and the 'established relationship' type, which the majority of comparative linguists actually pursue. I remember discussing the matter with splitter friends, who were steadfastly unwilling to entertain – even apparently unable to imagine – the desirability of the 'probable truth' type of classification. Only the 'established relationship' type made any sense to them. The other possibilities were evidently inconceivable. It makes no difference that the established relationship can be taxonomically false, as Greenberg and I and others have pointed out repeatedly; for example, it can be established that German and French are genetically related and that Italian and Greek are genetically related, but it would be taxonomically false to set up these pairs as genetic groupings.

What kind of cognitive pattern is operating here? This kind of absolutism is appropriate to some kinds of thought process; for example, to jury deliberations in criminal cases, where the accused is considered innocent until proven guilty beyond a reasonable doubt. I recall being on jury duty recently and being instructed by one of the attorneys that we were not even supposed to *believe* that the accused was guilty unless the guilt was proved beyond a reasonable doubt. I found that kind of thinking impossible to accept. For surely it makes more sense, if the weight of evidence makes it, say, 90% likely that the accused is guilty, to believe that he probably is guilty; but by the rules of the court one must nevertheless return a verdict of 'not guilty' since the case has not been proved beyond a reasonable doubt. In other words, there is a clear difference between what one comes up with as a verdict, following the established procedural rules, and what *one believes is probably true*. People of a certain cognitive style may find this line of reasoning impossible to accept. Some of them, perhaps most of them, also find it impossible to accept the hypothesis that certain languages, say those comprising Dene-Caucasian, are *probably* genetically related even if their relationship has not been established beyond a reasonable doubt.

Let us be explicit about an idea implied above: It seems to be the case, for many people, that their cognitive style so pervades their thought patterns and belief system that no amount of further thinking can change the situation. If this is so, we are not talking

about matters that can be settled by discussion or debate, even though the points are debated – in a futile exercise – repeatedly. It makes as much sense as expecting Protestants and Catholics in Northern Ireland (or elsewhere) to settle their theological differences by a process of rational debate; or expecting an orthodox Jew to be persuaded by intellectual argument that Jesus was the Messiah or that the Qur'an is the source of ultimate Truth.

A person's cognitive style leads to certain thought patterns and beliefs, and the diversity of cognitive styles provides a diversity of opinions and kinds of investigation. We could say that human life is enriched by such diversity. But, in ignorance or rejection of this point of view, we get continuation of the debates, and more: We get the involvement of animus. It evidently arises since cognitive style is not only basic to thinking but is connected to emotion, even to ideas of good and evil. In reading the invective that has been directed at Greenberg's classification of the languages of North America one discerns that those who have produced it believe themselves to be not only superior in knowledge or intellect (a questionable belief at best!) but also morally superior. I am suggesting here that a more enlightened response would be to see that Greenberg was operating with a different cognitive style, and that within the rules of that style he was doing so very skillfully, even admirably. By the same token, those who work with the fine details of close-range linguistic taxonomy are operating with their cognitive style, and their work is also often admirable. To get a full picture of language on this planet we need both the broad strokes and the finer details. Those who work on the one or the other would do better to be glad that others are working in complementary areas, rather than finding fault with them.

Why, then, do different people operate with different cognitive styles? As we are talking about basic patterns of thought, the answer has to come from a consideration of brain structure and function. That is not to say that we have the answers yet, only to point to where we should be looking for them. It appears that people can favor one portion of the brain over others in their mental activity. We have those who prefer to work visually, others auditorily, some depend heavily on verbal thinking, others prefer visual imagery. The important difference between left-brain and right-brain thinking also plays a role, to some extent overlapping with the differences just mentioned: Those who work well with visualization are likely also to use their right-brains effectively; those whose thinking is heavily language-dependent are likely to favor their left brain.

The absolutist – all or nothing – type of thinking characteristic of splitters appears to be a product of heavy reliance on left-brain operation. From a growing body of evidence it is becoming more and more apparent that there is a sound basis for the casual observations and conjectures we often encounter about left-brain and right-brain thinking: that left-brain thinking is more analytical, while right-brain thinking is more intuitive and holistic. The left hemisphere is the home of that part of linguistic structure – most of that traditionally studied by linguists – that operates by making sharp contrasts. (Even the majority of left-handed people use the left hemisphere for this kind of information.) The importance of contrast is perhaps most evident in phonology, which is subserved largely by Broca's area (mainly for production) and Wernicke's area (basically for recognition), in the left cerebral cortex (Lamb 1999). Thus it is very important for linguistic communication that /p/ be distinct from /b/, and even segments which are intermediate phonetically between these two are perceived as one or the other – binary perception of

continuously varying phenomena. On the other hand, some aspects of phonology are subserved by corresponding portions of the right hemisphere. These include intonation, stress, duration of vowels – phenomena where sharpness of contrast plays a more peripheral role. Related to this observation is the fact that much of musical processing goes on in the right hemisphere – but for trained musicians, who are perhaps more likely to intellectualize their musical experiences, there is also left hemisphere involvement (Wallin 1991, Lamb 1995).

Although linguists have traditionally concentrated on those aspects of language that are processed by the left hemisphere, increasing attention is being given to properties that are subserved mainly by the right hemisphere. These include not only prosodic features but also pragmatics and much of semantics. A person with damage (perhaps resulting from a stroke) in the upper posterior left temporal lobe will be unable to understand what is being said to him, but he may get the emotional import. On the other hand, a person with right-brain damage is likely to understand every word, yet not get the point. For example, if you tell him a joke, he will understand every sentence but won't 'get the joke'. And he will often be unable to relate what is being said to the discourse or situational context.

People whose thinking is dominated by the left brain, it appears, are more likely to engage in thinking patterns that involve sharp contrast. They make good analytical linguists. On the other hand, too much dominance of the left brain, or rather insufficient balance from right hemisphere activity, can lead to extremism in operating with contrast. Here we find thinking patterns that can be given more or less self-evident names, like *EITHER-OR THINKING*, *ALL-OR-NOTHING THINKING*, the *SINGLE-CAUSE FALLACY*, the *SINGLE-PURPOSE FALLACY*, and thinking patterns that rely heavily on words rather than concepts, such as the *ETYMOLOGICAL FALLACY*, in which one purports to find the true meaning of a word in its etymology, for example *dis-ease*. In the sphere of religion, such heavy reliance on language renders them more likely to be fundamentalists. If their 'religion' is an academic one, there may be a similar type of fundamentalism that makes them sticklers for following the rules of standard methodology, unreceptive to innovative methods or imaginative ideas.

The left brain is also very 'concerned' with maintaining the coherence of the belief system, even to the point of rejecting data that doesn't fit the existing system of beliefs. Ramachandran and Blakeslee (1998) provide some striking examples of people with brain damage, demonstrating that the right brain is very accepting of new data while the left brain operating without benefit of the right brain (because of damage) rejects even very obvious new data that doesn't fit into the previously existing belief system.

Further evidence of the heavy dependence of the left brain on verbal forms of thinking, as opposed to the more empirical right brain, comes from some experiments conducted in Russia using simple logical syllogisms (Dolinina 2001). In this experiment, subjects underwent temporary suppression of either right or left hemisphere. One syllogism used in the experiment was

All monkeys climb trees.
The hedgehog is a monkey.
Does the hedgehog climb trees, or not?

Subjects with suppressed right hemisphere usually said "Yes", along with some of the controls. Those with suppressed left hemisphere gave answers like, "The hedgehog is not a monkey", sometimes with indignation.

The right brain appears to subserve holistic thinking as well as intuition, and to be less concerned with the need to find overwhelming evidence to support beliefs. An excessive reliance on right-brain thinking with too little left hemisphere involvement, results in a person who is excessively gullible and perhaps wildly imaginative.

It is clear that both of our hemispheres have much to offer our thinking, and that the best thinkers among us will be those who use both to advantage.

From a neurological point of view, it appears that the contrast between the types of processing of the two hemispheres is related partly to the fact that there is a greater prevalence of inhibitory structures in the left hemisphere, so that a phoneme, for example, upon being recognized, inhibits its competitors, resulting in a heightened sense of contrast (cf. Lamb 1995, 1999). While this finding is interesting it only accounts for part of the difference between the processing activities of the two hemispheres.

Although the foregoing remarks have concentrated on differences between right-brain and left-brain thinking, this is only one of the contrasts operating to provide the variety of cognitive styles that we observe among our colleagues and students – Sternberg (1997) identifies no fewer than thirteen thinking styles. I have emphasized it here not only because it is especially interesting but because I believe it plays a key role in the differences between splitter-thinking and lumper-thinking. That is, lumper-thinking is right-brain driven, while splitter-thinking, like most thinking that goes on in academic circles, relies heavily on left-brain activity.

It must also be emphasized that we are not dealing here with a simple dichotomy. It is not necessarily the case that a person characteristically uses either left-brain or right-brain activity. The best brains are evidently those with facility in both hemispheres. The extreme form of left-brain thinking is that of people with a deficiency of right brain activity; and vice versa.

It is to be hoped that people of different cognitive styles will become more accepting of one other. The difference between the two kinds of thinking discussed here is not a matter of knowledge nor of morals. Therefore, the differences will not be resolved by ordinary intellectual argument, nor is there any justification for animosity toward ideas which may seem unacceptable within the narrow confines of one mode of thinking.

References

Claxton, Charles S. & Patricia H. Murrell. 1987. *Learning Styles: Implications for Improving Educational Practices.* ASHE-ERIC Higher Education Report No. 4.

Dolinina, Inga. 2001. "Verbal reasoning and brain hemispheres." *LACUS Forum XXVII.*

Lamb, Sydney. 1959. "Some proposals for linguistic taxonomy." *Anthropological Linguistics* 1: 33-49.

----------. 1995. "Meaning in language and meaning in music." *LACUS Forum XXI:* 269-277.

----------. 1999. *Pathways of the Brain: The Neurocognitive Basis of Language.*

Amsterdam & Philadelphia: John Benjamins.
Ramachandran, V.S. & Sandra Blakeslee. 1998. *Phantoms in the Brain: Probing the Mysteries of the Mind.* New York: Morrow.
Sternberg, Robert J. 1997. *Thinking Styles.* Cambridge University Press.
---------- & Li-Fang Zhang (eds). 2001. *Perspectives on thinking, Learning, and Cognitive Styles.* Mahwah & London: Lawrence Erlbaum Associates.
Wallin, Nils. 1991. *Biomusicology: Neurophysiological, Neuropsychological, and Evolutionary Perspectives on the Origins and Purposes of Music.* Stuyvesant, NY: Pendragon Press.

To what extent is Paleolinguistics an Art and to what extent a Science? What Constitutes Scientific Evidence in Linguistics?

by Paul Whitehouse
London, England
email: paul_whitehouse@talk21.com

As a history graduate who writes novels I have always argued that science is where you deal with reality and art is where you make it up. On this basis "arts" subjects such as history belong with the "true sciences," and the differences between them are simply matters of precision. Each strives for objective truth; where subjective judgments have to be made this is in direct proportion to the shortfall between the data required for objective judgment and the data available.

Paleolinguistics is a science too by this criterion. Always we are seeking the best possible explanation of the data. Perceived differences in methodology are simply adjustments to the limitations of the data. The end goal is the same: to account as precisely as possible for the data you have. Never, under any circumstances, are you allowed to make *anything* up.

A more narrow definition of science was offered by Thomas Kuhn, who said that science was that which can be falsified. By this criterion too taxonomic paleolinguistics is a science, since a classification can always be falsified by producing a superior alternative. Where disputes arise is in the failure to agree whether such an alternative has been demonstrated, and these disputes are rooted in the deficiencies of the data not the logic of falsification.

Nevertheless, some paleolinguists claim that their work is intrinsically more "scientific" than that of other paleolinguists, and this is meant pejoratively. Such contentions give rise to a whole series of important questions. For instance, what exactly is meant by "scientific," what is it about science that makes paleolinguists so anxious to be considered scientific too, and to what extent does any or all paleolinguistic work meet the criteria applied by scientists to their own disciplines?

As implied above, what makes science so respectable (in every sense) is its objectivity, its testability, and its formality. Each of these serves to create strong boundaries that separate the speculative from the truly known, and keep the former from contaminating the latter. Where paleolinguistics imitates the sciences is in the formulation of reconstructions and sound laws. The presence or absence of these is taken as determining whether or not paleolinguistic work is scientific, and so respectable—or not. The term "sound law" is a conscious borrowing of the language of science, and is intended to make these laws sound as though they have the same validity as physical laws. In reality, however, they have only the terminology in common.

The laws governing optics, thermodynamics, and so on, are descriptions of physical reality; they apply without exception, can be quantified, and can be demonstrated experimentally. What's more, they can be falsified. There are outcomes that are logically impossible as long as these "laws" are valid, so experiments can be devised to achieve these impossible outcomes, whose success will invalidate these laws and whose failure will confirm them. What underpins these laws is that they have a physical basis, wherein physical

properties are determined by physical form. An atom of oxygen combined with two atoms of hydrogen will have the properties of water, and only water, always and everywhere—and these properties are a consequence of atomic structure.

"Sound laws," however, have no such physical basis. They describe systematic changes that have taken place over time, but on an arbitrary basis. There is nothing about the sounds involved to demand that these particular changes *had* to occur, or even that any change had to occur at all. Though trends may be identified, they are no more than that. There is no physical mechanism at work that can be tested experimentally or from which one can generalise about other sounds in other environments.

The upshot of this is, first, that sound laws are local in time and space. A sound law identified in Dravidian during a particular phase of its development does not automatically have to apply at other times, nor does it automatically have to apply in, say, Niger-Congo at all. Secondly, because they are not determined by the physical properties of the sounds, sound laws can not be universally binding. There can always be exceptions, and there nearly always are because their occurrence defies no physical stricture against them. Nor is there any requirement for these exceptions to be explicable (since the laws they break are underpinned by no binding mechanism). Indeed, there is no physical reason why there should even *be* sound laws at all! These exceptions are almost certainly more common than is traditionally accepted, if only because of the practice of dismissing non-regular correspondences as *ipso facto* accidental. Finally, although a sound law can be falsified by devising a better sound law, it cannot be falsified by the exceptions to it; nor does the absence of sound laws falsify a classification.

From this one could argue that there is nothing scientific about sound laws beyond the name, but such is the kudos attached to even the appearance of being scientific that some historical linguists argue that the purpose of historical linguistics *is* reconstruction. Certainly there is a value in sound laws, to the extent that they are based on patterns that do genuinely exist and on the basis of which it is often possible to extrapolate where data have been lost. Nevertheless, it is important to bear in mind the limitations of reconstruction, in theory and in practice, and the ways in which reconstructions limit the degree to which paleolinguistics can employ—and insist upon—the methodology of the physical sciences. The idealisation of reconstructions and sound laws creates false expectations as to what can—or should—be achieved. This results in the creation of a taxonomic "no go area" between the realms of absolute acceptance and absolute rejection, within which classification is declared impossible to achieve and heretical to attempt.

Yet the fact remains that classification precedes reconstruction both sequentially and hierarchically and is necessarily not subject to the same standards of "proof." Reconstruction is impossible without classification, but not *vice versa*. The failure to identify sound laws does not necessarily invalidate a classification. It may be that the relationship concerned is so ancient that there are no longer enough shared retentions on which to base sound laws. That does not "disprove" the relationship. It does limit the degree of confidence we can feel that the conclusions we reach are correct.

As I have said, a proposed relationship can always be falsified by producing equivalent evidence (or better) for an alternative relationship or relationships. Falsification is implied in claims that a body of evidence put forward is no more than chance resemblance; that a whole series of bodies of evidence—of equal weight—could be put forward in

support of other relationships. This is a statement that any scientist would recognise as "scientific." Sadly, it is rarely followed by what a scientist would recognise as the obvious next step, which is to go ahead and produce such bodies of evidence. Instead, the statement is couched in terms of a theory, rooted in the mathematical analysis of linguistic relationships, which is taken as negating the need to produce alternative evidence—and for which, consequently, no body of evidence exists to support or undermine it.

This brings us to the other arena in which the "more-scientific-than-thou" argument is fought out: mathematics. Nothing offers more in the way of certainty than mathematics. There is good reason for this, since mathematical relationships are capable of irrefutable proof, and if we did not prefer certainty to uncertainty it is unlikely that we would devote so much of our lives to the pursuit of knowledge. Hence the urge to quantify retentions from common origin in ways that seem to support our own points of view. This is certainly a scientific way of working, but it does beg several questions: how valid are the ways in which we quantify genetic relationships, what are the limitations of statistical analysis, and what are the implications for the precision of our taxonomic conclusions in terms of relative and absolute time depth?

The complexities of glottochronology, lexicostatistics, and so on, require a paper of their own, but one thing that unites them is that they all rest on a false premise: that the rate of linguistic change over time is a constant. It is clearly not, nor is it arbitrary in nature. Each sequence of language changes arises from a unique set of circumstances, and each set of genetic relationships that ensues is also unique. Thus the phenomena we seek to analyse mathematically are extremely complex, yet the methods we use are too often crude and simplistic.

The golden rule of mathematical analysis is that you must always compare like with like. In linguistics, however, this is nearly always impossible because no two languages or language families are alike in terms of data available, prevalence of synonyms, obviousness of borrowings, number of family members, and their relative diversity (temporal and spatial). Even attempts to get round these problems, such as the comparison of standard 100- or 200-word lists, are fraught with distortions. For instance, when a language has two words with the same meaning, each of which is cognate with the word in a different language, what do you do? You can ignore one or the other (but which, and for what reasons?), count neither or count both. Do the second two alternatives affect the percentages? Or, what about when words for "red" in one language are cognate with words for "blood" in another. Do you count these as one comparison, two comparisons—or even four (red-red, blood-blood, red-blood, blood-red)? Each option has different mathematical implications.

This sort of resemblance pattern is the norm not the exception, so any analytical system that ignores such anomalies risks simplifying the problem beyond the point where the results cease to be meaningful. This applies equally to the question of chance resemblance, which is is subject to exactly the same parameters. Also, on what basis (apart from convenience) do we say that all cognates have equal weight? As with sciences, the dilemma is whether to limit the analysis to calculations that may be too simple to be of any use or to seek to complicate the model to match the phenomenon under investigation and risk distortions of a different kind.

What all of this means is, first, that absolute dating of genetic relationships is impossible to achieve with any precision. Second, it means that even relative dating is problematic. Thirdly, it means that there can be no quick fix. This is not to dismiss all mathematical analyses out of hand—on the contrary, I believe they offer extremely valuable insights—but there is no one statistical measure that will give us the answer we seek. We need to analyse the data in as many different ways as possible, and hope that these tests will confirm the same classification again and again.

This in effect is what happens when a paleolinguist makes a judgment as to the merits of various taxonomic proposals. As became apparent when computer programmers first tried to model the human decision making process, the human brain is actually very good at prioritising the information it is bombarded with and distinguishing the significant from the insignificant. Unfortunately, this intuitive faculty is far too inconsistent a process to be truly scientific, even though scientists themselves have recourse to it when they think no one is looking.

It is more prevalent in paleolinguistics, because there are always so many variables to consider and so few analytic tools to fall back on, and it is for this reason that linguistics is sometimes called an art. That does not make artists of linguists, however, even those whose judgment is out of the ordinary. After all, these people create nothing. They simply account for data that already exist. Neither is it scientific when questions are decided on the basis of subjective judgments as to *whose* subjective judgment counts more in any given case.

Sadly this is the level at which classificatory debates tend to be settled, on the basis of consent rather than "proof." In maths or physics proof is a matter of irrefutable logic or repeatable experiment, whereas in paleolinguistics it is a question of, 'Does Professor X accept my case?' Not, 'Am I right?' but, 'Can Dr. Y bring himself to say that I am right?' Not, 'What can I prove?' but, 'Whom can I convince?' In this linguists have more in common with lawyers – or, worse, with politicians – than with either scientists or artists.

Inevitably decisions end up being made according to different sets of criteria, since Professor X's willingness to accept the evidence presented to him may be influenced by all manner of nonlinguistic considerations. Prof. X is human too after all, whatever his devotion to objective truth, and it is not always easy to give up something in which you have invested so much. Here perhaps is one reason for the popularity of reconstructions and sound laws. They have the appearance of objectivity, which deters people from attacking them, and emotional attachments to them are easier to justify and harder to spot.

Of course, there may be another motive at work here. Viewed from a psychological standpoint the word "law" is a major give-away. Laws offer security, a refuge from doubt. What's more, they protect the innocent and punish the guilty, while at the same time conferring legitimacy on those who administer them. Upholding a law carries visibly high status. Grimm's Trend does not have the same ring to it.

It is always regrettable when questions of legitimacy and status threaten to override questions of objective truth, but this is part of the human condition—and scientists are no more exempt from it than paleolinguists. More regrettable still is the attitude, encouraged by misleading comparisons with mathematics and the physical sciences, whereby 'that which is not 100% true is 100% false.' This misconception misses the essentially relative nature of

classification and discourages the spirit of enquiry without which no discipline will ever be taken forward.

I described the difference between history and science as a matter of precision, and it is with this subject that I would like to conclude. The cognate sets put forward by John Bengtson and Merritt Ruhlen[1] to demonstrate common origin for all the world's languages include MALIQ'A 'to suck(le), nurse; breast.' This is an excellent illustration of the complexity of the variables, phonetic and semantic, that have to be taken into account when analysing—statistically or otherwise—the evidence for genetic relationship. I am not aware of a statistical model of sufficient complexity having yet been devised that would either confirm or deny this otherwise solid cognate. Of all the global cognate sets put forward by Ruhlen and Bengtson, MALIQ'A is by far the strongest for the very good reason that it is not a global cognate. The forms presented are restricted to those from the phyla sometimes grouped into a larger Borean super-phylum (Amerind, Dene-Caucasic and Nostratic)—and a very powerful piece of evidence for Borean it is too. Borean is, necessarily, of more recent common origin than Proto-Human, and it is most reassuring to see that its relative "novelty" is reflected in the quality of the evidence that can be adduced for it. If a global subgrouping is ever to be identified it will be on the basis of sub-global cognate sets such as MALIQ'A, which gives us hope that others can be identified too.

1. Bengtson, John D., and Merritt Ruhlen. 1994. "Global Etymologies." In Merritt Ruhlen, *On the Origin of Languages: Studies in Linguistic Taxonomy,* pp. 277-336. Stanford: Stanford University Press.

What Constitutes Scientific Evidence in Paleolinguistics?

by Winfred P. Lehmann
Linguistics Research Center
Austin, Texas

The evidence required in paleolinguistics is comparable to that used in historical linguistics, although it is not attested in written documents. The phonological structures of reconstructed languages, their morphology and syntax, and their lexicon are examined with the procedures that are applied in historical linguistics. If such data have been determined rigorously, and if the procedures are applied with the same rigor, the results will be similarly credible. We may review the basic procedures, and then examine some applications of them. The principal procedure has been excellently described by Antoine Meillet (1967). As the title (*The Comparative Method in Historical Linguistics)* indicates, the basic method is the comparative method. Briefly, in its use elements of three or more languages are examined to determine "rules of correspondences." The elements compared must be from the same family and from comparable stages. The comparison is applied to items of phonology, morphology/syntax, and semantics.

The method was applied by Jacob Grimm when he determined rules of correspondence between items in Greek, Latin, and Germanic in words like Greek πατήρ, Latin *pater*, and Gothic *fadar* 'father'. He aligned the corresponding items, e.g. Gk π, Lat. *p*, Goth *f*. Thereupon an earlier form of Germanic was reconstructed with **p* corresponding to the same item in Greek and Latin. Historical rules were then proposed, with *p t k* labeled ***Tenues*** and the Germanic equivalents ***Aspiratae***, so that the correspondences could be stated between the parent language and Germanic as **T > A** (Grimm, 1822). The procedure often leads to supplementary rules, such as those relating the second element in Latin *spuo* etc. and English *spew*. On the basis of such rules the earlier form of the language is reconstructed. Often for historical purposes and clearly in paleolinguistics, languages are reconstructed for which there are no written texts. One such is Proto-Germanic, where the prefix indicates that the language is not attested. The prefix Pre- may also be used, either before Proto- or alone, to indicate a still earlier stage than the proto-stage. The comparative method has gained credibility because languages have been discovered that support reconstructions that were made before those languages were known. A well-known example is Hittite, in which items were found with elements now symbolized with *h* that correspond to segments in reconstructions that had been proposed earlier in accordance with the laryngeal theory, such as Hittite *hanti-* : Greek ἀντί 'against'; Hittite *pahhur :* Greek πῦρ 'fire'.

It is highly important to note that the method does not simply set side by side words of similar meaning. To illustrate the requirement of determining rules of correspondences Meillet used the example of Armenian *erku* 'two', for which rules can be established to indicate its relationship to Latin *duo*, etc. These are too extensive to give here but may be examined in his monograph (1967: 18, 46-7). He there asserts: "It is not with similarities of forms that we work when we compare languages of the same family, but solely with rules of correspondences."

A second procedure is the method of internal reconstruction. It involves items within one language. It is based on the observation that change does not affect all similar items. An example I have often used is the intervocalic maintenance in General American English of voiceless [t] in words like *button* as compared with the voiced counterpart in words like *bottle, bottom, batter*. By the rule, intervocalic [t] is maintained before unstressed syllables ending in an alveolar nasal, that is, a nasal of the same position as [t]. We then reconstruct earlier forms of words like *bottle*, etc. with voiceless [t]. Happily we here have corroboration by comparing the cognates in other forms of English. The method was used by Verner to explain the voiced consonant in Gothic *fadar* and many other Germanic words. He concluded that such intervocalic voiced fricatives resulted when the accent did not stand on the preceding syllable in an earlier form of the language. On the basis of the method **t* could be reconstructed in an earlier form of the Germanic word 'father', and also a variable pitch accent. Fortunately Verner's assumption could be supported by the type and the position of the accent in Sanskrit and Greek cognates, as may be noted from the forms of 'father' above. Internal reconstruction is highly important in paleolinguistics when we do not have related languages to compare.

By a third procedure we look for residues, a procedure especially important for reconstructing the morphology and the lexicon. The procedure is based on the observation that items learned by children before they are aware of the rules of their language often are treated as individual items and then are maintained rather than subjected to the general rules that children apply when they increase their vocabularies. English has many examples, such as the noun plurals *children, women, men*, and the past tense forms of verbs, such as *ate, sang, went*. In time such "irregular" forms may be lost as was *kine* as the plural of *cow*.

We may now examine the procedures applied in the reconstruction of two paleolanguages, Nostratic and Eurasiatic. We are fortunate in having two recent books: *Nostratic,* with the sub-title: *Examining a Linguistic Macrofamily* (Renfrew & Nettle, 1999), and *Indo-European and its Closest Relatives,* by Joseph H. Greenberg (2000). It may be unfair to compare their procedures with those applied in the reconstruction of Proto-Indo-European, in view of the two centuries of extensive scholarship dealing with the Indo-European family. But unless we apply the procedures rigorously, paleolinguistic research may never achieve credibility.

First we may note the requirement of comparable stages. Proto-Indo-European is reconstructed on the basis of languages attested in the second millennium B.C. It may then be dated in the third millennium, with possible extension to the fifth. No one assumes that date for Proto-Afroasiatic, since we have Egyptian and Akkadian texts from the third millennium. The two languages differ from one another considerably so that Proto-Afroasiatic must be dated from a much earlier time. For reconstructing Nostratic a far earlier form of Indo-European must then be reconstructed than that in the well-known handbooks. Bomhard (1999: 71) in his paper in *Nostratic* dates "the unified Nostratic parent language...to between 15,000 to 12,000 BCE." If Nostratic is to achieve credibility, early proto-languages of the six families: Indo-European, Afroasiatic, Dravidian, Altaic, Kartvelian, and Uralic-Yukaghir must be reconstructed.

Moreover, by Meillet's fundamental principle we must establish "rules of correspondences." The papers in *Nostratic* resulted from a symposium that centered on

examination of 124 words in *The Nostratic Macrofamily and Linguistic Palaeontology* by Aharon Dolgopolsky (1998). The words were taken from a far larger number he has examined. It may be noted at once that the basis for the reconstruction consists of lexical items, not rules of correspondence.

Moreover, in support of many of the entries, forms are cited from only some of these languages, such as #17 'cereals', with items from only three families. Further, like these, reflexes are taken from later dialects, not the reconstructed parent language. In support of the reconstruction of #17 'cereals', an Arabic word is cited for Afroasiatic and a Georgian for Kartvelian; in view of Meillet's example, *erku*, such a procedure hardly leads to credibility. Only for Indo-European is there a reconstructed form; it is based on Hittite reflex *halki* 'grain' and Anatolian Greek ἄλιξ 'spelt'. Jaan Puhvel (1991, Vol. 3: 35-9) states that "Indo-European root connections are improbable." Further, because the Greek form is found in Anatolia, it may be a borrowing from Hittite. The reconstruction of Nostratic **gaL* (plus an unspecified vowel) for 'cereals' then is scarcely credible. It is even more important to note that the requirements in use of the comparative method have not been followed. As in any scientific pursuit, credible results require strict observation of the established procedures.

The reconstruction for #22 'wattle' is more credible, because it is based on reconstructed forms from all families but Uralic. We may assume that when similar evidence is available for items in all six families, the proposed Nostratic forms will also be more generally credible.

As noted above, Meillet insists on determining rules of correspondences rather than relying on similarities between words. Yet for the reconstruction of words, he states one must "not be content with comparing small root elements" (1967: 51). This, however, is precisely what we find in Greenberg's book. For example, the section on Grammatical Evidence for Eurasiatic (pp. 61-239) consists of sections with labels like the following: "**1. First-Person M... 2. First-Person K... 3. First-Person N... 4. Second-Person T....**" (2000: 61-71). As with treatment of the lexicon, much work must be completed before credible reconstruction of the grammar is carried out. The material proposed in the books of Dolgopolsky and Greenberg then may be assessed in accordance with "what constitutes scientific evidence in paleolinguistics." For most portions of Nostratic or Eurasiatic, such evidence remains to be determined.

As an example of the use of the method of internal reconstruction, we may cite the inferences based on two forms for 'fire' in the Indo-European dialects, as exemplified by Hittite *pahhur* and Latin *ignis*, as well as comparable forms in other dialects, and for 'water', by Hittite *watar* and Sanskrit *āp-*. We find such pairs in active languages where nouns and verbs fall into two classes, active or animate and inactive or stative. We also find "opposite" meanings for reflexes of roots such as **nem-* in Greek νέμω 'distribute' and Gothic *niman* 'take', in accordance with a patterning known as version.[1] These and other items found in reconstructed Indo-European are characteristic of active languages. On the basis of such internal evidence in Proto-Indo-European we conclude that it developed from an active language. The conclusion derives support from the twofold

1. I.e., reversal or inversion of meanings. [Ed.]

gender classification of Hittite nouns into common and neuter. Further support is provided by the absence of the passive voice in it, as well as by other characteristics. We may now examine the proto-languages of other families for evidence of active structure, such as the twofold classification of nouns in some Afroasiatic languages. But such reconstruction has not yet been undertaken; work in Nostratic and in Eurasiatic has been largely confined to examination of the lexicon and the phonology.

Last, we may note briefly some examples of residues. These have been identified among the words that are assumed for time of transition from hunting-gathering to settled existence. Many of them are monosyllabic and differing in pattern from one another, such as the words for the earliest domesticated animals, the cow, *$g^{w}ou$-, the sheep *owi-, the pig *su-, and so on. Other such forms are found in compounds, such as Sanskrit *dámpati*- 'master of the house', where the reconstructed *dem- 'house' is probably equivalent to the verbal root *dem- 'tame'.

On the basis of these three types of evidence in paleolinguistics – correspondences, internal reconstructions, and residues – we reconstruct an early proto-language like Nostratic or Eurasiatic much as we do Proto-Germanic or Proto-Indo-European. Obviously we do not have the same extent of evidence as we do for the reconstruction of these and other written languages. We may take comfort from the knowledge that societies in the hunting-gathering stage do not have as extensive lexicon, or probably grammar, as do societies with more complex living conditions. With these conditions, the evidence that we apply is comparable in scientific validity with that applied in the reconstruction of proto-languages like Proto-Altaic and those of the other families assumed in the Nostratic and Eurasiatic macrofamilies.

References:

Bomhard, Allan R. 1999. "Review of Dolgopolsky's *The Nostratic Macrofamily and Linguistic Paleontology.*" In Renfrew & Nettle 1999, pp. 47-74.

Dolgopolsky, Aharon. 1998. *The Nostratic Macrofamily and Linguistic Palaeontology.* Cambridge (UK): McDonald Institute.

Greenberg, Joseph H. 2000. *Indo-European and its Closest Relatives: The Eurasiatic Language Family* (Vol. 1. Grammar). Stanford, CA: Stanford Univ. Press.

Grimm, Jakob. 1822. *Deutsche Grammatik.* Berlin: Dümmler.

Meillet, Antoine. 1967. *The Comparative Method in Historical Linguistics.* (Trans. by Gordon B. Ford, Jr.) Paris: Champion.

Puhvel, Jaan. 1991. *Hittite Etymological Dictionary.* Berlin, New York, Amsterdam: Mouton de Gruyter.

Renfrew, Colin, and Daniel Nettle (Eds.) 1999. *Nostratic: Examining a Linguistic Macrofamily.* Cambridge (UK): McDonald Institute.

On the Rules of Discourse for Paleolinguists in Relation to Notational and Transcription Systems

by Paul Whitehouse
London, England
email: paul_whitehouse@talk21.com

Colleague Bengtson's brief was that I should address the question, "What should be the rules of discourse and polemics for paleolinguists" in relation to notational and transcription systems? I took this to mean that someone else was tackling the other aspects of this question, such as: how should we frame our arguments and marshall our evidence? Or: How do we reconcile the need for candour with the conventions of civilised society? The microbiologist Francis Crick once remarked that, "Politeness is the poison of all good collaboration in science. The soul of collaboration is perfect candour, rudeness if need be ... [I]n science criticism is the height and measure of friendship. The true collaborator points out the obvious with due impatience. He stops the nonsense." While that does not mean we should rip out each other's throats like rottweilers, Crick has a point. And there are a few candid things that need to be said about the way we write down our data.

Years ago, when I first became interested in linguistics and was introduced to the IPA chart, I thought: What a brilliant idea. A standard system of notation must make life so much easier. I also thought: How obvious. After all, that's how they do it in chemistry. Years later, when I began to collect comparative data and learned that the "standard transcription principle" was more honoured in the breach than in the observance, I thought: How incompetent.

Nothing is ever that simple, however, and people rarely do things wrong without what seem like sound reasons at the time. An examination of what has happened since the first version of the International Phonetic Alphabet was published in 1888 bears this out.

The reasons for the failure of IPA (or any other system) to establish itself as an inviolable standard may be expressed in terms of supply and demand. On the supply side there were the physical constraints of typography. Any complex system of special symbols and diacritics like IPA was always going to pose problems for those without ready access to specialist printers, but in the absence of an alternative access was somehow obtained. Printed sources from before World War I tend to be consistent (and so predictable) in their transcriptions, allowing for the biases inherent in the languages of the various colonial powers.

The introduction of the typewriter, however, by allowing the local publication of linguistic evidence, accelerated the growth of local variation. It became more tempting to abandon a complex system of transcription that could only be printed by a specialist compositor thousands of miles away in favour of a revised system that could be accommodated by a standard typewriter, available locally.

At the same time there was a corresponding shift on the demand side. Whereas in the early days language data were collected and published for the benefit of scholars in Europe and North America, whose interest was in comparing newly discovered languages, the introduction of the typewriter coincided with an increasing desire to write languages down for the benefit of the language speakers themselves. This called for very different

orthographies, which did not need to be either phonetically precise (since the speakers already knew what their language sounded like) or similar to other transcription systems (since they were aimed solely at a single language community).

With this shift the needs of comparative linguists became secondary, and the reason for this is simple. The number of linguists studying wide ranges of foreign languages for taxonomic purposes has always been low. It does not take a very large language community to out-number us. Our priorities are different from those of the people who speak the languages we seek to classify. Thus, when compiling a dictionary of, say, Telefol it makes good commercial sense for it to suit the few thousand Telefol speakers who might buy it, at the expense of the dozen or so comparative linguists who might also want to use it.

Furthermore, in significant parts of the world (New Guinea, for instance) linguistic research is largely the preserve of missionary organisations like the Summer Institute of Linguistics. Their priority is to bring literacy to pre-literate societies; this involves devising writing systems that allow people to read the languages they already speak. The needs of academic linguists are secondary—and why not? The SIL is only doing what it exists to do, and if academic linguists want language data written down in a particular way, let *them* camp out in the jungle for months on end and elicit the data.

So, if there is only enough money to finance a one-way dictionary, it will be a Telefol-English dictionary that gets published. The only people able to use it will be those who already speak Telefol. Comparative linguists like you and me, wishing to know the Telefol word for 'finger nail', will have to trawl through the entire dictionary and hope we don't miss it the first time around. In a language where the plosive /g/ becomes a fricative intervocally and is always nasalised in word-final position, just the one symbol /g/ will be enough for the language speakers (who already know these things); linguists on another continent will just have to either make a mental note to this effect or retranscribe the lot, in each case hoping to limit the errors of transmission.

In this way comparative linguists became progressively marginalised—and so it will continue. There is great excitement at present regarding the potential for mass-dissemination of linguistic data via the internet, but let us not kid ourselves that comparative linguists will be the beneficiaries. All that will happen is that the "Telefol Dictionary Syndrome" will be repeated on a larger scale. It is the languages with dictionaries now available in print that will become available on the web too. These will select themselves by weight of numbers, without regard to taxonomic significance. Those languages with few speakers but massive taxonomic importance, for which there are no data currently available in print, will not be available on the web either.

The stark truth is that comparative linguistis lie at the bottom of the pecking-order, and if we want things done in a particular way we must must do it ourselves. No one will do it for us. But here too there are barriers to overcome. Even among themselves comparative linguists have contrived to transgress the "standard transcription principle", and here too it is in matters of supply and demand that the reasons are to be found.

Writing had already been established for thousands of years when linguists started comparing large numbers of languages, so it seemed obvious at the time to use and adapt pre-existing writing systems rather than create a wholly new one. "Multilateral comparison" was pioneered in Europe, so standard notational systems were based on either the Roman

or Cyrillic alphabets. These systems are of course biased towards a particular phonetic range and, as 'complete systems', were not designed to be modifiable.

In an ideal world a completely new set of symbols would have been devised, whose resemblances were guided by phonetic criteria and which would have been designed so that they did not protrude either above or below the base line (as with /**b**/ or /**p**/), leaving space for modification by diacritics. There was never a chance of such a system catching on; nor is there now.

Furthermore, there were those languages (Greek and Hebrew, e.g.) with long established alphabets and armies of specialists who simply had no need or desire to harmonise "their" scripts with any other. After all, a system that requires a special effort to master it represents a useful barrier to outsiders. The same "exclusion principle" may also explain the resilience of local transcriptional conventions. These arise as a consequence of isolation (Austroasiaticists do not look at African languages, nor Africanists at Austroasiatic), but once in existence they help to define the demarcation lines that some linguists seem to find so reassuring when non-specialists threaten to gate-crash the party. It should come as no surprise to learn that some of the hardest languages in which to obtain phonetically transcribed data are languages like Greek, Welsh, Chinese. The implication appears to be that if you don't already have these data you have no business to be looking for them; if you don't already know how these languages sound, you have no right to be asking.

No doubt paleolinguists employ similar 'masonic hand-shakes', and for the same reasons. All specialisations are conspiracies to exclude outsiders, it is said—even if we *do* have the right to want things done in ways that suit us. After all, now that computers and phonetic fonts are available the world over, we are finally in a position to control exactly how our work is presented. All that stands in the way of paleolinguists adopting a uniform system of notation is the will to do it.

But why does this matter? I have assumed from the outset that a single transcription for all language data is a desirable thing, but in view of how consistently this requirement is overlooked maybe a statement of the obvious is needed.

Language is primarily a tool for communication and it is absurd that linguists of all people should wilfully erect barriers to communication between one another, but this is what we do in violating the "standard transcription principle." The classification of thousands of languages involves obtaining data in thousands of languages and comparing them. In practice this means extra work, and anything that adds such a work-load is a barrier to research. Comparing a thousand languages in a single transcription is hard enough; how much harder with a hundred different transcriptions! The time to make the necessary adjustments between writing systems is before you start comparing, not while you are actually trying to do it. How much better if the adjustments did not need to be made at all.

So, a standard transcription is called for. What system should that be? Well, mine is as good a system as you will find—but you would expect me to say that. Everyone has his own preferences, and can be expected to champion them. Given the choice between a phonemic system and a phonetic I would argue for the latter. It is easy enough to add a brief guide as to what is contrastive and what is allophonic, and this is surely preferable to

being kept in ignorance as to what allophonic variation may lurk beneath the symbols presented to us.

I have particular reservations about the use of the symbols /**c**/ and /**j**/. In IPA /**c**/ is an unvoiced palatal plosive, and that is how it should be used. I understand why it should have been adopted as a shorthand for alveolar or post-alveolar affricates, but surely neither /**ts**/ nor /**č**/ are too time-consuming to write. The same applies to /**j**/ versus /**dz**/. I am even considering whether to revert to the IPA convention of showing both components of affricates (i.e. /**tš**/ and /**dž**/ instead of /**č**/ and ... whatever variant of /**ǯ**/ you prefer). These look rather horrible, but at least you know where you stand. With /**c**/ and /**j**/ you hardly ever know where you stand.

My other great preference is for /**aː**/ rather than /**aa**/ to denote long vowels. Here I am influenced by musical notation in which the distinction between a crotchet and two quavers is sacrosanct. If /**aa**/ is in any way bisyllabic it should marked as such.

In fact, it doesn't matter *what* transcription we use as long as everyone uses it and sticks to it. In any case, we remain free to write things down however we like in the privacy of our own homes. What matters is how we share information with others. It is for the people who plan to share the information to get together and agree among themselves.

Failing that, can we at least explain the symbols we use? Every moment spent trying to work out whether /j/ is a semi-vowel, or a voiced palatal plosive, or a postalveolar affricate is time wasted. And can we also explain our transcriptions in phonetic terms, and not by comparison to a sound in some other language we don't know either. I am particularly fed up with invocations of "English" pronunciations that you never hear spoken in *my* corner of England.

It would also help if we could free ourselves of "the curse of the one-way dictionary." I understand why linguists prefer to list data according to "subject language order," since there is never an exact semantic match between any two languages. But what you end up with is a dictionary that can be used only by someone who already speaks the language. Anyone wishing to classify fifty such languages has to waste weeks putting together something before they can even begin to compare. Even then you are always left feeling that you have missed something vital. Lists should be ordered alphabetically, by meaning. It doesn't matter what language the meanings are written in, as long as it is a language sufficiently commonly used to justify the effort of learning it.

One last point. The merits of sharing are many and obvious. It helps no one if every linguist is obliged to duplicate exactly the same preparatory work, when much of the drudgery that underpins comparative research only actually needs to be done by one person, once. Whether or not this comes under the heading "rules of discourse" is a moot point; it wants repeating regardless.

Finally, a disclaimer: however strongly I may advocate the use of a common system of transcription, in practice I will prove as reluctant to toe the party line as any other linguist. This should come as no surprise. With those who preach to others it is so often a case of 'do as I say, not as I do!'

The Ever-green 'Beech'-argument in Nostratic Perspective

by Václav Blažek
Faculty of Arts of Masaryk
University, Brno
Czech Republic
blazek@phil.muni.cz

The "beech" tree-name played and plays a prominent role in almost all discussions on the Indo-European homeland and Indo-European dialectology. The conclusion about the limited occurrence of the tree-name in the western Indo-European dialects had to correspond with the limited diffusion of the tree in Europe. Now we know more and it is evident that this conclusion is not valid. Let us analyze the linguistic material in detail on the basis of our present level of knowledge, first in the core beech-branches:

Greek

Greek φηγός (from Il.), Doric φᾱγός (Theocritus) f. "a sort of oak with edible acorns / Quercus Aegilops". In principle, this tree-name may (or may not) be identified in the toponym *pa-ko* known from Pylos (PY An 427.2), cf. the place-names as Φηγός in Thessalia or Φήγεια in Arcadia (Aura Jorro 1993, 75-76).

Italic

Latin *fāgus*, gen. *-ī* (& *-ūs* after *quercus*, gen. *-ūs*) f. "beech". The tree-name also appears in the city-name *Fagifulae* (Livy XXIV, 20) and *Fagifulani* (Plin. *Nat. Hist.* III, 107) in Samnium, today Montagano, while the original name lives in the name of the church S. Maria a *Faifoli* (*PRECA* 6, c. 1967). Direct Latin borrowings probably occur in Basque *bago* (Navarro, Labortano, Roncalés, Suletino), *pago* (Vizcaíno, Guipuzcoano, Navarro), *phago* (Salaberry), *fago* (Labortano, Lower Navarro) "beech", *bagasta* "young beech-tree", *bagodi* "beech forest", *bagalia* "mast" (Agud & Tovar 1991, 152-53; Löpelmann 1968, 140) and surely in modern Celtic designations of "beech": Irish *faghvile* id. : *bile* "tree", New Irish *fáigh*, Welsh coll. *ffawydd* : *gwydd* "wood", sg. *ffawydden* = Breton *favenn*, pl. *fav*, *fao* id. (Buck 1949, 529, #8.62). Let us mention the witness of Caesar about the absence of the "beech" in the British Islands: *Est materia cuiusque generis, ut in Gallia, praeter fagum atque abietem* (*BG* V, 12).

Germanic

Germanic (all f.) **bōkō* > Old Saxon *bōke* "fagus, aesculus" (also the **-ōn* & **-jōn* stems are thinkable), Old High German *buocha* (~ *p-*, *-o-/-oa-/-ua-..*, *-hh-/-h-*) gl. "fagus", also "quercus, bedullanea, vibex", Old English *bōc*, pl. *bēc* "beech", while Old Icelandic *bók* id., Swedish *bok*, older Danish *bog*, later *bøg*, represent the consonant stem; further **bōkjōn* > Old English *bēċe*, Old Saxon *bōkia*, Middle Low German *bōke*, Dutch *beuk* "beech", and **bōkja-* > Old Icelandic *bœki* in *ǫlbœki* (poet.) "Bierfass", i.e. from "beech-wood", *bœki-skógr* "beech-forest" (de Vries 1962, 69; *EWAhd* II, 437-438).

Celtic

Bertoldi (1931, 286, fn. 2), Hubschmid (1933, 254f) and Pokorny (1956, 279-81) identified the continental Celtic base **bāg(o)-* in the following toponyms:

**bāgāko-* attested in *Bagacum* (*Itinerarium Antonini Augusti* 376,2, 377,1, 378,1), *Bagaco Neruiorum* (ibid. 380,7), Βάγανον ~ Βάσακον = *Βάγακον (Ptolemaius, *Geogr.* II, 9.11), later *Bavacum* (Holder I, 329; Billy 1993, 22); today *Bavay* in France between St. Quentin and Brussel; the same origin is proposed for the forest *Beiach* in Walperswil near Bern etc.; the suffix **-āko-* is productive in plant names too, cf. Old Irish *dristen-ach* "dumetum", Breton *drezecq* "roncière" (*LÉIA*, D-197).

**bāgonā* attested in *Val Bavona* in Tessin (Swiss) and *Baons-le-Comte*, lit. "beech forest of a count" (Seine-Infér., France).

**bāgustā* attested in *Bagusta..in pago Ambianensi*, today Amiens (AD 662; see Holder I, 332).

**bāgodiās* ("Buchenwälder") > *Baioies* (1182, 1213, 1228, etc.), today *Bavois* (Swiss); the adjectival suffix **-odi̯o-* occurs e.g. in Gaulish-Latin *Carant-odius,-odia*, Middle Welsh *-eid* and Old Irish *-de* (cf. Thurneysen 1946, 220-22).

**bāgantiā* attested in the river names *Baganza* (North Italy) and *Pegnitz* (Germany: between Bayreuth and Nürnberg) - see Holder I, 329, who derived it from **b^{h}eg-* "to run", but Greek φέβομαι "ich fliehe" indicates **-g^u-* which would imply Gaulish 'virtual' *+Bābantia*.

Mann (1984-87, 61) offered an original Brythonic addition, etymologizing Welsh *baedd*, Cornish *bāth* "boar" from **b^{h}āg-ed-* "mast-eater" (let us add Old Cornish *bahet* "aper, uerres" with unetymological *-h-*; see Campanile 1974, 11 who concluded "senza etimologia").

Albanian

La Piana (1939, 102-03) compared Greek φᾱγός with Albanian *bung(ë)* m., pl. *-a* "chestnut oak / Quercus sessiflora", cf. also Rumunian *bunget* "oak-thicket" (Orel 1998, 42). La Piana proposed the starting-point **b^{h}āngos*, while Hamp (1973, 1095; first apud Friedrich 1970, 108) derives it from **b^{h}āg-no-*, cf. the metathesis in *peng* "Pfand" < Latin *pignus* (Demiraj 1997, 112-13).

Iranian

Henning (1963, 68-69, fn. 2) perhaps definitively rejected Bartholomae's comparison of Central Kurdish (Mukri) *būz* "elm" which together with Gurani *wiz*, Talish *vizim*, Eastern Persian *guzm*, Zābulistān *γuzbe*, Central Persian (Khunsār) *vizvā* reflect Iranian **u̯izu̯ă-* (the change *ī* > *ū* is known in Kurdic, cf. *tūž* "sharp" vs. older *tīž* id. - see Bartholomae 1898, 271), and offered a more convincing Iranian supplement, namely Gilani *faγ / fiγ* "hornbeam", with an analogous development as Persian *faγfūr* "divine son" < **baga-putra-* (Bailey 1936, 1054 = 1981, 302; he mentioned the East Iranian provenance of this title in confrontation with Persian *pus* "son").

On the other hand, the attempt of Paxalina (1983, 26) to include here Wakhi *bānǰ* "oak", Munjan *vánʒiya* id. is wrong; these tree-names reflect Indo-Iranian **van-* "tree", cf. Old Indic *ván-* "tree" : *vána-m* "wood, forest, grove"; Avestan *van-*, *vanā-* "tree", Zoroastrian Pahlavi *(b)wn*, Persian *bun*, Buddhistic Sogdian *wnh*, Pashto *wana* "tree", Khotanese *banhya-* id., besides the specialized meanings in Khotanese *banījă-* "willow" or

"oak" (with the same suffix as *barīja* "crop, harvest", f. adj. from *bara-* "fruit, crop"), Shugni *wān*, Bartangi *wanōč*, Sarikoli *wanūǰ*, Yazgulami *waneg* "willow", Ormuri *wunɔ*, Baluchi *gwan* "pistachio", Parachi *γǎn* "oak" and Kati *wónzī*, Prasun *wőzu*, Pashai *wanǰ́*, Ningalami *banzí*, Wotapuri *banz* f., Kalasha *bōñǰ-*, Khowar *banǰ* "holly-oak", Western Pahari *bañǰ* "sp. tree", Hindi *bā:jh* m., *bañjī* f. "oak / Quercus incana" (Turner 1966, # 11209: **vañjha-*; Bailey 1979, 268-69).

Indo-Aryan

There are remarkable forms in the modern Indo-Aryan languages from Northwest: West Pahari *phāgu* "fig tree", Panjabi, Phalura, Shina *phāg*, Sawi *phåg*, Wotapuri *phāu* id. < **phāgu-*; Panjabi *phagvāṛā* m. "Ficus cariocoides", *phagvāṛī*, *phagūṛā* f. "its fruits", West Pahari *phɛgõ* "fig". The unexpected *ph-* could be caused by the influence of Old Indic *phála-* n. "fruit", cf. Old Indic (lex.) *phalgu-* "Ficus oppositifolia" (Turner 1966, #9063; Buddruss 1967, 120) with *-l-* indicating the contamination *phála-* x **bhāgu-*.

Baltic

Karulis (I, 304) sees a Baltic cognate in Lithuanian *gúoba & guobà* "Ulme, Rüster / Ulmus campestris", *gúobas* "weisser, heller Baum, der Ulme nahestehende Gattung" (Leipalingis), *guobýnas* "Ulmenhain, mit Ulmen bewachsene Hain", Latvian *guôba* "Ulme, Rüster / Ulmus campestris", "Hainbuche, Hornbaum / Carpinus betulus", *guôbājs*, *guõbiens* (*LKŽ* III, 736; Fraenkel 1962-65, 176-77; Mühlenbach & Endzelin I, 688). The Baltic tree-name is derivable from $*g^{(h)}\bar{o}b^{(h)}o$-/$\bar{a}$. The are two differences from the common root $*b^h\bar{a}g$-: (1) The opposite order of stops. (2) The difference $*\bar{o}$ vs. $*\bar{a}$. Ad (1): For the metathesis *P...K* > *K...P* there are more examples in Baltic: (i) Lithuanian *kèpti* : *kepù* "backen, braten", Latvian *cept* "backen, braten, sengen, brennen" vs. Slavic **pekt'i* : **pekǫ* "to bake" < $*pek^u$- (Pokorny 1959, 798; Fraenkel 1962-65, 241). (ii) Lithuanian *gerbti* : *gerbiù* "achten, hochschätzen, (ver)ehren" vs. Slavic **bergt'i* : **bergǫ* "to protect, look after" (Trubačev 1977, 10). Ad (2): The Lithuanian-Latvian correspondence in the root vocalism indicates IE $*\bar{o}$ instead of expected $*\bar{a}$ > Lithuanian *ó* [$\bar{o}$] ~ Latvian *ã/â* [$\bar{a}$] (Stang 1966, 37-38). The following three solutions might solve this question: A) The specific dialectal development, concretely IE / proto-Baltic $*\bar{a}$ > Žemaitic *uo*; Latgalian & Southeast Vidzeme *uo* (Stang 1966, 37-38). B) In the standard model of Indo-European apophony $*\bar{a}$: $*\bar{o}$ are compatible with the help of the laryngeal theory, i.e. $*eH_2$: $*oH_2$. C) Irregular development of the root vocalism under the influence, e.g., of the verb of the type *guõbti* "aushöhlen".

Slavic

In Slavic a cognate has been sought in **bъzъ*, **buzъ*, **byzьje* "sambucus; syringa", indicating $*b^{(h)}u\hat{g}^{(h)}$-, $*b^{(h)}a/ou\hat{g}^{(h)}$-, $*b^{(h)}\bar{u}\hat{g}^{(h)}$-, in spite of the difference in the root vowel and in semantics. The third argument against the 'beech-elder' etymology follows from the fact that the Iranian (Gilani *faγ*) and Albanian data confirm the velar character of **-g-*. Summing up, it seems best to exclude the 'elder'-name from our consideration of the 'beech'-etymology. For Slavic **bъzъ /*buzъ /*byzьje* there is a promising cognate in Lithuanian *búožė* "reed-mace / Typha latifolia" = "švendras, vilkauodegė" (*LKŽ* 1, 1169), cf. also the river-name *Búož-upis* (Vanagas 1981, 74) and Latvian *bouze* or *buožu kuoks* "gekappter Baum im Walde". On the other hand, it is legitimate to ask, if the

homonymous roots—pre-Slavic **b$^{(h)}$uĝ$^{(h)}$-* "elder", and IE **b^{h}uĝ(o)-* "he-goat" (> Gypsy *buzni* "goat"; Avestan *būza-*, Persian *buz*, Ossetic *boz* "he-goat", Armenian *buc* "lamb", pre-Celtic **bugno-* > **bukko-* "he-goat" > Old Irish *bocc*, Welsh *bwch*, Cornish *boch* "caper", Breton *bouc'h* "bouc", pre-Germanic **bukna-* > Old Icelandic *bukkr*, *bokki*, Old English *bucca*, Old High German *boc,-ckes* "buck, he-goat" - see Abaev I, 264; Adams & Mallory, *EIEC* 229; Pokorny 1959, 174; Stokes 1894, 179-80)—are related? The most natural semantic bridge between the meanings "elder" and "(he-)goat" is the fact that "elder" is classified as *Loniceraceae* and there are more representatives of this subfamily with a name motivated by "(he-) goat": Greek αἰγίνη (Ps.-Diosc. 4.14); Medieval Latin *caprifolium*, German *Geissblatt*, and further Russian *žímolosť*, Ukrainian *žýmolosť*, Byelorussian *žýmoloć*, Polish *zimołza* "woodbine / Lonicera", etc. < Slavic **žimolztъ* & **žim̥lza* < **g^{h}aid-* "goat" & **melĝ-* "to milk" (Trubačev 1960, 84 and Id. apud Vasmer II, 55-56), and outside Indo-European, e.g., Nogai *eški tal* "sambucus", lit. "goat's willow" (Dmitrieva 1972, 181). The 'goat'-etymology allows us to include here Latin *sambūcus* (Lucil.) ~ *sabuncus* (gloss) "elder", if it reflects Gaulish **sam-buKnos* < **sm̥-b^{h}uĝno-*, cf. Celtic **bukko-* "he-goat" < **bugno-* (**sm̥-* is a prefix with a collective function which can be identified in other tree-names too).

The following appelatives designating various parts of chariot seem more promising:

**baga* > Serbo-Croatian dial. *b`àga* "Wagnergerät, Teil am Wagen", Macedonian dial. *bága* "part of a chariot"; cf. also Russian *bagán* "long thin pole; fork, wooden hook of a plough".

**bag(ъ)rъ* > Slovak *bahor* "wooden bent part of a wheel", Ukrainian dial. (Boikovian) *báh(o)r* "hinge of the wheel tire", Russian *bagór* "long wooden pole with the iron extension and hook" and **bagro* > Slovak dial. *bahro* "part of a wheel", *bahra* f. "wheel tire", Ukrainian *bahró* "hinge; part forming the wheel tire", Russian dial. *bagró* "metallic extension of *bagór*" (Berneker 1908, 38; Sadnik & Aitzetmüller 1963-73, 108; *SP* I, 179; Trubačev 1, 132-33, 130; Vasmer I, 101-02). Trubačev (l.c.) accepts Ondruš' etymology identifying here the derivative of the root **b^{h}eg^{u}-* "to run". The apophonic grade **b^{h}ōg^{u}-*, unattested in Slavic, is known e.g. from Lithuanian *bog□nti* "etwas Schweres eilig forttragen, fortschleppen" : *bĕgti* "laufen, rennen". The semantic motivation "wheel" = "runner" is quite natural, cf. Greek τροχός "wheel" : τρέχω "I run, go" or Latin *rota* "wheel" : Old Irish *retaid* "runs" (Huld 2000, 99-101), but the primary meaning of the quoted Slavic words was "wooden part of a chariot". In this case it is quite natural to expect the motivation of the designation to be based on the material, namely "wood", cf. Mycenaean *do-we-jo* = *dorwejos* (Aura Jorro 1985, 194) determining the wooden parts of a chariot (Gamkrelidze & Ivanov 1984, 730). In the *Iliad* (V, 838) the wood from the **b^{h}āg-*tree ("oak" in Greek) serves as a material for the axle: φήγινος ἄξων, cf. also the Latin adaptation *fāginus axis* (Verg. *Georg.*III, 172). For both alternative suffixes (i) **-rъ* and (ii) **-ъrъ* there are functionally close parallels in Slavic: (i) **dǫbrъ* "oak forest" : **dǫbъ* "oak", **grabrъ* "hornbeam" : **grabъ* id., **bobrъ* "bean" : **bobъ* id., cf. also Lithuanian *stuobras* "trunk, stem" : *stúobas* id. (Sławski 1976, 17-18); (ii) Slavic **stъbъrъ* "trunk, stem; column, post, pole" = Lithuanian *stùburas* "column; spine, backbone" = Latvian *stuburs* "column, pole" : *stubs* "trunk, stem", Old English *stybb* "trunk, stem" < **stubja-* (Otrębski 1965, 151; Sławski 1976, 27).

Armenian

Armenian *p'ekon* "beech" is apparently borrowed from Greek (Mann 1963, 11). As a continuation of the base **b^{h}āg(ó)-* one would expect Armenian **bak*. Such a word really exists - it means "stick, pole" (Džaukjan 1967, 102, fn. 37). The semantic shift may be confirmed by the following parallels: (i) Russian *dubína* "stick, club, cudgel" : *dub* "oak"; (ii) Russian *trost'* "stick" : Ossetic *tærs / tærsæ* "beech / Fagus silvatica ~ orientalis" < **tr̥šta-*, originally "hard", cf. Old Indic *tr̥ṣṭá-* "rough, harsh; hoarse" (Abaev 1979, 272-73).

Besides the tree-names and wooden objects there are two Indo-European languages where the base **b^{h}āg-* (or **b^{h}ag-* !) can be identified only in divine names, viz. Phrygian and Lydian. Ramat (1963, 50) mentions the relation of **perk^{u}u-* "oak" as a sacred tree and **Perk^{u}uno-* "oak- & storm-god" (see Pokorny 1959, 822-23). In our case one should suppose an opposite development: **b^{h}ago-* "god (as a distributor of happiness)" → **b^{h}āg(ó)-* "divine tree".

Phrygian

Beginning with Torp (1895, 193-94) a Phrygian cognate has been sought in the epithet of Zeus: Βαγαῖος· Ζεὺς Φρύγιος (Hesych.), cf. other divine epithets motivated by the root **b^{h}āg-* as Ζεὺς Φηγωναῖος after Steph. Byz. and Latin *Iuppiter Fāgutālis*. This idea was supported e.g. by Gusmani (1958, 853). On the other hand, there are also alternative solutions: (a) The emendation Βα<λ>αῖος based on one inscription from Bithynia-Paphlagonia where we read ΔΙΙ ΒΑΛΗΩ ΠΟΠΑΙΟΣ ΑΝΤΩΝΙΟΣ ΑΡΕΣΤΟΣ, and Hesychius' gloss βαλήν· βασιλεύς. Φρυγιστί (Schmitt, 1963, 44-46); (b) The Iranian origin of Βαγαῖος with respect to Old Persian *baga-* "god", *bagāya-* "divine" (Neroznak 1978, 137-38 with older references). But the witness of the Old Phrygian inscriptions speaks for the native origin of the form with *-g-*. In the dedicative inscription G-136 (Gordion, 6th cent.?) *tadoy iman bagun* the last word has been interpreted as the acc. sg. ntr. in **-ŏn* (Lejeune 1979, 224) meaning "deity, image" or "happiness" (Neroznak 1978, 104; Bajun & Orel 1988, 194 respectively).

Anatolian

Phrygian is not the only language of ancient Anatolia where the base *BaK°* is used in theonymy. Arkwright (1918, 62) compared Phrygian Βαγαῖος with Lydian *Baki-* "Bakchus" (see Gusmani 1964, 74-75). With regard to the absence of any etymology of Greek Βάκχος "Dionysos", it is legitimate to ask if this theonym is not borrowed from Asia Minor (Nilsson I, 578). The relation of Dionysos to the world of trees may be demonstrated by Dionysos' epithets δενδρίτης "zum Baum gehörig" (Plut. *Quaest. conviv.* V, 3.1) or ἐνδενδρος· παρὰ Ῥοδίοις Ζεύς· καὶ Διόνυσος ἐν Βοιωτίᾳ. (Hesych.).

On the other hand, some scholars seek a continuant of the IE "beech" in Lydian μυσός "beech", deriving it from **b^{h}ugo*. The "beech" is and was known in Asia Minor, but only in the northern coastal zone, not inland, including Lydia. The most natural solution of this discrepancy was proposed by Lane (1967, 210-11), viz. μυσός = "the Mysian (tree)", cf. the typological parallels in two Finnish designations for "beech": *saksan tammi* "Saxon (= German) oak", *saksan saarni* "Saxon ash-tree" (Feist 1913, 495).

Etymology

There have been several etymological attempts to interpret the Indo-European dendronym "beech".

1. **b^{h}āg(ó)-* from **b^{h}ag-* "zuteilen; als Anteil, als Portion erhalten" → "essen" (Pokorny 1959, 107; Schirmer & Kümmel, *LIV* 65):

1.1. The idea of the 'beech' as a tree with edible nuts was first proposed by Apion (1st cent. AD) and kept by many scholars in the 19th cent. and later (e.g. Kuhn 1855, 84; Hirt 1892, 483). Support for this identification can be sought in Old English *mæst* "mast, beechnuts, acorns", Old & Middle High German *mast* "Futter, Mast, Mästung, Eichelmast" < West Germanic **masta*; Old Irish *mess* "glandée, pâture", Welsh *mes* "glandée" etc. < **med-tu-*, vs. Gothic *mats* "food", *matjan* "to eat" etc. (Torp & Falk 1909, 318, 305; *LÉIA*, M-43; Mann 1984-87, 724, 761-62; probably derivable from **med-* "voll werden, satt werden" reconstructed by Zehnder, *LIV* 423-24). It is tempting also to add Greek μόστηνα κάρυα (Ath. II, 52b) and Spanish *mesto* "Zerreiche" with Basque *ametz* "Steineiche / Quercus robur; Hainbuche / Carpinus betulus" (Meyer-Lübke 1935, #420; Löpelmann 1968, 49-50), although the question of the donor-language remains open. A semantically less supported idea was proposed by Leumann (1930, 190) who speculated about the interpretation "Losbaum", cf. Avestan *baγa-* "Anteil, Los".

1.2. Ramat (1963, 49-51) concluded that the tree-name **b^{h}āgós* meant "belonging to **b^{h}ago-*", i.e. "consecrated to **b^{h}ago-*".

2. Krogmann (1955, 19f) speculated about the derivation from the root **b^{h}ā-/*b^{h}ō-/*b^{h}ə̄-* "glänzen, leuchten, scheinen" (Pokorny 1959, 104-05; Schirmer, *LIV* 68-69: **b^{h}eH$_2$-*). The proposed derivational pattern is in principle possible, although the suffix **-go-/*-gā-*, extending the bare verbal root, is not too productive, but cf. Lithuanian *eigà* "Gang" : *eĩti* "gehen", Latvian *nirga* "Taucherente" : *nirt* "tauchen", and Old Church Slavonic *struga* "Strömung" : Lithuanian *sraviù*, *sravěti* "gelinde fliessen" (Brugmann 1906, 507). Concerning semantics, there is a parallel e.g. in German *Weissbuche*, Turkish *ak gürgen* "beech", lit. "white hornbeam" (Dmitrieva 1972, 181; Räsänen 1969, 151 has recorded *gürgän ~ gülgän* "eine Buchenart / Chadara tenax).

3. Starostin (1988, 124) assumed a borrowing of IE **b^{h}āg(ó)-* from a substratal source of the type proto-North Caucasian **p̄ŏlnqqwe* "oak; wood", reconstructed on the basis of West Caucasian **pə̄q^{w}a ~ *p(p)ə̄x́a* > Adyghe, Kabardin *pχa* "wood", Ubykh *məx́ā-ç$_w$* "spoon", Abkhaz *a-mħáç$_w$* id.; East Caucasian **mŏlqqwe* "oak, acorn" > Avar *mikk*, Chamalal *nik^w*, Akusha *mig*, Kubachi *mik^w*, Tabasaran *maql^w*, Tsakhur *moql* etc. Later the West Caucasian material was separated and connected with the East Caucasian etymon "birch", all from North Caucasian **mħĕrqwĕ* (*NCED* 810), while the reconstruction of the East Caucasian "oak" was given with more precision to **mħŏqwe* (*NCED* 811-12). With respect to these new reconstructions, it is difficult to seek any relation between IE **b^{h}āg-* and North Caucasian **mħŏqwe* "oak". The North Caucasian origin is very probable for the isolated designation of "oak" in Old Georgian *muqa*, Georgian *muxa* (Fähnrich 1988, 35). On the other hand, a substratal language related to North Caucasian could be a source of a pre-Romance **mūgus* "Zwergtanne" preserved in Trientinian *mugo*, Friulian *muge*, Puschlav (Graubünden) *müf*, Bormio (Lombardia) *muf* (Meyer-Lübke 1935, #5721).

4. Until the present time nobody paid attention to possible external relations. There are promising parallels in 3 or 4 branches of the Afroasiatic (macro)family:
Semitic **baky-* > Hebrew *bākā* "Pistacia lentiscus", *b^{e}kāyyīm* pl. "sorte d'arbre", Arabic *baka^{n}* "arbuste épineux", maybe also *baka^{n}* "plante ligneuse comme le *bašam* ["arbuste odoriférant"] qui, lorsqu'il est coupé, laisse couler une sève laiteuse" (Cohen 1976, 64).
Egyptian (New Kingdom) *bkj* "Art Baum in Syrien (Terebinthe?)", maybe also *bkj* "eine grössere Frucht" (Erman & Grapow I, 482), although the West Semitic origin cannot be excluded.
Berber: Ayr *abăka* "jujube, jujubier" (Alojaly 1980, 5), Ahaggar *ăbăka* id. < **ā-bakāH* (Prasse 1974, 228-29).
Chadic (East): Bidiya *bàgú* "fig-tree" (Alio), Migama *bóo* "fig-sycamore", Jegu *buugaye* "fig-tree" (Jungraithmayr).

In Dravidian there are at least three promising candidates, all trees with edible fruits or nuts:
**pākk-u* > Tamil *pākku* "areca nut, areca palm", *pāku* "areca nut", Malayalam *pākku* "a raw areca nut", Tulu *pākuṭṭi* "knife for cutting betel nuts", Telugu *põ:ka* "areca nut; areca tree / Areca catechu", Kuwi *pōka mrānu* "areca tree" (*DEDR* # 4048);
**pāk-al* > Tamil *pākal / pāval* "balsam pear / Momordica charantia", Malayalam *pāval* id., Kannada *hāgal(a)* id. (*DEDR* # 4045);
**pakk-ay* > Kannada *pakke* "tamarisk tree / Tamarix indica", Telugu *pakke*, *pakkiya* "tamarisk tree / Tamarix gallica ~ indica" (*DEDR* # 3812).

In Altaic promising cognates occur in one of the Mongolian languages, namely Kalmyk *buγ$^{\mu}$r* "Buche, Buchenholz" (Ramstedt 1935, 58)—while Written Mongolian *eberling modun* and Khalkha *everleg mod* reflect the metaphor "horny tree" (Dmitrieva 1972, 182)—and in one of the Turkic languages, namely Chulym-Turkic *pāγu* "forest in a swampy country consisting of no great pines" (Rassadin 1971, 162).

Conclusion

Friedrich & Mallory (*EIEC* 58-60) summarize the discussion about "beech" in the following assumptions:
1. That **b^{h}āgó-* actually meant "beech" rather than any of its other reflexes;
2. That **b^{h}āgó-* could be attributed to proto-Indo-European antiquity rather than a later dialectal status;
3. That the assumed distribution of the "beech" in prehistory was correct.
The facts analyzed here allow the following interpretation. Let us start with the third question: **Ad 3.** The best answer was given by Friedrich & Mallory (*EIEC*, 59-60):

> "The spread of the beech after the end of the last Ice Age can be traced across Europe where the initial finds are confined to southern and central Europe. By *c* 6000 BC the beech was largely confined to northern Greece, the Balkans and the Alpine region with expansions westward into northern Italy and towards south and central France. by *c* 4000 BC the beech may be found as far north as southern Germany and Romania. By 3000 BC, the beech would have penetrated further north into southern Poland and by 2000 BC the beech would have reached the Baltic Sea and northern France. Despite many claims to the contrary, this temporally dynamic spread of the beech offers little comfort to any putative solution to the IE homeland problem. In the area of Asia where we believe the beech was quite native, i.e. Anatolia to northern Iran, it is

linguistically unattested[1]. Its heartland in Europe was largely confined to Greece and the Balkans, the very territories that provide the meaning 'oak' rather than 'beech'. In those regions where IE stocks do attest the meaning 'beech', the tree itself seldom appears earlier than the Bronze Age, i.e., clearly after the period of PIE disintegration / IE expansions. Its sensational spread north and westwards from its original core area during the Bronze Age (*c* 2500-1000 BC) may correspond roughly to the expansion of some IE peoples into western Europe as the vast primeval forests of beech and oak had been established in Gaul and Germania but the concatenation of assumptions required to press the "beech line" into an argument concerning the earlier location of the Indo-Europeans would appear to be exceedingly dubious."

Note 1) The situation is not so hopeless. Besides Modern Turkish *ak gürgen* "beech", lit. "white hornbeam" (Dmitrieva 1972, 181) we know the Lydian designation of the "beech", viz. μυσός, μῦσος, lit. probably "a Mysian [tree]" (see above), Armenian *gadčaradzar*, *gadčari*, *gadži* "beech" (Schrader & Nehring 1917-23, 171), and several Iranian terms: Ossetic *tærs(æ)*, lit. "hard" (see above), Modern Persian *ălăš*, Talysh *ălăš*, Gilani *răš*, Kurdic *dara reš*, all from Iranian **raxša-* "red" (> Persian *raḫš* "mixed red and white, between black and fusc", Armenian lw. *erašḫ* "rötlich", besides Kurdic *raš* "black", Khotanese *rrāṣa* "dark-colored", Wakhi *rakṣ* grey, brown"; Iranian > Mari *rakš(ə)* "braun": Old Indic *rāga-* "redness" - see Bailey 1979, 362; Joki 1973, 306), cf. German *Rotbuche* and Turkic idiom from Kagānrūd *qizilaγaǰ*, lit. "red tree" (Henning 1963, 68, fn. 1), further Modern Persian *čeker*, Nūr *čil(h)ar* and Mazenderani *mirs* (Henning 1963, 68). The "beech" is also well-known in the aboriginal languages of the Caucasus: Kartvelian **ċipl-* or **ċip-* "beech" > Georgian *ċipeli*, Mingrelian & Laz *ċipuri*, Swan *ċipra* (Klimov 1964, 244; Fähnrich & Sardshweladse 1995, 503 respectively);
East Caucasian **pĭrpi* "beech" > Chechen-Ingush *pop*; Andi *pipi*; Ginukh *pepi*, Bezhta *pipe* id., Gunzib *pibe-s* "plane-tree", Khvarshin *pepe* "cudgel"; Akusha, Urakhi *purpi*, Kubachi *pūpe* "beech"; Lezgin *pipi-n ttar*, Tabasaran *pirpu-n har*, Budukh *pip*, Udin *pup*, *pup-na χod* id., Kryz *pip* "id., poplar" (*NCED* 872-73);
West Caucasian **t-ǵʷə* "beech" > Kabardin *txʷey*, Adyghe *tfăyə*, Abkhaz *á-šʷ*, dial. Bzyb *á-śʷ* id., Abaza *šʷ-čʷə* "plane-tree"; cf. the related tree-names in Nakh *pχā* "hornbeam" > Ingush *pχän* id., Chechen *pχa* "id., plane-tree"; Avar-Andi **λ(ʷ)VrV* > Avar *λor* "hornbeam", Tindal *λiri-ǟ:a roha* "asp"; Lak *x:iri* "lime-tree" (*NCED* 1064: **xwīrʕi*);
West Caucasian **pč:V* > Kabardinian *bžey* "plane-tree", Adyghe *pčăyə* id., dial. Shapsug *p:č:əyă* "beech"; Abkhaz *á-ʒ́* "oak", Abaza *ʒ́-čʷə* id., probably related to Nakh **nač* "oak" > Chechen *naž*, Ingush *naž*, dial. *nač:* id. (*NCED* 857: **nV̆čē*).

Ad 1. The common semantic denominator for the Indo-European dendronym **bʰāg(ó)-* may be determined as a "tree with edible fruits / nuts and hard wood, of the family *Fagaceae*", continuing as "oak" in Greek, "chestnut-oak" in Albanian, and "beech" in Italic and Germanic; probably also Baltic and Iranian, where the substitution "beech" → "hornbeam" is more probable than "oak" → "hornbeam", taking into account a common occurrence and the similar grey bark of both the "beech" and "hornbeam". Regarding the rich "oak"-terminology in Continental Celtic, cf. **derua* > Old French *dervée* "oak-forest", **d(e)rullia* > French *drouille*, *drille* "oak", **kassano-* > Old French *chasne*, French *chêne* "oak", South French *kasañú* "young oak-tree", Sicilian *cássinu* id., Piemontic (Cuneo) *kasna*, Catalonian (Ribagorça) *kase* "oak", **kassīko-* > Gallego *caxigo*, Asturian *caxigu*, Spanish *quejigo* "oak" (Meyer-Lübke 1935, ##2585b, 2778a, 1740;

Wolf 1997, 1013-32), it is quite natural to expect a different meaning of the Continental Celtic **bāg(o)-*, and "beech" is a promising pretender. Summing up, the meaning "beech" seems the most probable.

Ad 2. The occurrence of the **b^{h}āg(ó)-*dendronym and its derivatives in Italic, Germanic, Albanian, Greek and Iranian, probably also in Celtic, Baltic and Slavic, implies that this tree was among the most widespread tree-names in the Indo-European languages. The theonym Βάκχος with its Lydian (and Phrygian ?) parallels could indicate the presence of the **b^{h}āg(ó)-*tree in ancient Anatolia too, especially with respect to the epithets δενδρίτης and ἐνδενδρος of the synonymous theonym Dionysus. There is even an external witness of a deep antiquity in the arboreal terms outside Indo-European: Afroasiatic **baky/H-* and Dravidian **pākk-*pāk-/*pakk-*. Together with Indo-European **b^{h}āg(ó)-* they designate various trees with edible fruits. Regarding the regular correspondences in agreement with the Nostratic historical phonology, these forms represent a common heritage rather than mutual borrowings. On the other hand, the external parallels do not allow us to determine the IE laryngeal (if we do not take into account the glides in the position of the third radical in Afroasiatic). The evidence could indicate the originally unapophonic **-a-*, inherited from the older stage of the Indo-European protolanguage (cf. Gamkrelidze & Ivanov 1984, 161).

Summing up, **b^{h}āg(ó)-* ± "beech" belongs to the most archaic part of the Indo-European lexicon. Accepting this (and there is both internal and external evidence for it), this tree-name was used in the territory occupied by Indo-Europeans before their dispersal. The exact determination of a homeland is impossible; the described spread of the "beech" indicates that we must take in account the Balkans, north Asia Minor, the Caucasus and north(east) Pontic area, all the regions around the Black Sea (let us mention that in the Boreal this sea was significantly smaller). Paradoxically, the early dispersal of the "beech" excludes the area westward and northward to the Alps and Carpathians, contrary to attempts to localize the Indo-European homeland (e.g. Hirt, Thieme).

References

Abaev, Vasilij I. 1979. *Istoriko-ėtimologičeskij slovaŕ osetinskogo jazyka*, III. Leningrad: Nauka.

Agud, Manuel & Tovar, Antonio. 1991. *Diccionario etimológico vasco*, III. San Sebastian: Diputación Foral de Gipuzkoa.

Alojaly, Ghoubeïd. 1980. *Lexique touareg-français*. Copenhague: Akademisk Forlag.

Arkwright, Walther G. 1918. Lycian and Phrygian Names. *Journal of Hellenic Studies* 38, 45-73.

Aura Jorro, Francisco. 1985-93. *Diccionario micénico*, I-II. Madrid: Consejo superior de investigaciones científicas.

Bailey, Harold W. 1936[81]. Persia II: Language and dialects. In: *Encyclopaedia of Islam* III, 1050-58, reprinted H.W. Bailey's *Opera Minora. Articles on Iranian Studies*, Vol. I, ed. by M. Nawabi. Shiraz: Forozangah Publishers 1981, 299-307.

Bailey, Harold W. 1979. *Dictionary of Khotan Saka*. Cambridge: University Press.

Bajun, Lilija S. & Orel, Vladimir É. 1988. Jazyk frigijskix nadpisej kak istoričeskij istočnik. *Vestnik drevnej istorii* 1988/1, 173-200.

Bartholomae, Christian. 1898. *Arica* X: 56. Kurd.*būz* - lat. *fāgus*. *IF* 9, 252-283, espec. 271-272.

Berneker, Erich. 1908. *Slavisches etymologisches Wörterbuch*. Heidelberg: Winter.

Bertoldi, V. 1931. Casi di sincope nel galico e nel gallo-ligure. *Revue celtique* 48, 281-292.

Billy, Pierre-Henri. 1993. *Thesaurus Linguae Gallicae*. Hildesheim - Zürich - New York: Olms & Weidmann.

Brandenstein, Wilhelm. 1936. *Die erste "indogermanische" Wanderung*. Wien: Gerold.

Brandenstein, Wilhelm. 1962[68]. Das Indogermanenproblem. *Forschungen und Fortschritte* 36, 45-49, reprinted in Scherer 1968, 523-537.

Brugmann, Karl. 1906. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*$_2$, II.1. Strassburg: Trübner.
Buck, Carl D. 1949. *A dictionary of selected synonyms in the principal Indo-European languages*. Chicago: University Press.
Buddruss, Georg. 1967. *Die Sprache von Sau in Ostafghanistan*. München: Kitzinger.
Būga, K. 1922. *Kalba ir senovė*, I. Kaunas: Švietimo Ministerijos leidinys; quoted after *Rinktiniai raštai*, II, ed. Z. Zinkevičius. Vilnius: Valstybinė 1959.
Campanile, Enrico. 1974. *Profilo etimologico del cornico antico*. Pisa: Pacini.
Chantraine, Pierre. 1968-80. *Dictionnaire étymologique de la langue Grecque*. Paris: Klincksieck.
Claudii Ptolemaei *Geographia*, ed. C.F.A. Nobbe. Hildesheim: Olms Verlagsbuchhandlung.
Cohen, David. 1976. *Dictionnaire des racines sémitiques*, II. Paris-La Haye: Mouton.
DEDR = *A Dravidian Etymological Dictionary*$_2$ by Thomas Burrow & Murray B. Emeneau. Oxford: Clarendon Press.
Demiraj, Bardhyl. 1997. *Albanische Etymologien (Untersuchungen zum albanischen Erbwortschatz)*. Amsterdam - Atlanta: Rodopi.
Dmitrieva, L.V. 1972. Nazvanija rastenij v tjurkskix i drugix altajskix jazykax. In: *Očerki sravnitel'noj leksikologii altajskix jazykov*. Leningrad: Nauka, 151-223.
Dolgopolsky, Aron. 1988. The Indo-European homeland and lexical contacts of Proto-Indo-European with other languages. *Mediterranean Language Review* 3, 7-31.
Džaukjan, Gevork B. 1967. *Očerki po istorii dopiśmennogo perioda armjanskogo jazyka*. Erevan: Izdatel'stvo Akademii nauk Armjanskoj SSR.
EIEC = *Encyclopedia of Indo-European Culture*, eds. James P. Mallory & Douglas Q. Adams. London - Chicago: Fitzroy Dearborn Publishers.
Eilers, W. & Mayrhofer, Manfred. 1962. Kurdisch *būz* und die indogermanische „Buchen"-Sippe. Zugleich ein Beitrag zur Ulme und algemeinen Pflanzennamenkunde. *Mitteilungen der Anthropologischen Gesellschaft in Wien* 92, 61-92.
Erman, Adolf & Grapow, Hermann. 1971. *Wörterbuch der aegyptischen Sprache*, I. Berlin: Akademie-Verlag.
Ernout, Antoine & Meillet, Antoine. 1932. *Dictionnaire étymologique de la langue latine*. Paris: Klincksieck.
ESJS = *Etymologický slovník jazyka staroslověnského*, 1-10, eds. Eva Havlová & Adolf Erhart. Praha: Academia 1989-2000.
ÉSUM = *Étymologičnyj slovnyk ukrajinskoj movy*, I (**A-G**). Kyjiv: Naukova dumka.
EWAhd = *Etymologisches Wörterbuch des Althochdeutschen*, Bd. II, eds. Albert C. Lloyd, Rosemarie Lühr, Otto Springer. Göttingen - Zürich: Vandenhoeck & Ruprecht 1998.
Fähnrich, Heinz. 1988. Der Anteil der nachisch-daghestanischen Sprachen an der Prägung des georgischen Wortschatzes. *Georgica* 11, 34-36.
Fähnrich, Heinz & Sardshweladse, Surab. 1995. *Etymologisches Wörterbuch der Kartwel-Sprachen*. Leiden - New York -Köln: Brill.
Feist, Sigmund. 1913. *Kultur, Ausbreitung und Herkunft der Indogermanen*. Berlin: Weidmann.
Fraenkel, Ernst. 1962-65. *Litauisches etymologisches Wörterbuch*, I-II. Göttingen: Vandhoeck & Ruprecht - Heidelberg: Winter.
Friedrich, Paul. 1970. *Proto-Indo-European Trees. The Arboreal System of a Prehistoric People*. Chicago - London: University of Chicago Press.
Gamkrelidze, Tamas V. & Ivanov, Vjačeslav V. 1984. *Indoevropejskij jazyk i indoevropejcy*, I-II. Tbilisi: Izdatel'stvo Tbilisskogo universiteta.
Gusmani, Roberto. 1958. Studi sull' antico frigio. La popolazione, le glosse frige presso gli antici. *Rendiconti dell' Istituto Lombardo di Scienze e Lettere, Classe di lettere e scienze morali e storiche* 92, 835-869.
Gusmani, Roberto. 1964. *Lydisches Wörterbuch*. Heidelberg: Winter.
Hamp, Eric P. 1973. Review of Friedrich 1970. *American Anthropologist* 75, 1093-96.
Henning, W.B. 1963. The Kurdish elm. *Asia Major* 10, 68-72.
Hesychii Alexandrini *Lexicon*, editionem minorem curavit Mauricius Schmidt$_2$. Jena: Dufft 1867.
Hirt, Herman. 1892. Die Urheimat der Indogermanen. *IF* 1, 464-485.
Hirt, Herman. 1905. *Die Indogermanen. Ihre Verbreitung, ihre Urheimat und ihre Kultur*, I-II. Strassburg: Trübner.
Holder, Alfred. 1896-1907. *Alt-celtischer Sprachschatz*, I-III. Leipzig: Teubner.

Hubschmied, Johann. 1933. *Bāgāko-, *Bāgon(o)- 'forêt de hêtres'. Étude de toponymie suisse. *Revue Celtique* 48, 254-271.
Huld, Martin. 1981. Albanian corrigenda to Friedrich's *proto-Indo-European trees*. *KZ* 95, 302-308.
Huld, Martin E. 1984. *Basic Albanian Etymologies*. Columbus: Slavica Publishers.
Huld, Martin E. 2000. Reinventing the Wheel: The Technology of Transport and Indo-European Expansions. In: *Proceedings of the Eleventh Annual UCLA Indo-European Conference* (LA, June 1999), eds. K.Jones-Bley, M.E. Huld, A. Della Volpe. Washington D.C.: Institute for the Study of Man (*Journal of Indo-European Studies* Monograph Series No. 35), 95-114.
Joki, Aulis. 1973. *Uralier und Indogermanen*. Helsinki: *Memoires de la Société Finno-Ougrienne* 151.
Karulis, Konstantinas. 1992. *Latviešu etimoloģijas vārdnīca*, 1-II. Riga: Avots.
Klimov, G.A. 1964. *Étimologičeskij slovaŕ kartveĺskix jazykov*. Moskva: Nauka.
Kluge, Fridrich & Seebold, Elmar. 1989. *Etymologisches Wörterbuch der deutschen Sprache*$_{22}$. Berlin-New York: Walter de Gruyter.
Krogmann, Willy. 1954-55. Das Buchenargument. *KZ* 72, 1-29 & 73, 1-25.
Kuhn, Adalbert. 1855. Die Sprachvergleichung und die Urgeschichte der indogermanischen Völker. *KZ* 4, 81-124.
La Piana, Marco. 1939. *Studi linguistici albanesi*, 1: *Prolegomeni allo studio della linguistica albanese*. Palermo: S. Pezzino.
Lane, George S. 1967. The Beech Argument: A Re-Evaluation of the Linguistic Evidence. *KZ* 81, 197-212.
Lehmann, Winfred P. 1986. *A Gothic Etymological Dictionary*. Leiden: Brill.
LÉIA = Lexique étymologique de l'irlandais ancien (lettre B, D, M), ed. Pierre-Yves Lambert. Dublin: Institute for Advanced Studies - Paris: CNRS 1980, 1996, 1960.
Lejeune, Michel. 1979. Regards sur les sonores i.e. en vieux phrygien. In: *Florigium Anatolicum. Mélanges offerts à Emmanuel Laroche*. Paris: Boccard, 219-224.
Leumann, Ernst. 1930. 'Suppletivwesen' im Nordarischen. *KZ* 57, 184-200.
Leumann, Manu & Hofmann, J.B. 1928. *Lateinische Grammatik*$_5$. München: Beck.
LIV = Lexikon der indogermanischen Verben. Die Wurzeln und ihre Primärstammbildungen$_2$. eds. Helmut Rix et al. Wiesbaden: Reichart.
LKŻ = Lietuvių kalbos žodynas, I$_2$, III, eds. J. Kruopas, J. Kabelka, K. Ulvydas, B. Vosylytė. Vilnius: Minties 1968 (I), Lietuvių kalbos ir literatūros institutas 1956 (III).
Löpelmann, Martin. 1968. *Etymologisches Wörterbuch des baskischen Sprache*. Berlin: Walter de Gruyter.
Mallory, James P. 1992. *In Search of the Indo-Europeans. Language, Archaeology and Myth*. London: Thames & Hudson (reprint of the 1st edition from 1989).
Mann, Stuart E. 1963. *Armenian and Indo-European (Historical Phonology)*. London: Luzac.
Mann, Stuart E. 1984-87. *An Indo-European Comparative Dictionary*. Hamburg: Buske.
Melchert, H.Craig. 1994. *Anatolian Historical Phonology*. Amsterdam-Atlanta: Rodopi.
Meyer-Lübke, W. 1935[92]. *Romanisches etymologisches Wörterbuch*$_6$. Heidelberg: Winter.
Morgenstierne, Georg. 1930. The Waṇetsi Dialect of Pashto. *Norsk Tidsskrift för Sprogvidenskap* 4, 156-175, quoted after *Irano-Dardica*. Wiesbaden: Reichert 1973, 187-201.
Mühlenbach, K. & Endzelin, J. (ed.). 1922-25. *Lettisch-deutsches Wörterbuch*, 1. Riga: Lettische Bildungsministerium.
NCED = A North Caucasian Etymological Dictionary, by Sergei L. Nikolayev & Sergei A. Starostin. Moscow: Asterisk 1995.
Neroznak, Vladimir P. 1978. *Paleobalkanskie jazyki*. Moskva: Nauka.
Nilsson, M.P. 1940-50. *Geschichte der griechische Religion*, I-II. München.
Orel, Vladimir. 1998. *Albanian Etymological Dictionary*. Leiden - Boston - Köln: Brill.
Otrębski, Jan. 1965. *Gramatyka języka litewskiego*, II: *Nauka o budowie wyrazów*. Warszawa: Państwowe wydawnictwo naukowe.
Paxalina, Tat'jana N. 1983. *Issledovanija po sravnitel'no-istoričeskoj fonetike pamirskix jazykov*. Moskva: Nauka.
PC = Ptolemäus Cosmographia. Das Weltbild der Antike. Stuttgart: Parkland 1990.
Pokorny, Julius. 1956. Zum keltischen Buchennamen. *Beiträge zur Namenforschung* 7, 279-281.
Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch*. Bern-München: Francke.
Prasse, Karl-G. 1974. *Manuel de grammaire touaregue* (tăhăggart), IV-V: *Nom*. Copenhague: Akademisk Forlag.

PRECA 2, 6 = *Paulys Real-Encyclopädie der classischen Altertumswissenschaft*, Bd. 2, 6, ed. Georg Wissowa. Stuttgart 1896, 1909.

Puhvel, Jaan. 1991. *Hittite Etymological Dictionary*, Vol. 3 (H). Berlin-New York: Mouton de Gruyter.

Ramat, Paolo. 1963. I problemi della radice indoeuropea **bhăg-*. *Annali di Istituto orientale di Napoli*, Sezione linguistica 5, 33-57.

Ramstedt, Gustav J. 1935. *Kalmückisches Wörterbuch.* Helsinki: Lexica Societatis Fenno-Ugricae III.

Räsänen, Martti. 1969. *Versuch eines etymologisches Wörterbuch der Türtsprachen.* Helsinki: Lexica Societatis Fenno-Ugricae XVII.1.

Rassadin, V.I. 1971. *Fonetika i leksika tofalarskogo jazyka.* Ulan-Ude: Burjatskoe knižnoe izdatel'stvo.

Sadnik, Linda & Aitzetmüller, Rudolf. 1963-73. *Vergleichendes Wörterbuch der slavischen Sprachen*, I. Wiesbaden: Harrassowitz.

Schmitt, Rüdiger. 1963. Der angebliche Phrygergott *Bagaios*. *Sprache* 9, 38-47.

Schrader, Otto & Nehring, Alfons. 1917-23 / 1929. *Reallexikon der indogermanischen Altertumskunde*, I-II. Berlin -Leipzig: Walter de Gruyter.

Sławski, Franciszek. 1976. Zarys słowotwórstwa prasłowiańskiego, II. In: *SP* II, 13-60.

SP = Słownik prasłowiański, I-II. Wrocław - Warszawa - Kraków - Gdańsk: Ossolineum 1974-76.

Stang, Christian S. 1966. *Vergleichende Grammatik der Baltischen Sprachen.* Oslo - Bergen - Tromsö: Universitetsforlaget.

Starostin, Sergei A. 1988. Indoevropejsko-severnokavkazskie izoglossy. In: *Drevnij Vostok: Étnokul'turnye svjazy.* Moskva: Nauka, 112-163.

Stokes, Whitley. 1894. *Urkeltischer Sprachschatz.* Göttingen: Vandenhoeck & Ruprecht.

Stolz, Fr. 1894. *Historische Grammatik der lateinischen Sprache*, Bd. I. Leipzig: Teubner.

Thieme, Paul. 1953[54]. *Die Heimat der indogermanischen Gemeinsprache.* Wiesbaden: Steiner (*Akademie der Wissenschaften und der Literatur, Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse*, jg. 1953, Nr. 11, 535-613.

Thurneysen, Rudolf. 1946. *A Grammar of Old Irish.* Dublin: Institute for Advanced Studies.

Torp, Alf. 1895. Βαγαῖος. *IF* 5, 193-94.

Torp, Alf & Falk, Hjalmar. 1909. *Wortschatz der Germanischen Spracheinheit.* Göttingen: Vandenhoeck & Ruprecht.

Trubačev, Oleg N. 1960. *Slavjanskie nazvanija domašnix životnyx.* Moskva: Nauka.

Trubačev, Oleg N. (ed.). 1974-99. *Étimologičeskij slovaŕ slavjanskix jazykov*, 1-26. Moskva: Nauka.

Trubačev, Oleg N. 1977. Slavjanskie i baltijskie ėtimologii. *Étimologija* 1975, 3-12.

Turner, Ralph L. 1966. *A Comparative Dictionary of the Indo-Aryan Languages.* London: Oxford University Press.

Udolph, Jürgen. 1979. *Studien zu slavischen Gewässernamen und Gewässerbezeichnungen.* Heidelberg: Winter.

Vanagas, Aleksandras P. 1981. *Lietuvių hidronimų etimologinis žodynas.* Vilnius: Mokslas.

Vasmer, Max. 1950-56. *Russisches etymologisches Wörterbuch.* Heidelberg: Winter; quoted after the Russian translation of O.N. Trubačev: *Étimologičeskij slovaŕ slavjanskix jazykov*, I-IV. Moskva: Progress 1986-87.

de Vries, Jan. 1962. *Altnordisches etymologisches Wörterbuch*$_2$. Leiden: Brill.

WLS = Wörterbuch der litauischen Schriftsprache, I (A-K), eds. Max Niedermann, Alfred Senn & Franz Brender. Heidelberg: Winter.

Wolf, Heinz Jürgen. 1997. Die gallische Eiche. *Zeitschrift für celtische Philologie* 49-50, 1013-32.

On the Origin of Affricates in Austric

La Vaughn H. Hayes

1. Introduction.

1.1. Background. In 1973, Paul K. Benedict rejected the existence of Austric and inclusion of Austroasiatic in Austro-Tai, the superphylum which he had created in 1966 to link the South East Asian language families of Austronesian, Kadai, and Miao-Yao.[1] He did so on grounds of insufficient lexical evidence, but the writer has since presented a massive quantity of lexical and phonological evidence (cf. Hayes 1992, 1997b, 1999, 2000a) which confirms that Austric is a viable taxonomic entity comprising the AA and AN languages, as Wilhelm Schmidt had first proposed in 1906.[2] Pending future confirmation, Austric may be assumed to be a subgroup of Austro-Tai on the basis of Benedict's classification of Austronesian.

In 1975, Benedict reconstructed a two-phoneme affricate series for Austro-Tai, which he enlarged in 1990 to six phonemes. If Austric does belong within Austro-Tai, then the question arises as to whether or not the AT affricates were continued in Austric. At first glance, that would seem to be the case, for Benedict set up in 1973 two provisional affricates for Proto-Austroasiatic and.proposed in 1990 that Proto-Austronesian possessed reflexes of all six of the AT affricates. Closer examination of the available comparative data reveals, however, that an entirely new hypothesis of affricate origin is needed for Austric and Austro-Tai.

1.2. Purpose and Objectives. The purpose of this presentation is to review pertinent AT and Austric reconstruction and examine relevant lexical, morphological, and phonological correspondence. Its objective is to determine whether or not the AT affricates were continued in Austric and what the origin of the corresponding Austric phonemes really was.

2. Review of Affricate Reconstruction.

2.1. Austro-Tai. In his 1975 reconstruction of PAT phonology, Benedict set up the affricates and nasal-affricate clusters shown in Table 1. Their primary AA and AN correspondents are also displayed, the former introduced by the writer, the latter taken from Benedict 1975:155, 168.

In 1990, Benedict added the palatal affricates */tś, dź, ntś, ndź/ and alveolo-palatal affricates */tš, dž, ntš, ndž/ to the AT phoneme inventory.[3] He also argued that all of these consonants and

---

1. Kadai is also known as Tai-Kadai, Miao-Yao as Hmong-Mien. In 1990, Benedict attempted further to include Japanese-Ryukyuan in Austro-Tai.

2. Abbreviations used here are AA (Austroasiatic), ACD (Austronesian Comparative Dictionary), AJ (Austro-Japanese), AK (Austro-Kadai), AN (Austronesian), AT (Austro-Tai), C (consonant), CF (composition form), KY (Khmu' Yuan), MK (Mon-Khmer), MP (Malayo-Polynesian), MUK (Mường Khến), NK (Nyah Kur), OK (Old Khmer), OM (Old Mon), PK (Proto-Katuic), PM (Proto-Mon), PMN (Proto-Mnong), PNB (Proto-North Bahnaric), PSB (Proto-South Bahnaric), PVM (Proto-Viet-Muong), PW (Proto-Waic), R (resonant), V (vowel), VN (Vietnamese).

3. Square brackets denote phonetic, slashes phonemic representation. In reconstructed forms, square brackets also denote uncertainty and parentheses optionality.

clusters are reconstructible at the PAN level, a position not endorsed by any Austronesianist to the writer's knowledge. AA */c, j, ñc, ñj/ appear to correspond respectively to these new phonemes and clusters on the basis of the few comparative examples found thus far (see appendix).

Table 1. The AT Affricates and their Austric Correspondents

Austro-Tai	Austroasiatic	Austronesian
*ts	*c	*s
*nts	*(ñ)c	*(ñ)s
*dz	*j	*z
*ndz	*(ñ)j	*(ñ)z

2.2. Austronesian. Otto Dempwolff did not reconstruct affricates for Proto-Austronesian, but the "palatals" */t′, d′, ńt′, ńd′/ which he established are pertinent to our discussion.[4] Isidore Dyen resymbolized these as */s, z, ńs, ńz/ (*ñ replaces *ń here), and these symbols have been adopted in mainstream AN historical linguistics, as represented by the work of Robert A. Blust, for example. Dyen also introduced the retroflex affricate */C/ and palatal stop */Z/ to account for certain irregular correspondences.

Although called palatals, the exact phonetic nature of */s, z/ (Dempwolff's */t′, d′/) has always been debatable. The preponderant modern reflex of */s/ is /s/, but some languages have /c, t, ts/; the most common reflex of */z/ is /d/, but some languages have /r, ts, dz/. In light of these facts, it is quite possible that Benedict was correct in proposing that PAN */s, z/ were really */ts, dz/. Nevertheless, */s, z/ will be used here in keeping with common practice in AN linguistics.

Dempwolff did not reconstruct the denti-alveolar sibilants */s, z/, but evidence from Formosan, one primary subgroup of Austronesian, indicates that these phonemes were present at the PAN and PFormosan levels, shifting to */h, D/, respectively, in Proto-Malayo-Polynesian, the other primary AN subgroup. Dyen sought to rectify this situation partially by creating PAN */S/, ostensibly a voiceless apical sibilant, though Otto Dahl's discussion of this proto-phoneme (Dahl 1973:32-5, 101) leaves its exact phonetic nature open to debate. In this study, */h, D/ will be used in the exemplary data in section 4 and the appendix per usual AN linguistic practice.[5]

2.3. Austroasiatic. Affricates are very rare, palatal stops very common in the modern AA languages. Accordingly, Heinz-Jürgen Pinnow reconstructed */c, j/, but no affricates for Proto-Austroasiatic, a position with which the writer concurs. Benedict tentatively proposed PAA */[ts], [dz]/ in 1973, but Gérard F. Diffloth (1977:48ff.) showed that */c, j/, respectively, can be reconstructed in their place, thus confirming that affricates did not exist in Proto-Austroasiatic..

The frequency of palatals in Austroasiatic is partially explained in Austric II (Hayes 1997b, also cf. 1997a). In this article, the writer shows that a number of phonemes and clusters was transformed into palatals, to include */t, k, q, x/ > */c/ and */d, g, G, γ, R/ > */j/, due to environmental conditioning and that some palatals are reflexes of final clusters composed of the final

---

4. Dempwolff's PAN palatals, */ḱ, g′/, which Dyen resymbolized as */c, j/ and Benedict interpreted as reflexes of AT clusters, such as */kr, gr/, respectively, are not discussed here because they appear to have a history distinct from that of the affricates, but in a few cases, it is clear that */j/ is a reflex of Austric */dz/, cf. e.g. PAN */baluj/ 'dove species'.

5. Note, however, that */S/ appears in those examples from the ACD where Blust uses */S/ apparently to denote a voiceless sibilant. Also note that */D/ (Dempwolff's /ḍ/) represents a voiced retroflex stop.

stops */t, d/ and the suffix */s/. But these processes do not account for all of the AA palatals; some predate those changes, others are more recent innovations. Nor do they completely explain the many irregular correspondences seen in the available Austric comparative data involving palatal stops and other phonemes. It was these anomalies which provided much of the impetus for and interest in conducting this study.

The discovery that some Austric affricates were actually clusters of final stops and a sibilant suffix raised the possibility that affricates might not be as ancient in Austro-Tai as Benedict had led us to believe. It also indicated that additional research was needed to determine how non-final Austric affricates and their palatal correlates had originated, which is our main task here. That research began with a closer look at sibilants in "The Austric Denti-alveolar Sibilants", published in the last volume of this journal (Hayes 2000a). In this article, the writer demonstrated that AA */s, z/ and AN */h, D/ are reflexes of Austric */s, z/, respectively, and contended that the AA/AN sibilant correspondence provides "irrefutable proof" of the existence of Austric. The usage of that phrase drew sharp criticism from certain quarters, and it may be that this censure has obscured the value of the discovery of the AA/AN sibilant correlation. If so, it is hoped that publication of this sequel to Hayes 2000a will do much to counter any negative influence arising from the criticism, for this paper demonstrates that the development of the sibilants is only one part of a larger evolutional pattern involving the Austric sibilants and affricates.

3. A New Hypothesis of Affricate Origin.

3.1. Phonological Considerations. The correspondence of AA */c, j/ to AT */ts, dz/ and AN */s, z/, respectively, shown in Table 1 and the exemplary data presented below indicates that if Austric possessed dental affricates, then they changed generally to palatal stops in Austroasiatic. This indication would also be valid in the case of the palatal and alveolo-palatal affricates, if they also existed. The phonetic ambiguity discussed in subsection 2.2 makes it unclear whether the affricates were continued or changed to palatal stops in Austronesian.

However, the AA denti-alveolar sibilants */s, z/ also correspond to the AT affricates and their AN correlates in a number of cases. The duality of this correspondence suggests phonemic splitting, as in Austric */ts, dz/ > AA */c, j/ and */s, z/, respectively. But AA */c, j/ also correspond to the AT sibilants and their AN correlates in other cases. Here, too, splitting is suggested, but on the AN side of the house, as in Austric */ts, dz/ > AN */s, z/ and */h, D/, respectively.

This irregular correspondence could be accounted for by positing identical splits in the AA and AN proto-languages or a single split in Austric prior to emergence of the dialects which would become Austroasiatic and Austronesian. However, this solution is less than satisfactory because the conditions under which those splits could have occurred are not obvious in either modern Austroasiatic or reconstructed Austronesian.

An alternate solution would be to set up two pairs of affricates, e.g. Austric */ts_1, dz_1/ and */ts_2, dz_2/—or */ts, dz/ and */tś, dź/ (or */tš, dž/) could be used— with the former shifting to AA */c, j/ and AN */s, z/, respectively, and the latter merging with AA */s, z/ and AN */h, D/, respectively. Similar solutions have been used in similar cases in AT and AN historical reconstruction, and this explanation would be acceptable, even though we would have no means of telling precisely which pair of Austric proto-phonemes corresponds to which pair of AA and AN reflexes.

However, a more economical solution suggests itself upon review of the phonology and morphological model reconstructed for Proto-Austroasiatic in Austric I (Hayes 1992:162ff., 165ff, also cf. Hayes 2000a). Since AA */s, z/ and AN */h, D/ are the regular reflexes of Austric */s, z/, respectively, one may simply infer that Austric */s, z/ became AA */s, z/ and AN */h, D/, respectively, and splitting of affricates or mutation of an affricate series did not occur, as speculated upon above. Instead, it seems more likely that Austric */s, z/ developed under certain conditions new allophones which evolved into AA */c, j/ and AN */s, z/, respectively.

In this case, the conditions under which that development could have occurred are easily visible in the existence of the nasal-oral clusters that have been reconstructed for Austro-Tai, Austroasiatic, and Austronesian. It is virtually assured that the cluster set included */ns, nz/, and given the fact that epenthesis is known to occur in such clusters in other languages, one may infer further that the changes */ns, nz/ > */n^{t}s, n^{d}z/ > */nts, ndz/ must have taken place. This epenthesis must also have taken place prior to the emergence of the AA and AN dialects, thus during the Austric era if not the AT or a preceding era.

3.2. Morphological Considerations. In Austric I, a morphological model was presented for Proto-Austroasiatic, which may also be valid for Proto-Austric (cf. Hayes 1992:167ff.). This model includes three affixal complexes, a prefixal complex with the canonic form (C)(V)(R), an infixal complex with the form (V)(R)C, and a suffixal complex (R)C(V)(C). R represents probably several ancient affixes which coalesced over time as */N/, a homorganic nasal.

Juxtaposition of */N/ to consonants in the affixal complexes created a set of nasal-oral clusters, some of which have already been discussed in the Austric series, and this set doubtlessly included */ns, nz/. Epenthetic reconfiguration of the latter to */nts, ndz/ as described in the previous subsection would not be unexpected in these particular cluster environments. Thus, the morphology of the ancient language, whether Proto-Austric or Proto-Austro-Tai, can be seen as playing a crucial role in setting up the conditions under which the affricates could develop, cf. e.g. AN */bahaq/ 'flood(ed)' and */basaq/ 'wet', where */N/ was apparently applied to a verbal form */basaq/, whence AN */bahaq/, to create an attributive */baNsaq/, whence AN */basaq/.[6]

3.3. The New Hypothesis. Comparison of the AA and AN data presented in section 4 indicates that a dental affricate series did not exist originally, but developed at some point in time primarily due to epenthesis in nasal-sibilant clusters. Prefixation of sibilant initials by stop affixes and suffixation of stop finals by sibilant affixes were secondary sources. The point in time at which these changes began to take place may have been during the AT stage or earlier, if Austric is an AT subgroup, but pending confirmation of that relationship, development of the dental affricates is shown in Table 2 as having occurred during the Austric stage.

Table 2. The Origin of the Austric Dental Affricates

Austric	**Austrosiatic**	**Austronesian**
*s	*s	*s > *h
*ns > *n^{t}s > *(n)ts	*(ñ)c	*(ñ)s
*z	*z	*z > *D
*nz > *n^{d}z > *(n)dz	*(ñ)j	*(ñ)z

6. Blust has replaced Dempwolff's PAN */bahaq/ and */basaq/ with PAN */baSaq/ 'flood(ed)' and PMP */besaq/ 'wet, wash', respectively.

The existence of the palatal and alveolo-palatal affricates proposed by Benedict cannot be verified at this time on the basis of the Austric comparative data available to the writer.

3.4. Subsequent Developments

3.4.1. Overview. In Table 3, phonological developments which can be posited as occurring subsequent to those shown in Table 2 are displayed.

Table 3. Evolution of the Austric Sibilants and Affricates

Austric	Austroasiatic	Austronesian
*s	*s > *s > *ns > *nt	*s > *h, *S
*z	*z > *z > *s > *nz > *nd	*z > *D
*ts	*c > *c > *ñc	*s
*dz	*j > *j > *ñj	*z > *z, *Z?
*nts	*ñc > *c > *ñ'c	*ñs
*ndz	*ñj > *j > *ñ'j	*ñz > *z, *Z?

3.4.2. Austroasiatic. The changes depicted in Table 3 took place in the old prefixal and infixal complex environments; they do not include those discussed in Austric II involving palatalization of spirants and stops and associated transformations, with the exception of the devoicing of */z/.

Although it is believed that affricates developed in all of the affixal complexes, it is presently unclear what became of their reflexes in the old suffixal complex environment. Nasal-oral clusters do not occur finally in the Austric daughter languages; hence, it is likely that final */nts, ndz/ were simply reduced to */c, j/ in Austroasiatic, */s, z/ in Austronesian, respectively (or perhaps more properly PAN */s, j/ since */z/ does not occur finally). However, there are some indications that at least in a few instances, final */nts, ndz/ > */ñ'c, ñ'j/ > */ñ'j/ > */y?/ in Austroasiatic.

The */N/ affix apparently remained in use during and after the PAA stage, and this usage contributed to creation of new nasal-oral clusters. Some of these were retained into the modern era, others simplified to palatal stops or nasals. As a result of the cluster recreation, two distinct sets of cluster reflexes may be distinguished. One set apparently differs little from the original cluster, cf. Biat /njiŋ/ 'to weigh', AN */ziŋziŋ/ 'balance something on hands'. The other may differ significantly, cf. Katu /ʔəʔyɑh/ 'blouse', AN */bazuh/ 'clothes/clothing'. The latter set is believed to reflect the earliest AA nasal-oral clusters, i.e. those inherited from Austric, the former clusters created at a later date.

It appears that the earliest AA nasal-stop clusters generally changed to prenasalized implosives, such as the */ñ'c, ñ'j/ depicted in Table 3. In the case of the latter, this analysis is based primarily on correlation of PAA */ñ'j/, with which */ñ'c/ appears to have merged at least partially, to modern /y, ʔy, ʔ, ñ, ʔñ/, which appear to reflect dissimilation and/or reduction of imploded */'j/ and/or */ñ'j/. But it is possible that the early AA nasal-stop clusters merely coalesced as prenasalized unit phonemes, with implosion coming later, if at all. Whatever the nature of this

change was, its timing and distribution cannot yet be determined precisely, in part because the same sort of development has apparently recurred several times at more recent dates in the AA subgroups.

3.4.3. Austronesian. Early AN developments are more obscure, and the only certain one is the shift of PAN */s, z/ to PMP */h, D/, respectively, as mentioned in subsections 2.2 and 3.1.

Nevertheless, it is quite possible as well as plausible that other early developments occurred and remain as yet undetected. Some of the post-Dempwoff phoneme reconstructions and the phonemic alternation reconstructed both by Dempwolff and his successors may conceal such developments. Dyen's */C/, the existence of which some Austronesianists question, probably has nothing to do with the Austric sibilants and affricates. */Z/, which Dyen used to account for alternation of */z/ with */d, D/ in certain synonomous doublets reconstructed by Dempwolff, may reflect the shifts of Austric */dz/ to */z > Z/ and Austric */ndz/ to */ñz > Z/, transformations quite similar to those seen in Austroasiatic. The frequent correlation of */d/ to */D/—note that these proto-phonemes also often alternate—may well reflect PAN */nz > nd > d/ in some cases while PAN */T/, a retroflex apical stop, may reflect the same */ns > nt/ shift seen in Austroasiatic. These matters need much further study.

3.5. Reconstruction Notes. Many of the exemplary data sets in the appendix contain reflexes which evidentiate within the same set more than one of the changes depicted in Table 3, and it is often difficult to account for all of this contrastive phonological development, even by using the usual qualifier of dialectal divergence. AN */[t]un[D]uk/ 'to bow, bend down/over', for example, reflects Austric */z/ > AN */z > D/. AA correspondents to this etymon reflect three different evolutionary paths of Austric */z/: 1) */z > z > s/ in Pacoh /sɨk/ 'butt, gore'; 2) */z > nz > ndz > dz > j/ in Katu /pəjuk/ 'lower head to butt'; and 3) */z > nz > nd/ in Sora /duj/ 'bend'. Additional examples can be cited from the writer's files, with the latter change also exhibited in Pacoh /diək/ 'bend backward (finger)', /dɤy?/ 'bent a little ways backward', /pidɤk/ 'bend backwards'. Also cf. Vietnamese /cuk/ 'bend down, point downward', which reflects */zok > [k]zok > cuk/ and coalescence of a prefix and a word initial, a common phenomenon in AA linguistic history.

Contrastive morphological development can also be detected. AN */Dekuh/ 'bent, to bow' probably does not reflect */zok/, the Austric root underlying the above-cited forms, but rather a variant of the same lexical base, */zVk/, underlying */zok/. Pacoh /dɤy?/ and /pidɤk/ probably derive from the same variant as AN */Dekuh/. Similarly, Pacoh /sɨk/ and Katu /pəjuk/ apparently descend from Austric */zok/, but their morphological history is different, the former reflecting the bare root, the latter affixed */Nzok/.

As a result of such developments, it is often necessary to reconstruct at the PAA level two or more variant proto-forms for certain etymologies.[7] In doing so, care has been taken to exclude later developments, such as loss or accretion of the nasal element in nasal-oral clusters. But it is also often difficult to determine the age of certain changes; hence, some error in reconstruction is unavoidable at the present state of our knowledge of AA linguistic history.

---

7. A rigorously reductionist approach would probably eliminate many of the markers of contrastive phonological and morphological development on which the writer has focussed in order to discover explanations for the origin and evolution of the Austric affricates. The basic problem here is that the writer's position requires the assumption that the AA languages retained an unusually high number of parallel lexical forms which differed minimally with respect to phonology, morphology, and semantics. In this view, even a single language might retain reflexes of such forms as

4. Exemplary Data.[8]

4.1. Regular Correspondences.

4.1.1. Reflexes of Austric *s. Only examples from the appendix are cited; for additional ones, see Hayes 2000a.

Austric	**Austroasiatic**	**Austronesian**
*s	*s	*s > *h, *S

Austroasiatic	**PAA**	**Austronesian**
PW *s?ʏm 'rotting'	*sa?om	AN *ha[r]um 'aroma'
Bahnar kəsaay? 'sprinkle'	*saqi	AN *baSaq 'flood(ed)'
Katu saak 'corpse'	*sa[?]ak	AN *Sawak 'waist'
Jeh saw mat 'wash (face)'	*saw	AN *SawSaw 'wash'

4.1.2. Reflexes of Austric *ts.

4.1.2.1. Initial and Medial Position.

*ts	*c	*s
Jeh cayh 'to sneeze'	*caqi	AN *kesaq 'breathe loudly'
VN coy 'broom'	*cuqi	AN *sapuh 'broom'
Pacoh kəciik 'to comb'	*cikat	AN *sikat 'brush'
Khmer chkay 'stand straight up or out'	*cakai	MP *sakay 'climb'
Sora (*c- > s-) sar 'comb'	*caɣ	WMP *saRu 'comb'
PW *cak 'sambhar deer'	*Rucak	MP *Rusa 'deer'
Bahnar cun 'garden'	*cu[?]an	AN *suqan 'digging stick'
PW (*c- > s-) *sɔ? 'dog'	*cuq	AN *hasuh 'dog'
Sora (*c- > s-) s?ii-n 'arm'	*ciq	AN *siku[?,h] 'elbow'
Khmer caas 'contrary'	*caqi	AN *mesaq 'enemy'
Kharia (*ac >) aj 'sprinkle'	*ac	WMP *asu 'fetch water'
Semelai cəruus 'claw (nail)'	*caḷus	AN *[s]ilu[h] 'fingernail'
Pearic cu(u)c 'meat'	*cəci	AN *Sesi 'flesh, meat'
VN (*mbɔɔc >) ?bɔt 'foam'	*buc	MP *busa 'foam'
Kurku (*-c > -j) laaj 'penis'	*lac[u]	AN *lasu 'genitals'
Chrau (*c- > *s-) sec 'tooth'	*nguci[q]	AN *gusi(?) 'gums'
VN con 'bury'	*con	AN *susun 'heap up/pile up'
PNB *cuh 'set fire'	*cuq	AN *lasuq 'hot'
Brou ?əcəw'knife'	*c[ə]w	AN *pisaw 'knife'
Bahnar cɔɔr 'dig ditch'	*cu?aḷ	WMP *qasual 'lever up'

---

*/zok, nzok, jok, zək, nzək, jək/. The reductionist approach, much practiced in both AA and AN historical linguistics, would probably reject as untenable the idea that so many minimally divergent forms could be retained and reduce them to */zVk/ or perhaps */[z,d,j]Vk/ or */(z,d,j)Vk/, thereby missing the diachronic developments proposed here. It will be interesting to see in future which side of this methodological debate proves to be the more accurate.

8. See appendix for comments on font characters.

Pearic cɛɛŋ 'shine'	*caŋ	AN *ga(ñ)saŋ 'to light'
MUK moc 'one'	*muaca	AN *asa 'one'
Khmer caak 'pierce'	*cak	AN *saksak 'prick'
Semai (Sakai Ra.) cərus 'rib'	*cuɣoki	AN *Rusuk 'rib'
Pearic (mə)cok 'to smoke'	*cuk	MP *qasu 'smoke'
Gutob (*c > s) vusɑɑm 'bitter'	*cəm	MP *qe(n)sem 'sour'
VN cɔk 'pierce'	*cuk	MP *suksuk 'stab'
Katu ʔəcak 'body'	*ca[ʔ]ak	AN *Sawak 'waist'
Katu cəcah 'wash sore'	*caq	WMP *biseq 'wash'
Pearic pəcak 'wet'	*[b]acaq	MP *besaq 'wet'

In the following examples, the nasal cluster */ñc/ was recreated, and the stop became voiced, with the nasal element usually dropping off.

*ts	*c > *ñc > *j	*s
Birhor joʔ 'broom'	*cuq	AN *sapuh 'broom'
OM jik 'to harrow'	*cak	AN *[s]aka[h] 'cultivate'
Kĕnaboi rajak 'deer'	*Rucak	MP *Rusa 'deer'
Chrau jaar ka 'fish with a pole'	*cuʔaḷ	WMP *qasual 'lever up'
Kharia tijɔʔ 'worm'	*tecəwaq	AN *sawa 'giant snake'
Jeh jam 'salty'	*cəm	MP *qe(n)sem 'sour'

4.1.2.2. Final Position.

Stieng tec 'break (as a string)'	*tac	AN *ge(n)tas 'break'
OK kac 'break off, snap'	*kac	AN *gas 'broken in two'
MUK ʔbac 'cut to a point'	*mpac	AN *(C,t,T)a(ŋ)pas 'cut (off)'
Katu tac 'chop wood'	*tac	WMP *qutas 'cut through'
PW *ɲac 'wet'	*[ŋh]ac	AN *leŋas 'damp'
Bahnar moc 'dive in water'	*moc	AN *lemes 'dip, dive'
Nicobar (Teressa) paic 'venomous snake'	*pac	MP *upas 'poisonous, as a snake'
Katu (ʔə)ruəyʔ 'field and unhusked rice'	*uɣac	AN *beRas 'rice between harvesting and cooking'
Sengoi (*suuc >) suit 'wash'	*zoc	AN *kaDus 'rub'
Jehai (*-c > -j) mpɔj 'salt'	*mpuc	AN *[t,T]imus 'salt'
MUK pɛc 'fish scales'	*pac	AN *kupas 'scale off'
Che' Wong kikɔc 'scratch'	*kuc	AN *kuskus 'scrape, scratch'
Mendriq kac 'scratch'	*kac	AN *kaskas 'scratch'
Chrau gac 'chop small wood'	*nkac	AN *ragas 'sever'
PM *k_ləc 'slip'	*ḷac	AN *dulas 'slide, slip'
Khmer pluuc 'slip through a passage'	*puluc	AN *pelus 'slip off'
Khmer sroc 'sprinkle'	*ɣoc	AN *diRus 'spray, sprinkle'
Kurku roj(-ki) 'wring, squeeze'	*ɣəc	MP *peRes 'squeeze'
MUK məc 'honey	*mec	MP *amis 'sweet'
Katu ʔɑɑc 'thin'	*uc	WMP *kurus 'thin, lean'
Bahnar prach 'splash on water'	*saɣac	AN *SuRas 'wash body parts'
Sre rəpoc 'stroke'	*puac	AN *qapus 'wipe'
Khmer kuəc 'knot'	*kəc	AN *ta(R)kes 'wrap around'

4.1.3. Reflexes of Austric *z. Only examples from the appendix are cited; for additional ones, see Hayes 2000a.

*z	*z > *s	*z > *D
Katu sənɛɛt 'belt for skirt'	*zanit	AN *genDit 'belt'
Pacoh sɨk 'butt, gore'	*zok	AN *[t]un[D]uk 'to bow'
Katu (High) ʔəsʏʔ 'leaf'	*za(n)qa	AN *[d,D]aqan 'branch'
Stieng saw 'see'	*zaw	AN *tin[D]aw 'look at closely'
Pacoh saat 'wipe off, rub'	*zat	AN *DasDas 'rub off'
OM saŋ 'conch'	*zaŋ	AN *qudaŋ 'shrimp'
Chrau sʏʏr 'take out, up'	*zeɣ	MP *ma-diRi 'stand'
Mang zum 'water'	*zom	MP *danum 'water (fresh)'
Brou sɔɔy 'tail'	*uzay	AN *huDay 'worm'

4.1.4. Reflexes of Austric *dz.

4.1.4.1. Initial and Medial Position.

*dz	*j	*z, *Z?
VN (*jəh >) zɨ 'wicked'	*jəq	MP *zaqat 'bad'
Sora jɑkkɑŋ 'firm'	*jag	AN *zagzag 'balance'
Biat njiŋ 'to weigh'	*jeŋ	AN *ziŋziŋ 'balance something on hands'
Palaung jən 'be heavy'	*jən	MP *qezen 'bearing down'
Thavung [*ʔajəəp >] ʔəcoɔɔp məlooɔy 'lightning'	*jəp	WMP *kizap 'blink'
Sengoi jʌŋka 'chin'	*jaŋka	AN *zaŋgut 'chin'
Sora ji-ji 'wear as a cloth'	*juh	AN *bazuh 'clothes/clothing'
Sora jəruu-n 'deep'	*jar[ʔ]u	AN *zahuq 'far'
Sora onjər-neeb-ən 'the fig tree'	*jari	AN *zabih 'ficus (species)'
Chrau juuʔ 'black'	*jaw	WMP *hizaw 'fighting cock with greenish feathers on light background'
Bahnar jrɔwʔ 'make stew'	*juɣoq	AN *zuRuq 'fluid/liquid'
Chrau kənji 'grass'	*je	AN *bali(j,z,Z)i 'grass'
Katu rəjɔɔl 'bamboo (thick, small)'	*jəlay	AN *zelay 'unspecified grass species'
Palaung jɔk 'lift'	*juk	AN *ta(ñ)zuk 'jut out'
Mundari jal 'lick'	*jeḷat	AN *Zilat 'lick'
Pearic (*-j > -c) laac 'lightning'	*ḷajap	Formosan (Paiwan) *ḷadzap 'lightning'
PSB *jɔɔŋ 'long'	*[u]jaŋ	WMP *kazaŋ 'long'
VN (*jaak >) *cak 'tie, rope'	*jak	AN *ta(ñ)zak 'rope (for boat)'
Chrau daq ju 'saliva'	*joɣ	WMP *qizuR 'saliva'
Kharia taraju 'balance [noun]'	*juq	AN *tarazuh 'scales'
Palaung juuur 'buy'	*ju[ʔ]aḷ	AN *zuhal 'sell'
Katu jəəc 'weave by machine'	*jəqit	AN *zaqit 'sew'
Jeh pla jei 'sharp'	*rajay	Formosan (Paiwan) *[r]adzay 'sharp'
Khmer khjaak 'spit out'	*jaqi	MP *luzaq 'spit'
PW *jɛŋ 'to stand'	*jəŋ	AN *zeŋzeŋ 'stand (up)'
Mon jak 'march'	*jak	MP *bezak 'step'

Bahnar jar 'pitch of tree'	*jər	AN *pizer 'to stick'

4.1.4.2. Final Position.

*ds > *dz	*j	*j
Chrau (*-j > -c) sum ʔɔɔc 'sparrow'	*[ʔ]uj	AN *baluj 'dove species'

4.1.5. Reflexes of Austric *nts.

*nts	*ñc > *ñ'c > *ñ'j	*ñs > *s
VN ñɐy 'to sneeze'	*ñcaqi	AN *kesaq 'breathe loudly'
Katu ʔyiik 'gather plants'	*ñcak	AN *[s]aka[h] 'cultivate'
KY təyaak 'sambhar deer'	*Ruñcak	MP *Rusa 'deer'
Pacoh ʔyəʔyuul 'fish with baited hook over a hole'	*ñcuʔaḷ	WMP *qasual 'lever up'
Kotua muyʔ 'one'	*muañca	AN *asa 'one'
Katu gəyɨək 'smoke'	*ñcuk	MP *qasu 'smoke'
Katu kəʔyɛp 'suck'	*ñcep	AN *qesep 'sip, suck'

4.1.6. Reflexes of Austric *ndz.

*ndz	*ñj > *ñ'j	*ñz > *z, *Z?
Boriwen kəyaʔ 'bad'	*ñjaq	AN *zaqat 'bad'
PW *yɨn 'press down'	*ñjən	MP *qezen 'bearing down'
Bahnar ʔyeel 'smooth'	*ñjeḷ	AN *ŋa[z,Z]el 'blunt'
Sora rəyɑm-ən 'a loan for a short time'	*ñjam	AN *hinzam 'borrow'
VN ñɐm 'close eyes'	*ñjəm	AN *kezem'close the eyes'
Katu ʔəʔyɑh 'blouse'	*ñjuh	AN *bazuh 'clothes/clothing'
Brou yərəw 'deep'	*ñjar[ʔ]u	AN *zahuq 'far'
Temoq gəyɔ̃w 'night'	*ñjaw	WMP *hizaw 'fighting cock with greenish feathers on light background'
Bahnar həʔyɤɤl 'type of bamboo'	*ñjəlay	AN *zelay 'unspecified grass species'
PM *yuk 'lift'	*ñjuk	AN *ta(ñ)zuk 'jut out'
Pacoh ʔyiəl jal 'lick'	*ñjeḷat	AN *Zilat 'lick'
PK gəʔyoŋ 'high, long'	*[u]ñjaŋ	WMP *pañzaŋ 'long'
PMN *yoy 'lip(s)'	*ñjoy	Formosan *ŋudzuj 'mouth'
OM kyaal 'wind'	*ñjal	AN *quzaN 'rain'
Temiar ʔəñuh 'heavy'	*ñjuq	AN *tarazuh 'scales'
Katu ʔyiək 'weave fish net'	*ñjəqit	AN *zaqit 'sew'
PK *ʔəʔyɨŋ 'stand'	*ñjəŋ	AN *zeŋzeŋ 'stand (up)'

4.2. Irregular Correspondences. The irregularities in phonological correspondence exhibited by the examples in this subsection appear to reflect primarily differences in the usage of the */N/ affix, either in the Austric dialects at the time when affrication of the sibilants began or in Austroasiatic and/or Austronesian and their subgroups at much later dates.

4.2.1. Reflexes of Austric *s.

*s *ns > *nts > *ts	*c	*s > *h, *S
Sora (*c- > s-) serum 'to smell'	*car?om	AN *ha[r]um 'aroma'
Pearic cu(u)c 'meat'	*cəci	AN *Sesi 'flesh, meat'
Pearic chaak 'seed'	*caq	AN *bineSiq 'seed rice'
Katu ?əcak 'body'	*ca[?]ak	AN *Sawak 'waist'
Katu ncɑ 'wash hair'	*caw	AN *SawSaw 'wash'

*s *ns > *nts > *ts	*s	*s
Thavung suh 'nest'	*suq	AN *pu(ñ)suh 'anthill'
Santali sar 'arrow'	*saɣ	AN *busuR 'bow'
Bahnar səpuuyh 'sweep'	*sapuqi	AN *sapuh 'broom'
Bonda sɔsɔ 'rotten'	*su	MP *busuk 'rotten'
Jehai kasak 'deer'	*Rusak	MP *Rusa 'deer'
Khmer suən 'garden'	*su[?]an	AN *suqan 'digging stick'
PVM *si 'arm, hand'	*siq	AN *siku[?,h] 'elbow'
Pacoh son 'push'	*son	AN *susun 'heap up'
PSB *rəsɔh 'lungs'	*suq	AN *pusuq 'heart'
PSB *sɔh 'kindle'	*suq	AN *lasuq 'hot'
Thavung ksaŋ 'tooth'	*saŋ	AN *hi(ñ)saŋ 'jaw'
Katu sɔɔr 'replant'	*su?aḷ	WMP *qasual 'lever up'
Santali basaŋ 'warm'	*saŋ	AN *ga(ñ)saŋ 'to light'
Sora siŋgɛr 'ginger'	*siqaŋ	AN *saqaŋ 'sharp-tasting'
Mundari si?b 'to smoke'	*sep	AN *qesep 'sip, suck'
Kharia sɔw 'husband'	*saw[a]	MP *qasawa 'spouse'
Bahnar kəhrɛt 'tie securely'	*sirat	AN *si[r]at 'tie (together/up/on)'
OM sah 'wash the face with'	*saq	WMP *biseq 'wash'

*s *ns > *nts > *ts	*s > *ns > *nt > t, ṭ	*s
Mundari tuka 'nest'	*suq	AN *pu(ñ)suh 'anthill'
PM *tɔh 'breast'	*suq	WMP *susu 'breast'
Khmer kaṇṭaas 'to sneeze'	*saqi	AN *kesaq 'breathe loudly'
PW *te? 'hand'	*siq	AN *siku[?,h] 'elbow'
Pacoh tərta? 'fight each other'	*saq	AN *mesaq 'enemy'

4.2.2. Reflexes of Austric *z.

*z *z	*z > *nz > *nd	*z > *D
Sora duj 'bend'	*zok	AN *[t]un[D]uk 'to bow'
VN ?dam 'have an intense desire for'	*zam	WMP *kidam 'miss, crave'
Pacoh ndɔɔy 'little finger/toe'	*uzay	AN *huDay 'worm'

*z *nz > *ndz > *dz	*j	*z > *D
Katu pəjuk 'lower head to butt'	*juk	AN *[t]un[D]uk 'to bow'
Katu (High) ʔəjɨɨh 'snag'	*jaqa	AN *[d,D]aqan 'branch'
Palaung joh 'drop down'	*juq	AN *tuDuq 'drip'
Santali joha 'cheek'	*juqa	AN *[d,D]ahay 'forehead'
Kurku jimu 'name'	*jam[u]s	AN *(q)a(d,D,j)a(n,N)[9] 'male personal name'
VN (*jəm >) zəm 'lustful'	*jam	WMP *kidam 'miss, crave'
VN (*jat >) zɐt 'wash, launder'	*jat	AN *DasDas 'rub off'
PSB *juun 'deer'	*jeŋi	AN *sala[d,D,j]eŋ 'ruminant'
VN (*jəp >) cəp toy 'at nightfall'	*jəp	WMP *si(n)dep 'set, of the sun'
Khmer khjaak 'spit out'	*jaqi	AN *ludaq 'spittle'
Kharia juraʔ 'thorn'	*ju[R]aq	AN *[d,D]uRih 'thorn'
Khasi jhum 'vapor'	*jom	MP *danum 'water (fresh)'
VN zɔy 'worm, larva'	*ujay	AN *huDay 'worm'

*z *nz > *ndz	*ñj > *ñ'j	*z > *D
Katu ʔyiic 'belt'	*ñjit	AN *genDit 'belt'
VN ñɐm 'close eyes'	*ñjəm	WMP *kidem 'close the eyes'
Pacoh tənʔyɔh 'drip (water)'	*ñjuq	AN *tuDuq 'drip'
Sengoi tinyaw 'to watch'	*ñjaw	AN *tin[D]aw 'look at closely'
Pacoh kəʔyur 'weak or limber and shaky place'	*ñjuɣ	AN *kenDuR 'loose'
Sora əñɑm 'name'	*ñjam[u]s	AN *(q)a(d,D,j)a(n,N) 'male personal name'
Sora kəyeem 'be fond of'	*ñjam	WMP *kidam 'miss, crave'
Katu (High) ʔyaat 'scrub'	*ñjat	AN *DasDas 'rub off'
Pacoh ʔəʔyɤñ 'barking deer'	*ñjeŋi	AN *sala[d,D,j]eŋ ' ruminant'
Santali əyuʔb 'evening'	*ñjəp	WMP *si(n)dep 'set, of the sun'
Khmer khyaŋ 'shellfish'	*ñjaŋ	AN *qudaŋ 'shrimp'
Sora ub-yeer 'rise up'	*ñjeɣ	MP *ma-diRi 'stand'
Sora ñum-ñum 'pass urine'	*ñjom	MP *danum 'water (fresh)'

*z *nz > *ndz > *dz	*z > *nz > *nd	*z
VN ʔdɐk 'thick, strong'	*zag	AN *zagzag 'balance, stand firm'
Bonda ḍonḍeŋ 'difficulty'	*zeŋ	AN *ziŋziŋ 'balance something on hands'
Katu dil 'smooth'	*zeḷ	AN *ŋa[z,Z]el 'blunt'
Chrau ndəy 'different'	*zay	AN *zayu[h] 'foreigner'
Stieng ndoy 'lip(s)'	*zoy	Formosan *ŋudzuj 'mouth'
PSB *dəŋ 'to stand'	*zəŋ	AN *zeŋzeŋ 'stand (up)'

9. Blust's replacement for Dempwolff's PAN */['ʼ]ag'an/ = */[h]ajan/ 'name'.

*z *nz > *ndz > *dz	*z > *s	*z
Pacoh sa?(-si) 'demons'	*zaq	MP *zaqat 'bad'
Pacoh nsiil 'very dull'	*zeḷ	AN *ŋa[z,Z]el 'blunt'
Che' Wong nisɯŋ 'long'	*zaŋ	WMP *kazaŋ, *panzaŋ 'long'
PM *sroo? 'rice plant'	*zarawaq	AN *zawa[h] 'millet'
Pearic khsal 'wind'	*zal	AN *quzaN 'rain'
Ong kser 'sneeze'	*zoɣ	WMP *qizuR 'saliva'
Bahnar sit 'sew'	*zaqit	AN *zaqit 'sew'
Bahnar kəsɔh 'spit'	*zaq	MP *luzaq 'spit'
PSB sa:k 'go, return'	*zak	MP *bezak 'step'
Khmer phsaar 'join (two member) with gum or glue'	*zər	AN *pizer 'to stick'

In a few examples, AA /c/ corresponds to AN */D/ or */z/. The AA palatal probably reflects cluster coalescence, as in */z > s > [k]s/ > /c/, where [k] denotes a prefix of indeterminate nature.

MUK cɐt 'wash clothes'	*zat	AN *DasDas 'rub off'
VN cɐk 'firm'	*zag	AN *zagzag 'balance'
VN cʏp 'blink'	*zəp	WMP *kizap 'blink'
Khmer caŋka 'chin'	*zaŋka	AN *zaŋgut 'chin'
Souei cuəl 'pay a salary'	*zu[?]aḷ	AN *zuhal 'sell'
KY kəcuh 'spit'	*zuq	MP *luzaq 'spit'

These AA examples reflect Austric */z/ and */dz/, respectively. The AN */j/ is irregular in both cases, but probably reflects Austric */dz/ or */ndz/.

Bonda si? 'fever, pain'	*ziq	AN *Sapejiq 'smarting, stinging pain'
PNB *ji? 'sick'	*jeq	AN *Sapejiq 'smarting pain'

5. Conclusion.

As the preceding demonstration reveals, there is good reason to believe that affricates did not exist at an early stage in the history of the Austric languages, but came into being due to epenthesis in nasal-sibilant clusters and coalescence of initial and final stop/sibilant clusters. The former development led to a split of the Austric denti-alveolar sibilants into sibilant and affricate reflexes which can be reconstituted on the basis of the correspondence of reflexes of those original sibilants found in both the AA and AN languages. It cannot be presently determined at what precise stage that split occurred, but one may reasonably assume that Proto-Austric possessed an affricate series which changed to palatal stops in Proto-Austroasiatic, but may have been continued in Proto-Austronesian.

This presentation, together with "The Austric Denti-alveolar Sibilants", provides some of the most significant and compelling linguistic evidence for the genetic relatedness of Austroasiatic and Austronesian ever published. The quantity and quality of the supportive lexical evidence may be less than optimal in comparison to more well-established language families, but then it may be all that is or ever will be available for Austric. Perhaps more important is the fact that this evidence is amenable to the analysis and interpretation shown above, which depict a

coherent and credible pattern of diachronic phonological evolution in the languages concerned.

REFERENCES

Banker, John and Elizabeth and Mơ. 1979. *Bahnar Dictionary, Plei Bong-Mang Yang Dialect.* Huntington Beach: SIL.

Barker, Milton and Muriel. 1976. *Mường-Vietnamese-English Dictionary.* Microfiche. Huntington Beach: SIL.

Benedict, Paul K. 1975. *Austro-Thai: Language and Culture, with a Glossary of Roots.* New Haven: HRAF Press.

Benedict, Paul K. 1976. "Austro-Thai and Austroasiatic." *AS*:1–36.

Benedict, Paul K. 1990. *Japanese/Austro-Tai.* Ann Arbor: Karoma Publishers.

Bhattacharya, Sudhibhushan. 1968. *A Bonda Dictionary.* Poona: Deccan College Post-Graduate and Research Institute.

Biligiri, H.S. 1965. *Kharia: Phonology, Grammar and Vocabulary.* Poona: Deccan College, Postgraduate and Research Institute.

Blažek, Václav. 2000. "Comments on Hayes, The Austric Denti-alveolar Sibilants." *Mother Tongue V*:15-7

Blood, Henry F. 1968. *A Reconstruction of Proto-Mnong.* Grand Forks: SIL-University of North Dakota.

Blust, Robert.A. Undated. *Austronesian Comparative Dictionary.* Manuscript (work in progress).

Blust, Robert A. 2000. "Comments on Hayes, The Austric Denti-alveolar Sibilants." *Mother Tongue V*:19-21

Costello, Nancy A. 1971. *Katu Vocabulary.* Saigon: Bộ Giáo Dục.

Dahl, Otto Christian. 1973. *Proto-Austronesian.* Scandinavian Institute of Asian Studies Monograph Series No. 15. Lund: Studentlitteratur.

Diffloth, Gérard F. 1977. "Mon-Khmer Initial Palatals and 'Substratumized' Austro-Thai." *MKS VI*:39-57.

Diffloth, Gérard F. 1980. "The Wa languages." *LTBA* 5.2.

Diffloth, Gérard F. 1984. *The Dvaravati-Old Mon Language and Nyah Kur. Monic Language Studies 1.* Bangkok: Chulalongkorn University.

Ferlus, Michel. 1974. "Lexique souei-français." *ASEMI* V(1):141–59.

Ferlus, Michel. 1979. "Lexique thavung-français." *CLAO* 5:71–94.

Ferlus, Michel. 1983. "Essai de phonétique historique du Môn." *MKS* XII:1–90.

Ferlus, Michel. 1991. "Vocalisme du proto-viet-muong." Paper circulated at the 24th International Conference on Sino-Tibetan Languages and Linguistics, Bangkok and Chiang Mai, Thailand.

Fleming, Hal. 2000. "LaVaughn Hayes and Robert Blust Discuss Austric." *Mother Tongue V*: 29-32

Gradin, Dwight and Thong. 1979. *Jeh Vocabulary.* Huntington Beach: SIL.

Gregerson, Kenneth and Marilyn. 1977. *Rengao Vocabulary.* SIL.

Haupers, Lorraine and Ralph. 1991. *Stieng-English Dictionary.* Manila: SIL.

Hayes, La Vaughn H. 1992. "On the track of Austric, part I: introduction." *MKS* XXI:143–77.

Hayes, La Vaughn H. 1997a. "Another look at final spirants in Mon-Khmer." *MKS* XXVI:51–64.

Hayes, La Vaughn H. 1997b. "On the track of Austric, part II: consonant mutation in early Austroasiatic." *MKS* XXVII:13–41.

Hayes, La Vaughn H. 1999. "On the track of Austric, part III: basic vocabulary comparison." *MKS XXVIII*:1-34.

Hayes, La Vaughn H. 2000a. "The Austric Denti-alveolar Sibilants." *Mother Tongue V*:1-12.

Hayes, La Vaughn H. 2000b. "Response to Blažek's Comments." *Mother Tongue V*:33-4.
Hayes, La Vaughn H. 2000c. "Response to Blust's Comments." *Mother Tongue V*:35-7.
Hayes, La Vaughn H. 2000d. "Response to Fleming's Comments." *Mother Tongue V*:39-40.
Headley, Robert K. Jr. 1977. "A Pearic vocabulary." *MKS* VI:69–149.
Jacob, Judith M. 1974. *A Concise Cambodian-English Dictionary*. London: Oxford University Press.
Lindell, Kristina. 1974. "A vocabulary of the Yuan dialect of the Kammu language." *Acta Orientalia* 36:191–207.
Means, Nathalie and Paul B. Means. 1986. *Sengoi-English, English-Sengoi Dictionary*, ed. Gordon P. Means. Toronto: Joint Centre on Modern East Asia, University of Toronto-York University.
Miller, Carolyn and John. 1967. *English-Brou Language Familiarization Manual*. Saigon: SIL.
Nguyễn Đình Hoà. 1971. *Vietnamese-English Student Dictionary*. Carbondale: Southern Illinois University Press.
Nguyễn Phú Phong, Trần Trí Dõi and Michel Ferlus. 1988. *Lexique Vietnamien Ruc-Français, Parler d'une minorité ethnique des montagnes de Quang Binh, Vietnam,avec Notes ethnographiques et introduction linguistique*. Paris: Sudestasie.
Peiros, Ilia. 1996. *Katuic Comparative Dictionary*. Pacific Linguistics Series C-132. Canberra: Pacific Linguistics.
Pinnow, Heinz-Jürgen. 1959. *Versuch einer historischen Lautlehre der Kharia-Sprache*. Wiesbaden: Otto Harrassowitz.
Rabel, Lili. 1961. *Khasi, a Language of Assam*. Baton Rouge: Louisiana State University.
Ramamurti, G.V. 1938. *Sora-English Dictionary*. Madras:Superintendent, Gov't Press. Reprinted 1986, Delhi:Mittal Publications.
Shorto, H.L. 1962. *A Dictionary of Modern Spoken Mon*. London: Oxford University Press.
Shorto, H.L. 1971. *A Dictionary of the Mon Inscriptions from the Sixth to the Sixteenth Centuries*. London: Oxford University Press.
Sidwell, Paul J. 2000. *Proto South Bahnaric, A reconstruction of a Mon-Khmer language of Indo-China*. Canberra: Pacific Linguistics.
Singh, U. Nissor. 1906. *Khasi-English Dictionary*, eds. Major P.R.T. Gurdon, U Dohory Ropmay, and U Hajom Kissor Singh. Reprinted 1966, New Delhi: Gyan Publishing House.
Skeat, Walter W. and Charles O. Blagden. 1906. *Pagan Races of the Malay Peninsula, Volume II*. London: Macmillan. Reprinted 1966, New York: Barnes & Noble.
Smalley, William A. 1961. *Outline of Khmu' Structure*. New Haven: American Oriental Society.
Smith, Kenneth D. 1972. *A Phonological Reconstruction of Proto-North-Bahnaric (Language Data, Asian-Pacific Series, No. 2)*. Santa Ana: SIL.
Theraphan L. Thongkum. 1984. *Nyah Kur (Chao Bon)-Thai-English Dictionary, Monic Language Studies 2*. Bangkok: Chulalongkorn University.
Thomas, David D. and Thổ Sáng Lục. 1966. *Chrau Vocabulary*. Saigon: Bộ Giáo Dục. Reprinted 1977, Huntington Beach: SIL.
Thomas, David and Sophana Srichampa, eds. 1996. *Phraya Prachakij-karacak, Some Languages of Siam*. Translated and annotated. Bangkok: Mahidol University.
Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects, Two Thousand Years of Language Contact and Change*. Honolulu: University of Hawai'i Press.
van Driem, George. 2000. "Four Austric Theories." *Mother Tongue V*:23-7
Watson, Richard and Saundra and Cubuat. 1979. *Pacoh Dictionary*. Huntington Beach: SIL.
Wurm, S.A. and B. Wilson. 1983. *English Finderlist of Reconstructions in Austronesian Languages (Post-Brandstetter), Pacific Linguistics Series C33*. Canberra: Australian National University.

Appendix: Austric Comparative Data Glossary[1]

This glossary contains 123 AA/AN lexical comparisons, 51 of which are published for the first time. Since the presentation is aimed at an audience including non-specialists in the AA languages, the normal orthographic representation has been changed to an IPA-based notation intended to make the phonology of the cited data clear to those unfamiliar with AA orthography. Note that doubled vowels, such as /aa/, represent a long vowel (all final vowels are long). Note also that AN */e/ is phonetically [ə].

Austroasiatic	**PAA**	**Austronesian**
Thavung suh, PK *so(o)h, Mundari tuka 'nest'	*suq[2]	AN *pu(ñ)suh 'anthill'
VN (*sr > *ś > th) thʏm 'be fragrant, smell good', PW *s?ʏm 'rotting', Sora (*c- > s-) serum 'to smell'	*sa(r)?om, *car?om	AN *ha[r]um 'aroma, scent', AT *s[a][r]om 'smell, fragrant'
Pacoh[3] sa?(-si) 'demons, evil spirits that live in specific places', VN (*jəh >) zɨ 'fierce, ferocious, wicked', Boriwen kəya? 'bad'	*zaq, *jəq, *ñjaq	MP[4] *zaqat 'bad'
VN cɐk, Khmer jak 'firm', VN ?dɐk 'thick, strong', PSB *kəjək 'heavy', Sora jɑkkɑŋ 'firm, well built'	*zag, *jag, *jəg	AN *zagzag, *zegzeg 'balance, stand firm'
Biat njiŋ 'to weigh', Khmer jañjiiŋ 'scales for weighing', OM sjiŋ 'have difficulty', Bonda ḍonḍeŋ 'difficulty'	*zeŋ, *jeŋ	AN *ziŋziŋ 'balance something on hands'
Palaung jən 'be heavy', PK *həʒən 'heavily (of rain)', PW *yin 'press down'	*(ñ)jən	MP *qezen 'bearing down, pressing out, as in defecation or childbirth'
Katu sənɛɛt 'belt for skirt', Chrau chɛ nich, Katu ?yiic 'belt'	*zanit[5], *ñjit	AN *genDit 'belt, girdle'
VN cʏp '(of heaven) lighten, blink, wink, lightning', Pacoh piciip 'go by light of torch', Thavung [*?ajəəp >]	*zəp, *jəp	WMP *kizap[6] 'blink, wink, flicker, flash'

---

1. The writer was criticized for not including reconstructions from Blust's ACD in Hayes 2000a. I did not have access to the ACD when that paper was written, but have since acquired a copy and incorporated Blust's ACD proto-forms in this paper insofar as possible. Note, however, that the ACD is a work in progress and far from complete; it does not include reconstructions applicable to all of the exemplary data sets cited here.

2. Replaces */sus/ cited in Hayes 1997a:60. Also cf. Stieng /chɔɔk/ 'placenta', PSB */rəsɔɔn/, Khmer /sampuk/ 'nest', Khmer /ṭampuuk/ 'anthill', which appear to be affixed derivatives of an Austric root */su/. Cf. further Proto-Chamic */srũh/ 'nest, swarm', which Thurgood Graham identifies as a MK loan.

3. Also cf. Pacoh /pəsu?/ 'teach evil, get another to do bad deed for you', indicating a variant */zuq/.

4. Blust's replacement for Dempwolff's PAN */zaqat/. Also cf. Proto-Chamic */jəhaat/ 'wicked, bad', but note also Cham /sa?/ 'wicked, bad' and Western Cham /kəjah/ 'bad', which appear to be MK loans.

5. */zanit/ is an infixed derivative of an unattested root */zit/.

6. Blust reconstructs several kindred forms in the ACD, WMP */ki(n)sap, izap, izep, ki(n)zep, kezep/, all meaning 'blink, wink'. Also cf. WMP */andap/ 'phosphorescent millipede', apparently Blust's replacement for Dempwolff's PAN */han[D]ap/ 'flicker', which reflects the same Austric root */zVp/. Also see the 'lightning' etymology.

ʔəcoooɔp məlooɔy 'lightning'		
Pacoh nsiil 'very dull', Jeh sɔl 'dull, blunt', Katu dil, Bahnar ʔyeel 'smooth'	*zeḷ, *ñjeḷ	AN *ŋa[z,Z]el 'blunt, dull'
Chrau ñaam 'borrow', Sora rəyɑm-ən 'a loan for a short time', PK *həʔyi:m 'price, debt'	*ñjam	AN *hinzam 'borrow, lend', AT *si(n)dzam 'borrow, loan'
Santali sar 'arrow', Pearic sar, Bahnar sraʔ 'crossbow'	*saɣ(aq)[7]	AN *busuR 'hunting bow', AT *[bo](n)tsoR 'bow, spear, arrow'
Pacoh sɨk 'butt, gore', Katu pəjuk 'lower head to butt', Sora duj 'bend'	*zok(i), *juk	AN *[t]un[D]uk 'to bow, bend down/over'
Katu (High) ʔəsʏʔ 'leaf', ʔəjɨih 'snag', Mon saka 'toothstick, twig or slip of wood for cleaning teeth'	*za(n)qa, *jaqa	AN *[d,D]aqan, AT *[da]Nqa 'branch'
Kharia tɛj 'break', Bahnar kətɛc 'of string, etc., to break', Stieng tec 'break (as a string)'	*tac	AN *ge(n)tas 'break'
PW *tis, PM *tɔh, Khmer ṭoh 'breast'	*suq	WMP *susu 'breast'
Jeh cayh, VN ñɐy, Khmer kaṇṭaas 'to sneeze'	*saqi, *(ñ)caqi	AN *kesaq 'breathe loudly'
OK kac 'break off, snap', Pearic kac 'break', Pacoh tikayh 'broken'	*kac	AN *gas 'broken in two'
VN coy, Birhor joʔ 'broom', Bonda suk', Khmer pos, Bahnar səpuuyh 'sweep'	*cuq(i), *sapuqi[8]	AN *sapuh 'broom, sweep, wipe'
Pacoh kəciik, Maleng Bro seek, Kharia kaḍ 'to comb'	*cikat[9]	AN *sikat 'brush'
Khmer caŋka, Sengoi jʌŋka 'chin', Kharia gucu 'mustachios, beard'	*zaŋka, *jaŋka, *goc[u][10]	AN *zaŋgut 'chin, beard', AJ *dzaŋgo[t,c] 'chin'
Khmer chkay 'stand straight up or out', Pearic khəcay 'climb', VN ŋɐy 'straight, erect'	*ca(n)kay[11]	MP *sakay 'climb', AT *[(n)tsa]kai 'climb, ascend'
VN ñɐm, ñim, Pacoh ñiim 'close eyes', Chrau ñʏʏm mat 'shut eyes'	*ñjəm	AN *kezem, WMP *kidem 'close the eyes'
Sora ji-ji 'wear as a cloth', Katu ʔəʔyɑh 'blouse', Jeh ʔəyɔh 'shirt, clothes'	*(ñ)juh	AN *bazuh 'clothes/clothing, garment'

---

7. The AA and AN forms are probably not reflexes of the same proto-form, but of allomorphs of the same Austric root */sVɣ/. Also cf. PSB */gəsɔ:r/ 'quill' as a possible AA reflex of the */suɣ/ variant underlying the AN forms.

8. Replaces */ca(m)pusi/ cited in Hayes 1997a:61. Also cf. Bahnar /hədruuyh/ 'broom' from */sunruqi/. Austroasiatic has retained reflexes of the Austric root */suq/ only on the periphery of its domain in Vietic and Munda. Elsewhere, reflexes of infixed derivatives appear, as also in Austronesian.

9. Replaces */(in)kat(i)/ cited in Hayes 1997b:29,32. Pacoh /ciik/ and Kharia /kaḍ/ constitute what Benedict called split stems, each apparently reflecting a different syllable of */cikat/. An alternate explanation is that each reflects an independent root morpheme, which could be compounded as */(n)sikkat/, whence AN */sikat/.

10. It is unknown if this was an independent root morpheme once used in composition with */zaŋka/ or a truncated reflex of a single lexeme */zaŋ[ka]goc[u]/. Cf. PAN */ŋu[s]u/ 'lip(s)' as a possible correlate of */goc[u]/.

11. Replaces */(n)caka(i)/ and */ca(n)ka(i)/ cited in Hayes 1992:173 and 1999:23, respectively.

Sora (CF) sar 'comb', Khasi sar 'sweep', Stieng caas 'comb hair'	*caɣ(i)[12]	WMP *saRu 'comb'
OK jyak 'excavate', OM jik 'to harrow', Katu ʔyiik 'gather plants'	*(ñ)cak[13]	AN *[s]aka[h] 'cultivate'
MUK ʔbac 'cut to a point', Bahnar pɛc 'cut small wood', Kui pac 'slash, cut down/off with a slashing motion'	*(m)pac	AN *(C,t,T)a(ŋ)pas 'cut (off)', AT *[t]a(m)pats 'cut off/up'
Katu tac 'chop wood', OM ktaac 'to smooth, level with the hand', Kharia tɑj 'serve out (rice), distribute, share'	*tac	WMP *qutas 'cut through, sever, divide by cutting', AT *[kə]tats 'cut'
PW *ɲac 'wet', Bahnar hɲac 'light rain that continues for a long time', Jehai hej 'rain'	*[ŋh]ac	AN *leŋas 'damp'
Bonda sɔsɔ, Jehai ʔənsɔʔ 'rotten', Katu nsɔʔ 'decayed, spoiled'	*su(q)	MP *busuk 'decayed, rotten, putrid, stinking'
Jehai (Sem. Jarum, D68a) kasak, Kĕnaboi (I, D68b) rajak 'deer', PW *cak, KY təyaak 'sambhar deer'	*Rusak, *Ru(ñ)cak	MP *Rusa 'deer'
Bahnar cun, VN viən, Khmer suən 'garden'	*su[ʔ]an, *cu[ʔ]an[14]	AN *suqan 'digging stick', AT *tsu(w)an 'dig into'
Khmer muj 'sink, immerse oneself/ itself', Bahnar moc 'submerge oneself in water, dive in water', Rengao muyʔ 'submerge'	*moc	AN *lemes 'dip, dive'
PW (*c- > s-) *sɔʔ, Sabum cooʔ, Bonda (*c > s) gusɔʔ 'dog'	*cuq	AN *hasuh 'dog'
Chrau sum ʔɔɔc 'sparrow', Kharia ud 'owl', Khmer kruac 'quail'	*[ʔ]ud, *[ʔ]uj[15]	AN *baluj 'dove species'
Palaung jɔh 'drop down', Chrau juuyh 'drop, let fall', Pacoh tənʔyɔh 'drip (water)'	*(ñ)juq(i)	AN *tuDuq 'drip, trickle'
PVM *si, Sora (ə-)sʔii-n (CF sii-n) 'arm, hand', PW *teʔ, Kharia tiʔi 'hand'	*si(q), *ci(q)	AN *siku[ʔ,h][16] 'elbow'
Khasi pərʃaʔ 'adverse', Khmer caas 'contrary, against', Pacoh tərtaʔ 'fight each other, war'	*saq, *caq(i)	AN *mesaq 'enemy', AT *[m][u](ŋ)tsak 'enemy, war'

---

12. Initial */c/ > /s/ in both Sora and Khasi. The AA etyma were previously compared only with PAN */sisi[r]/ 'comb, harrow', but it seems likely that all of these forms are derived from the same Austric root */sVɣ/.

13. It is unclear whether the cited forms reflect */(ñ)ciak/ or */i(ñ)cak/, whence */ñciak/.

14. Also cf. Pearic */chew/, Jah Hut /cəwəm/ 'dig', Kharia /ən/ 'plant'.

15. The final development was */d > ds (by suffixation) > dz > j/.

16. PAN */siku[ʔ,h]/ appears to reflect an ancient compound formed of the Austric roots */si/ 'arm' and */ku/ 'bend, curve'. The former is apparently unattested elsewhere in Austronesian, but the latter appears in a number of AN reconstructions, cf. e.g. PAN */(C,t,T)iku/ 'bend, curve'. The suspect meaning was 'bend or curve of arm' = 'elbow'. In AA, Nicobar (Car) /sɛɛkɔŋ/ and Sora /kuŋ-sii-n/ 'elbow' represent the same compound, though it is uncertain that they date back to the Austric level. But direct diachronic correspondence of the compositions is unimportant because related languages retaining reflexes of common roots can be expected to form similar lexical compositions.

Sora jəruu-n, Khmu jru?, Brou yərəw 'deep'	*(ñ)jar[?]uq	AN *zahuq 'far'
Kharia (*ac >) aj 'sprinkle', Bonda a?i, PM *saac 'bale out (water)'	*ac, *sac	WMP[17] *asu 'fetch water, scoop up water'
Sora onjər-neeb-ən 'the fig tree', PM *jrəy 'ficus (banyan, pipal, Bo)', PSB jəri 'banyan tree'	*jari[18]	AN *zabih 'ficus (species)'
Chrau juu? 'black', Katu kəjɑk 'dark', Temoq gayɔ̃w 'night'	*(ñ)jaw	WMP *hizaw 'fighting cock with greenish feathers on light background', AT *(q)hi[dz]aw 'green'
Semelai cəruus, Jehai cənrɔs 'claw (nail)', PM *krmpuus 'finger, toe, breadth of finger'	*caḷus, *caḷumpusi	AN *[s]ilu[h] 'fingernail'
Bahnar sɛc, Rengao (*sərniś >) hə?nih, Pearic cu(u)c 'meat'	*səc(i), *cəc(i)	AN *Sesi, AJ *śətši 'flesh, meat'
Bahnar kəsaay? 'sprinkle, splash water', Pacoh saay? 'splash water on self or other', NK cha? (daak) 'pond'	*saq(i)	AN *baSaq 'flood(ed), (fresh) water', AT *ba[saq] 'overflow, flood'
Sora kuṛu 'porridge', Katu jəruə? 'mix meat with salt, rice to store', Bahnar jrɔw? 'make stew, mix up food'	*juɣoq	AN *zuRuq, AJ *džuɣuq 'fluid/liquid'
Thavung buuc 'make bubbles in water (fish)', VN (*mbɔɔc >) ?bɔt 'foam, bubble, suds', Palaung (Panku) buh 'foam'	*bus, *buc	MP *busa 'foam'
Santali joha, Sedang ŋiə 'cheek', Khmer thŋaas 'forehead'	*j(a,u)(n)qa(si)	AN *[d,D]ahay 'forehead', AT *(q)(n)dza[q]ai[s] 'face, forehead'
Chrau ndəy 'different', Pacoh nday 'other, different', Bahnar (?ə)naay 'different, (an)other'	*zay	AN *zayu[h] 'foreigner, stranger'
Nicobar (Car) ku-lɔic, Thavung looc, Kurku laaj 'penis'	*lac[u]	AN *lasu 'genitals'
Chrau kənji 'grass, weeds', Sora jiiŋ-ən 'weeds'	*je(nje)	AN *bali(j,z,Z)i 'grass'
NK chəlay 'Homonoia riparia', Katu rəjɔɔl 'bamboo (thick, small)', Bahnar hə?yɤɤl 'type of bamboo'	*(ñ)jəlay	AN *zelay 'unspecified grass species'
Chrau sec, PMN *sek, Sengoi lʌŋseit 'tooth'	*nguci[q]	AN *gusi(?) 'gums'
Pacoh son 'push, crowd others', Pearic so(o)n 'accompany', VN con 'bury'	*son, *con	AN *susun 'heap up/pile up', AT *tsontson 'pile up'
PK sɑh, PSB *rəsɔh, PVM *psoos 'lungs'	*suq(i)	AN *pusuq, AT *()pots[oq] 'heart, lung(s)'
PSB *sɔh 'kindle', PNB *cuh 'set	*suq, *cuq	AN *la[c,s]uq 'hot'

17. Also cf. Proto-Chamic */sac/ 'bail (water to catch fish)', which Graham identifies as a MK loan.

18. The second syllables of the AA and AN proto-forms may reflect different roots in composition with */za/. Also cf. PK /?ərii/ 'banyan', where */za/ has been replaced by a prefix.

fire', Ruc kucɔh 'fire'		
Thavung ksaŋ, Bahnar sənɛñ, PW[19] *hraŋ 'tooth'	*saŋ	AN *hi(ñ)saŋ 'jaw, jawbone'
Palaung jɔɔk, Stieng juɯk 'lift', PM *yuk 'lift, support, carry'	*(ñ)juk	AN *ta(ñ)zuk 'jut out, stand out'
Brou ʔəcəw 'knife', Pacoh ʔəciiw paat 'a large-bladed knife'	*c[ə]w[20]	AN *pisaw 'knife'
Katu sɔr 'replant', Bahnar cɔɔr 'dig ditch', Chrau jaar ka 'fish with a pole', Pacoh ʔyəʔyuul 'fish with baited hook over a hole'	*suʔaḷ, *(ñ)cuʔaḷ	WMP *qasual 'lever up', AT *(n)tsu(w)aḷ 'lever, lift (net), uproot'
Mundari jal, Pacoh ʔyiəl, Black Riang liɛt 'lick'	*(ñ)jeḷat[21]	AN *Zilat[22] 'lick'
Santali basaŋ 'warm, hot, to boil', Pearic cɛɛŋ 'shine', Pacoh cɔŋ 'burn off field'	*saŋ, *caŋ	AN *ga(ñ)saŋ 'to light', AT *[ga](n)ts[a]ŋ 'light (give light), glow, glitter, shine'
Pearic laac, Bahnar kəmlaat 'lightning', Pacoh lilaayʔ 'lightning flash'	*ḷad, *ḷajap[23]	Formosan (Kuvalan) *ḷidzap, (Paiwan) *ḷadzap, AT *ḷidzap 'lightning'
Che' Wong nisɯŋ, PSB *jɔɔŋ 'long', PK gəʔyoŋ 'high, long'	*zaŋ, *[u](ñ)jaŋ	WMP *kazaŋ, *panzaŋ 'long'
Stieng saw, Kharia yɔ 'see', Sengoi tinyaw 'to watch, look'	*zaw, *ñjaw	AN *tin[D]aw 'look at closely', AT *[ti](n)dz[a]w 'look at, see'
Pacoh kəʔyur 'weak or limber and shaky place', Bahnar ʔyiir-ʔyuur 'to describe loose-fitting trousers'	*ñjuɣ	AN *kenDuR 'loose'
Sora sɑroo 'paddy', PM *sroo? 'rice plant, paddy', Brou sərɔ 'unhusked rice'	*zaraw(aq)	AN *zawa[h] 'millet'
VN ʔdam 'have an intense desire for', (*jəm >) zəm 'lustful, sexy, lewd', Sora kəyeem 'be fond of, long for (as food, music, women)'	*zam, *(ñ)jam	WMP *kidam 'miss, crave, long for'
Stieng ndoy, PMN *yoy, PS *ñəyñooy 'lip(s)'	*zoy, *ñjoy	Formosan *ŋudzuj 'mouth, nose', AT *gu(ń)dźuy 'mouth'
Kurku jimu, Sora əñɑm 'name', Chrau tənhya 'to name'	*(ñ)jam[u]s	AN *(q)a(d,D,j)a(n,N) 'male personal name', AT *()a(ń)ja(n) 'name'
Khmer muoy, PM *mway, Kotua muyʔ, MUK moc 'one'	*muaca, *muañca	AN *asa, AT *atsa 'one'
Bonda siʔ 'fever, pain', PW *siʔ 'pain, disease', PNB *jiʔ 'sick'	*zeq, *jeq	AN *Sapejiq 'smarting, stinging pain'
Nicobar (Teressa) paic 'venomous snake', Katu kəbac 'snake'	*(m)pac	MP *upas 'poisonous, as a snake'

19. PW */s-/ > /h/; the liquid is an infix or metathesized prefix.

20. Found thus far only in Katuic. */c[ə]w/ may be a reflex of a variant */sVw/ of the Austric root */zaw/, which underlies PAN */maŋ(d,D,j)aw/ 'war sword' and various AA reflexes, such as Kharia /daw/ 'big flat sickle'.

21. Replaces */[d]iḷat/ cited in Hayes 1992:160.

22. Dyen's replacement for Dempwolff's PAN */dilat/ 'lap, lick'; no equivalent found in the ACD.

23. From */ḷadzap/, a composition of the Austric roots */ḷat/ or */ḷad/ and */zap/, cf. the 'blink' etymology.

'prick, sting', Santali caca? 'strip, tear off'		*tsak(tsak) 'prick, stick, pierce'
Pearic khsal, OM kyaal, Brou kuyaal 'wind'	*zal, *ñjal	AN *quzaN, AT *(q)[u](n)dzal 'rain'
Semai (Sakai Ra.) cərus, Sengoi cərʌs 'rib', PM *cruus 'chest'	*cuɣoki[24]	AN *Rusuk 'rib'
Katu (?ə)ruəy? 'field and unhusked rice', Sora dɑrɑj-ən 'cooked rice or millet', Kharia koṛaj 'dolichos'	*uɣac	AN *beRas 'rice between harvesting and cooking'
VN (*jaak >) *cak 'tie, rope', Ruc, Sach cəak 'rope'	*jak[25]	AN *ta(ñ)zak 'rope (for boat)'
Sengoi (*suuc >) suit 'wash, clean', Pearic dus skip 'rub', Chrau đuyh 'rub vigorously'	*(n)zoc(i)	AN *kaDus 'rub, scratch'
Pacoh saat 'wipe off, rub', MUK cɐt 'wash clothes', VN (*jat >) zɐt 'wash, launder', Katu (High) ?yaat 'scrub, clean, rub'	*zat, *(ñ)jat	AN *DasDas 'rub off'
PSB *juun 'deer', Pacoh ?ə?yʏñ 'barking deer', NK khəñɔɔŋ 'mouse deer'	*(ñ)jeŋ(i)	AN *sala[d,D,j]eŋ 'wild ruminant'
PM *sɔɔr 'blow (mucus) out of one's nose', Ong kser 'sneeze', Chrau daq ju, Sora oleen-ən (CF ool-ən) 'saliva'	*zoɣ, *joɣ	WMP *qizuR 'saliva, spittle'
Jehai mpɔj, Semnam ?əmpɔɔj, Kuy pos 'salt'	*qa(m)puc(i)	AN *[t,T]imus 'salt'
MUK pɐy, pɐc, Semang (Pa. Max., S36) kəlipɛh 'fish scales', Katu mpɛh 'scabies'	*(m)pac(i)	AN *kupas 'scale off'
Kharia taraju 'balance [noun]', Temiar ?əñuh, Brao kayut 'heavy'	*(ñ)juq	AN *tarazuh 'scales'
Che' Wong kikɔc 'scratch', Katu kɑɑc 'carve', Kharia kɔj 'peel off'	*kuc	AN *kuskus 'scrape, scratch', AT *kutskuts 'scratch, scrape, dig'
Mendriq kac, Thavung akayh 'scratch', Pearic khač 'scrape'	*(qa)kac(i)	AN *kaskas, AT *kats(kats) 'scratch'
Pacoh ca? 'root', Pearic chaak 'seed, grain', Bahnar hədrɛc 'seed rice'	*caq(i), *c[i]nraqi	AN *bineSiq 'seed rice', AT *b[i]nsaq 'seed'
Souei cuəl 'pay a salary', Chrau joor 'hire, call', Palaung juuur 'buy'	*zu[?]aḷ, *ju[?]aḷ	AN *zuhal 'sell', AT *dzu(w)aḷ 'sell, buy'
VN (*jəp >) cəp toy 'at nightfall', Khmer yup 'night', Santali əyu?b 'evening, nightfall'	*(ñ)jəp	WMP *si(n)dep 'set, of the sun'
Chrau gac 'chop small wood', MUK kac 'to cut, reap, harvest', Che' Wong	*(n)kac	AN *ragas 'sever'

24. Metathesis is apparent, but in which family is unclear.

25. Also cf. Proto-Thai */jïak/ 'rope'. Possibly a loan word, but it is unclear who borrowed from whom. Chamic does not possess this etymology. Cf. further PAN */hizuk/ 'bundle of fibres' and Proto-Mon */jook/ 'creeper, vine, rope, string, cord, lashing', which possibly reflect an allomorph of the same root underlying Austric */dzak/.

kac 'to cut'		
Bahnar sit 'sew', Katu jəəc 'weave by machine', ʔyiək 'weave fish net'	*zəqit, *(ñ)jəqit	AN *zaqit 'sew'
Pacoh raayʔ 'sharpen to a point', OK rac 'come to an end or point', Jeh pla jei 'sharp'	*rajay	Formosan (Paiwan) *[r]adzay 'sharp (as blade)', AT *r[a]dzay 'sharp'
Sora síŋgɛr 'ginger', Khasi kha sʔiŋ 'dried fish with ginger', PMN *haŋ 'peppery'	*siqaŋ	AN *saqaŋ 'sharp-tasting', AT *[ts]iqaŋ 'sharp (tasting), ginger, pepper'
OM saŋ 'conch', Rengao kəsoŋ 'shrimp', Khmer khyaŋ 'shellfish'	*zaŋ, *ñjaŋ	AN *qudaŋ 'shrimp, crayfish, lobster', AT *qu(n)zaŋ 'crustacean, shrimp'
Mundari siʔb 'to smoke', Khmer jip 'sample (with pursed lips), sip', PK *həʔyeep 'suck'	*sep, *ñcep	AN *qesep, *sepsep 'sip, suck', AT *tsəptsəp 'suck'
PM *k_ləc 'slip', Chrau lac 'fall out through hole', Mundari jilad 'slip, slippery'	*l̥at, *l̥ac	AN *dulas 'slide, slip'
Khmer pluuc 'slip through a passage', MUK tlɔc 'slip into', Katu kəlɑɑc 'crawl'	*puluc	AN *pelus 'slip off', AT *(q)lutsluts 'slip off/away'
Pearic (mə)cok 'to smoke', Khmer jak 'suck in, smoke (tobacco)', Katu gəyiək 'smoke'	*(ñ)cuk	MP *qasu 'smoke, fumes, steam; to smoke (as a fire)', AT *(qa)(n)tsu[b](an) 'smoke'
Kharia tijɔʔ 'worm', PM *tñjuuʔ 'worm, caterpillar, maggot', Temoq tejow 'snake'	*te(ñ)cəw(aq)	AN *sawa 'giant snake'
Gutob vusɑɑm 'bitter', Jeh jam 'salty', VN (*jəm >) zəm 'vinegar'	*(ñ)cəm	MP *qe(n)sem, AT *(q)atsəm 'sour'
Bahnar kəsɔh, KY kəcuh 'spit', Khmer khjaak 'spit out'	*zaq, *zuq, *jaqi	MP *luzaq 'spit', AJ *l̥udaq 'spittle/saliva'
Kharia sɔw 'husband', Katu səsaaw 'father's cousins, sister's husband, father's sister's children', Proto-Semai *bnsaaw 'wife's elder brother'	*saw[a]	MP *qasawa 'spouse', AT *(qa)(n)saw[a] 'woman, wife, spouse'
Palaung rət, Khmer sroc 'sprinkle', MUK rɔc 'pour'	*ɣot, *ɣoc	AN *diRus 'spray, sprinkle', AT *[di]ɣots 'sprinkle, bathe'
Kurku roj(-ki) 'wring, squeeze', Kui ŋac 'squeeze, knead, massage', Chrau ret 'squeeze'	*ɣət, *(n)ɣəc	MP *peRes 'squeeze'
Sora suj, VN cɔk 'pierce', Sengoi cok 'stab, pierce'	*cuk	MP *suksuk 'stab', AT *(n)tsuk((n)tsuk) 'stick/push into'
Chrau sʏʏr 'take out, up', Sora ub-yeer 'rise up', Pacoh yoor 'get up, arise'	*zeɣ, *ñjeɣ	MP *ma-diRi 'stand', AT *(q)(n)ź[u]ɣ[i] 'stand'
PSB *dəŋ, PW *jɐŋ, PK *ʔəʔyiŋ 'to stand', Sora ñooŋ-ñooŋ-jeeŋ 'stand on tip-toe'	*zəŋ, (ñ)jəŋ[26]	AN *zeŋzeŋ 'stand (up)', AT *()dzəŋ(dzəŋ) 'stand'

26. Cf. Proto-Chamic */dəŋ/ 'stand', which Graham identifies as a MK loan.

PSB sa:k 'go, return', Mon jak 'march, travel', Jeh juə? 'stamp on, tread, step'	*zak, *jak, *jaq	MP *bezak 'step, tread, stamp on'
Khmer phsaar 'join (two members) with gum, glue or the like', Bahnar jar 'pitch of tree', Khmer jaar 'sap, resin'	*zər, *jər	AN *pizer 'to stick'
Kharia uḍ 'drink, suck', Chrau huuc 'drink', Proto-Semai *huc 'to sip'	*ud, *quj	WMP *qudud 'to suck'
MUK məc, Katu mit 'honey', Sengoi cʌʔɛt 'sweet'	*[ʔ]et, *met, *mec	MP *amis 'sweet'
Katu ʔɑɑc 'thin', Khmer ṣtuac 'slender, slight', Jeh roh 'lean'	*uc, *rus	WMP *kurus[27] 'thin, lean'
Kharia juraʔ 'prickle, thorn', PM PM *jrlaaʔ, Sengoi jʌrlaa 'thorn'	*ju[R]aq	AN *[d,D]uRih, AT *dzuRi(an) 'thorn'
Bahnar kəhrɛt 'tie securely', Pacoh sət 'tie in a bundle'	*sirat	AN *si[r]at 'tie (together/up/on)', AT *(ts)[i]rat 'bind, tie'
Bahnar kiak 'ghost, corpse', Katu saak 'corpse', ʔəcak 'body'	*[ʔ]ak, *sa[ʔ]ak, *ca[ʔ]ak	AN *Sawak[28] 'waist', AT *[a](w)ak 'chest, body'
OM sah 'wash the face with', Pacoh sah 'wash, mop, scrub (dishes)', Katu cəcah 'wash sore'	*saq, *caq	WMP *biseq, AT *(m)b[a](n)tsaq 'wash'
Jeh saw mat 'wash (face)', Katu ncɑ 'wash hair', Chrau təcɔɔ 'rub between hands'	*saw, *caw	AN *SawSaw 'wash, rinse', AT *sawsaw, *ntsawntsaw, *[lu]tsaw 'wash, rinse'
Bahnar prach 'splash on water', Khasi (*-c > -it) pharait 'spatter, squirt', Chrau jraih 'sprinkle ceremonially'	*saɣac(i)	AN *SuRas 'wash body parts, cooking or eating utensils (but not clothes)', AT *(n)suɣa(t)s 'wash'
Mang zum 'water', Gutob dum 'drown', Khasi jhum 'vapor', Sora ñum-ñum 'pass urine'	*(n)zom, *(ñ)jom	MP *danum[29] 'water (fresh)', AT AT *()(n)zalom 'water'
Pearic pəcak, Chrau suh, Tampuan tšətšuyh 'wet'	*[b]acaq, *[u]c[a]q(i)	MP *besaq 'wet'
PM *smpɔɔt 'stroke, rub, pass hand over, wipe', Sre rəpoc 'stroke', Katu buəc 'wash face'	*mpuat, *(m)puac	AN *qapus 'wipe'
Brou sɔɔy 'tail', Pacoh ndɔɔy 'little finger/toe', VN (*jɔɔy >) zɔy 'worm, larva'	*uzay, *ujay	AN *huDay 'worm', AT *(q)u(n)zay 'worm, larva, penis'
Khmer kuəc 'knot', Chrau gɔɔc 'to lassoo', Katu (High) klɔc 'to tie'	*(n)kəc, *kələc	AN *ta(R)kes 'wrap around, encircle', AT *[b]ə[ɣ][ə](ŋ)kots 'tie, enlace'

27. In the ACD, Blust cites */kurus/ as a doublet of */kuru/ 'thin, lean'. The AA data suggests an Austric root */u/, from which a variety of affixed derivatives and compositions could be built.

28. Blust's replacement for Dempwolff's PAN */hawaq/ 'body'. The AA etyma indicate an Austric root */[ʔ]Vk/, which could be inflected in a variety of ways, */(n)sa[ʔ]ak/ being perhaps the most common.

29. Cf. Dempwolff's PAN */[d,D]anum/ 'water/fresh water', where the */z > D/ shift is clear. These forms apparently reflect */zalom/, an affixed derivative of the Austric root */[ʔ]om/ 'water'.

Comments on La Vaughn H. Hayes' "On the Origin of Affricates in Austric"

by Paul Sidwell
Australian National University

The paper before me seeks to demonstrate the Austric hypothesis of an Austroasiatic-Austronesian genetic relationship by illustrating a number of non-trivial correspondences. In effect Hayes claims to be showing an Austric "Verner's law" arising from a distinctive patterning among apical and laminal obstruents that is interpreted as indicating a series of Austric affricates.

GENERAL COMMENTS

My professional assessment is that the paper is unconvincing. Proof in this case would require a larger set of data and much more rigorous application of the comparative method. At present the paper should be characterized as highly speculative and not well supported empirically. The "Austric Glossary" appended to the text is a potentially useful source of comparabilia - however it must be used with great caution - in particular, in assessing comparisons for etymological merit the views of specialists in each language family/sub-family need to be sought and taken into consideration. Conclusions based on incomplete data, especially in respect of what seem to be otherwise straightforward resemblances, are extremely unsafe.

PARTICULAR COMMENTS

If for the moment one ignores any problems with the Austric Glossary, the following can be said about the discussion and interpretation of the correspondence sets.

1) Six distinct correspondence sets are proposed between AA (Austroasiatic: as reconstructed by Hayes) and AN (Austronesian: a generally received reconstruction). These can summarized as follows:

Austric	**AA**	**AN**
**s*	**s, nt*	**h, *S*
**z*	**z > *s, *nd*	**D*
**ts*	**c, *ñc*	**s*
**dz*	**j, *ñj*	**z, *Z?*
**nts*	**c, *ñ'c*	**ñs*
**ndz*	**j, *ñ'j*	**z, *Z?*

If these correspondences are accepted as real, the reconstruction remains problematic. The direction of change indicated in Hayes' paper requires that in AA or pre-AA there was a hardening of affricates to stops, while generally a weakening in AN. The latter is not a problem, but it would be more natural to reconstruct the stop series as original, and posit only a weakening in one branch (AN). This suggests Austric **s, *z, *c, *j, *nc, *nj.*

2) The proposed historical phonology of AA is highly problematic. The proposed voicing distinction between **s* and **z* is, according to my examination of the data, not supported. The reflexes in AA languages are generally /s/ for both, so it appears to be entirely dependent upon AN comparisons, and thus does not belong to the AA level (if at all).

3) The proposed prenasalised series within AA are also not well supported by the data. There is little attempt to distinguish morphological nasals from phonotactically conditioned epenthetic nasals. For example, in Sidwell 2000 I showed that intervocalic prenasalised stops in South Bahnaric languages are simply conditioned variants of plain stops. Also, I know that in many MK languages prenasalisation is frequently derived from affixation by glottal consonants. These kinds of crucial phonological considerations are absent from the paper, which tends to treat surface (phonetic) representations as (morpho-)phonemic. Generally the issue of level of representation is ignored, and this seriously weakens the analysis, independently of the quality of the lexical comparisons supporting the model.

4) Hayes acknowledges much of the methodological difficulties himself, with statements such as: "...it is often difficult to account for all of this contrastive phonological development, even by using the usual qualifier of dialectal divergence." and "...it is often necessary to reconstruct at the PAA level two or more variant proto-forms for certain etymologies." Such comments are welcome; more, they underline the fact that the comparisons do not show the degree of regularity that specialists would like to see in such a discussion.

Allowing any latitude when it comes to regularity of correspondence permits a characteristic inflation in the number of comparisons, and this may be more indicative of a flawed theory, rather than being evidence that we need to posit additional phonemes. Yet it appears that Hayes has taken the latter interpretation.

5) The source of AA glottalisation is not explained by Hayes' model. So far as I can tell, glottalised reflexes appear as apparently unconditioned variants of otherwise plain obstruents. It appears that Hayes has confused 'implosive' and 'glottalised.' While these are distinct articulations, they are often confused in the literature, and this poses problems for scholars working from secondary sources, such as Hayes. In the text he posits a phase in the history of AA with "prenasalised implosives". These are an articulatory impossibility, as one cannot create reduced airpressure in the oral tract if the velum is lowered (as one must to create a nasal sound). Whereas glottalised stops, particularly if voiced, are often automatically prenasalised because glottal tension can condition a lowering of the velum (so-called 'rhinoglottophilia'). However, we again have the problem that we need to distinguish the sources and phonemic status of nasals/prenasalisation.

6) The glossary of comparisons is very problematic—I have two issues with it: a) it is not possible to determine the model of word and syllable structure being used—it appears that any syllable in a multisyllablic word can be compared with any syllable in any other word, regardless of structural position; and b) in many cases the semantics are poorly confined, rendering many comparisons unconvincing. However minor semantic differences may be, I must caution that proving any sound changes requires a solid core of examples with unambiguous semantic unity, in addition to the usual requirement that every segment in the words being compared must be accounted for.

The last point is especially important. It is easy to calculate that if even one segment is disregarded in a word, this multiplies the already high probability of chance resemblances by 5, 10 or even 20 times, depending on the phonemic inventory and phonotaxis. On this basis alone as many as three-fourths of the comparisons presented by Hayes may be unrelated etymologically. Only a much more rigorous investigation will tell us which need to be thrown out, and which should be followed up further.

Response to Sidwell's Comments

by La Vaughn H. Hayes

Dr. Sidwell's observations are generally more to the point than most of the comments offered on the writer's previous paper, viz. "The Austric Denti-alveolar Sibilants," with comments/responses in *Mother Tongue* V:1-40, and that difference in approach is a welcome improvement. However, the writer finds much to disagree with in those observations, and he feels that this disagreement arises primarily because Sidwell, like some of the previous commenters, is laboring under the burden of a conflict between a subjective conception of what the writer's Austric comparison should be and the objective reality of what that comparison is.

The Austric comparison is not a textbook demonstration of the comparative method or a manual on how to reconstruct a proto-language; it is an exploratory and tentative work in progress that is very far from being perfected or finished. To be sure, conventional methodology is used in the comparison, but certain steps in the process have not yet been taken, either because the writer has not yet progressed that far or because he does not feel that the data is yet sufficient to taking those steps. As a result, the comparison contains a number of innate gaps and flaws.

The comparison also does not claim to present a definitive reconstruction of Proto-Austric (PAustric) or any other proto-language, though that objective may one day be attainable. To date, however, it has had only two limited objectives: First, to demonstrate that linguistic evidence does exist to support the Austric hypothesis, and second, to evidentiate the writer's findings about how Austric appears to have evolved into Austroasiatic (AA) and Austronesian (AN), such findings being useful to accomplishment of the first objective. In the latter regard, the affricates paper is another essay to demonstrate a certain set of AA/AN correspondences and to explain as briefly as possible how the writer believes that correlation came about, and that is all that the paper tries to do.

The writer would prefer that his work be evaluated within the parameters of the above definition, but Sidwell and other observers evidently take it for granted that the Austric comparison is a completed, exhaustive exercise in historical reconstruction, which it is not and cannot be until much more work is done. That assumption leads in turn to misunderstanding and unjustified criticism, as well as the sharp differences of opinion seen in these comments and responses.

Sidwell, for example, would like to have at his disposition 1) a larger set of data, 2) a much more rigorous application of the comparative method, 3) the views of specialists in the language families involved, 4) a distinguishing between morphological nasals and phonotactically conditioned epenthetic nasals, 5) an identification of the level of representation used, and 6) a model of the word and syllable structure being used. Now, that is quite a lot to demand of a 23-page paper, the central subject of which is none of the above, so I suppose that Sidwell really means that I should have taken care of those

matters elsewhere. Be that as it may, I shall attempt to address each of his desiderata briefly in the following paragraphs.

The writer finds the demand for a larger set of data a little preposterous, given the difficulty of finding potential AA/AN lexical comparisons and the fact that the writer is the only one who has ever had any success in this regard. The simple truth is that we are lucky to have the 123 lexical comparisons presented in the affricates paper; few others, if any, may ever be found.

A more rigorous application of the comparative method is always a desirable goal, but data availability and quality impose constraints on how far one may go in that application. The Austric comparison is a long-range comparison, after all. The data situation just is not up to the rigor of application Sidwell desires, a fact he and other short-range comparatists do not seem to ever appreciate. Not only is data availability limited, but data quality is also poor in a number of ways. For example, many of the AA/AN lexical comparisons do not involve directly corresponding reflexes of the same PAustric form, but rather indirectly corresponding reflexes of allomorphs of the same PAustric root. These allomorphs apparently differed morphologically, either through affixation or consonantal or vocalic alternation, and under such circumstances, it is impossible to apply the comparative method "much more" rigorously because we do not yet understand just exactly what happened in such cases. One must work with what one has, if one is to do any comparing at all, and hope that more and better answers will be found in the future.

The writer has tried over the years to obtain the views of other specialists in area languages, but the response has been disappointing. Most all of those he has contacted have been unwilling to comment or even engage in discussion, either because they claimed a lack of familiarity with the writer's data or because they view the Austric comparison in the very same terms that Sidwell uses to describe the writer's affricates paper, i.e. "highly speculative and not well supported empirically." That attitude will probably change only after much more data and findings are presented, which is another justification for publication of this paper.

Sidwell's desire for a distinction between nasals derives from his own work on Bahnaric, as he indicates, which has apparently led him to believe that the nasal element in all nasal-oral clusters and prenasalized consonants in Austroasiatic is an environmentally conditioned excrescence and for that reason such clusters and consonants cannot date back to the PAA level. This is also why he argues that the writer's "proposed prenasalized" series is not well supported by the data. The writer agrees that some of the AA nasal-oral clusters and prenasalized consonants are secondary, but disagrees that all of the nasal clusters are. Such clusters are widely distributed in the area, being found in both families (Mon-Khmer and Munda) of Austroasiatic, in Austronesian, and in other area languages, which may/may not be related to Austric. Hence, it appears probable that nasal-oral clusters existed at the PAA level and likely that some of them coalesced into the unitary phonemes described as prenasalized consonants. The crucial question is whether

or not this coalescence dates back to the PAA level, and it cannot be answered definitively at present.

As for levels of representation, the writer uses only two, phonetic and phonemic, and sees no pressing need to conjure up a third, morphophonemic, at this time.
The model Sidwell desires-a description of the canonic structure of Austric and a detailed mapping of how that structure evolved into the AA and AN canonic structures-is totally beyond the scope of the paper being discussed here, but in truth, no such description exists anywhere to the writer's knowledge. The writer addressed PAustric and PAA canonic structure briefly in "Austric I" (Hayes 1992:167-70, 172), and that is all that has been written on the topic.

Using the canonic structure gap as a launch point, Sidwell raises the same complaint that has been voiced by others about Paul K. Benedict's comparison of "split stems" in Austro-Tai, namely that the writer (like Benedict) is comparing any syllable in a multisyllabic word with any syllable in any other word, regardless of structural position, and that is bad procedure and error prone as well. But that is not what either Benedict or the writer has done. The comparison is predicated on partial correspondence of phonological and semantic nature, and if such correspondence does not exist, just any syllables will not do. The choice is never as arbitrary as the complainers would have us believe. The most common canonic form in both Austroasiatic and Austronesian appears to be the disyllabic CVCCVC; hence, *CVCCVC was presumably the most common Austric canonic form. This form can be reduced by syllable loss to either CVC0 or 0CVC, where 0 denotes the missing syllable, and Austroasiatic may retain CVC0 in some cases and 0CVC in others where Austronesian has retained CVCCVC (and vice versa, for Austronesian also evidences canonic reduction). When these forms are compared, the resulting impression may be that one is arbitrarily choosing structurally non-correlating syllables to compare, but in most cases, the AA syllable can correspond to only one of the AN syllables and the phonological details make it very clear which one that is. There is no bad procedure involved here.

Sidwell also complains that this method of comparison does not account for every segment of the words compared, but that is not correct, either. If a segment or syllable is not visible in the comparison, it has been lost, and if it has been lost, it has been replaced by /0/, which is normally left out of the citation. The missing element is thus accounted for; what might not be, are the causes and means of its loss, and those factors are not always discernible. The writer has discussed elsewhere (cf. e.g. Hayes 1992:158) the various conditions (stress shift, etc.) that have led to the canonic reduction in AA whereby initial, medial, and final syllables or segments thereof have been lost, thus necessitating the method of comparison Sidwell finds intolerable. The writer has not yet discussed in print the probability that some of the Austric lexical comparisons do not reflect stem splitting at all, but rather retention of an isolated stem in one language and an ancient composition containing that same stem in the other language, which means that in the case of the retained isolate, nothing has been lost and nothing needs to be accounted for.

On the basis of the alleged lack of accounting for all segments, Sidwell ingeniously extrapolates a prediction that 75% of the comparisons presented by the writer are unlikely to be related etymologically. He adds further that only a much more rigorous investigation will tell us which need to be thrown out and which should be followed up further. The writer can certainly agree with the second statement, but the first is hyperbole. The decision whether or not to reject any given comparison cannot be made solely on the basis of the fact that the number of segments or syllables do not agree, especially in a language family like Austroasiatic, which possesses such descendants as the monosyllabic Vietnamese where all non-stressed syllables have been lost and all clusters reduced to single phonemes.

Sidwell finds fault with the writer's proposal of Austric */s, z, ts, dz, nts, ndz/ and suggests that Austric */s, z, c, j, nc, nj/ should be reconstructed instead on the grounds that a stop system would be more natural. Perhaps so, but that suggestion seems to ignore or evade a salient point in the writer's discussion of the AA and AN palatals, to wit: The occurrence of the irregular palatal/sibilant correspondence clearly indicates that something other than palatal stops were the antecedents of the relevant phonemes.

Sidwell's rejection of the writer's PAA */s/ and */z/ on grounds that the AA data do not support such a voicing distinction reveals that he has not yet had the pleasure of reading all of the writer's publications. In Hayes 1997a:58 and 1997b:15, it is indicated that this ancient voicing distinction was lost everywhere in Austroasiatic except in the Vietic branch of Mon-Khmer where it was retained due to the influence of Middle Chinese. The Vietic data also indicate an old voicing distinction between */ś/ and */ź/, also lost everywhere else in Austroasiatic, but as the writer has shown (Hayes 1997a:57, 1997b:15), this distinction apparently does not date back to the PAA level. To explain the Vietic facts otherwise, one would have to argue that Vietic borrowed the Chinese voiced sibilants and used them to replace AA voiceless ones in random fashion, a development which the writer thinks highly unlikely.

Sidwell observes that the source of AA glottalization is not explained by the writer's model and this is a serious weakness. The writer is somewhat mystified by this statement because he does not use the term "glottalization" anywhere in the paper and the only "model" he refers to is the morphological one presented in Austric I, which does not explicitly account for any phonological facts. Apparently, Sidwell has in mind the writer's attempt to explain the diachronic behavior of the reflexes of the PAA nasal-oral clusters, for he also finds no data support for the writer's proposed prenasalized series and that the writer has apparently confused the terms "implosive" and "glottalized". Whether one or the other term is correct or not, the essential point is that the nasal-oral clusters of Austroasiatic appear to have two sets of reflexes which differ in a way that is difficult to concretize and define. The writer's explanation may or may not be correct; only time and further research will tell.

In conclusion, the writer feels that Sidwell's comments can be regarded as very useful if taken mainly as a recommendation for matters to address in future research or

publications, but as less helpful if viewed solely as a critique of the writer's affricates paper. Like the commenters on the sibilants paper, he focusses mainly on peripheral issues and never quite gets around to addressing the core issue, which in this case is the proposition that a series of nasal-sibilant clusters provided the environment for the development of a series of dental affricates, which in turn shifted to a series of palatal stops in Austroasiatic. As a result, Sidwell's comments do not serve to either affirm or disprove that proposition in any direct way.

Sub-Grammatical Survival: Indo-European s-mobile and its Regeneration in Germanic, by Mark R.V. Southern. (*Journal of Indo-European Studies* No. 34.) Washington, D.C.: Institute for the Study of Man, 1999.

Reviewed by Roger W. Wescott
Late Vice-President of ASLIP

Southern's monograph began as a doctoral dissertation entitled "The Wandering S" (Princeton University, 1997). In it, he reviews scholarly opinions on this movable sibilant expressed by Brugmann, Schrijnen, Siebs, Wood, Sturtevant, Benveniste, and Edgerton. He finds majority opinion favoring the view that this preposed *s* originated primarily as a sandhi phenomenon, carrying over the ergative or nominal masculine noun ending, across a word-boundary, to the onset of the verb that followed it.[1]

Southern himself, however, shows a sufficient sense of intellectual adventure to consider both what Harold Bailey called "z-mobile" (in Greek and Iranian) and the occurrence of preposed *s* in non-Indo-European languages.[2] In Salishan, for example, it has a nominalizing function. In Basque and Burmese, on the other hand, it makes verbs causative. Observing that, because of their stridency, sibilants are always marked speech-sounds (acquired slowly and with difficulty by children), Southern clearly implies that their general phonosemic force, in most languages, is one of intensification. He also recognizes parallelism between the preposed sibilant in Greek (σ)τέγος [(s)tégos] 'roof', and the postposed sibilant in Greek ἀμφί(ς) [amphí(s)] 'about', which, without the *s,* functions as a prefix or preposition but, with the *s,* as an adverb.

Since he treats both phonological and morphological aspects of his topic, I was surprised that Southern ignores the vexed subject of onset phonesthemes, like the *gl-* in *gleam, glow,* and *glitter.* Some linguists treat the consonant cluster in these words as non-morphemic and only phonically suggestive of visual saliency. Others, however, treat it as an asyllabic allomorph of the syllabic morpheme preceding the final obstruent in the word *gold.* When so regarded, it may be said to have the meaning 'to produce or reflect light'.

Because English has been, over two centuries, the most globally influential of Germanic languages, I was also surprised that Southern, with his well-merited focus on Germanic languages, did not treat 20th century English coinages containing prothetic *s.* Among these are popular slang terms, like *spudgy,* meaning 'very pudgy', and technical scientific neologisms like *sparticle,* meaning a subatomic particle so rotated as to become a carrier of force (and so to be classified as a boson rather than as a fermion).

---

1. For example, PIE **pótis pékyeti* > **pótis spékyeti* '(the) husband sees'. The former verb form is preserved in Sanskrit *páśyati* 'sees', the latter in Avestan *spasyeiti,* Latin *specit,* English *spy,* etc.

2. See also the discussion of Dene-Caucasian **s-* (transitive prefix), in *Mother Tongue* V (Bengtson 1999), pp. 175-176.

I would further recommend that any linguist following in Southern's footsteps also consider triphonemic prefixes in contemporary American English slang, such as the first syllables of *spaginzy*, 'Negro', and *skedaddle*, 'depart quickly'. In the first case, the preposed sequence of sibilant + stop + vowel seems to differentiate the word from *guinea*, 'any dark-skinned person'. And, in the second case, the preposed sequence seems to differentiate the word from *daddle*, 'to saunter'.[3]

Overall, however, I would say that Southern has provided Indo-Europeanists and other linguists with an excellent introduction to the enduring problem posed by *s-mobile*, and that *JIES* has done well to publish it.

---

3. See Wescott (1995).

References:

Bengtson, John D. 1999. "In Honor of Jürgen Pinnow: Comparanda Dene-Caucasica." *Mother Tongue* 5: 173-182.

Wescott, Roger W. 1995. "Asyllabic Sibilant Prefixation in English: A Neglected Structural Feature." In *Constructed Languages and Language Construction* (Proceedings of the October 1995 Conference of the American Society of Geolinguistics), ed. by J. Levitt, L.R. Ashley, and K.H. Rogers, pp. 125-133.

Indo-European and Its Closest Relatives: The Eurasiatic Language Family (Volume 1. Grammar), by Joseph H. Greenberg. Stanford, CA: Stanford University Press. 2000. viii + 326 pp.

Reviewed by John D. Bengtson

This is the first of two volumes detailing Greenberg's evidence for the Eurasiatic language family (or "superstock"). As explained by Greenberg (p. 5), the Eurasiatic hypothesis came about as a by-product of his work in classifying the native languages of the Americas and was formulated in complete independence of the largely overlapping Nostratic hypothesis of Russian scholars. The Eurasiatic family was briefly outlined in Greenberg's 1987 book, *Language in the Americas*, and there (p. 332) he promised to present the evidence in a "forthcoming book." This book eventually became two volumes, the first of which is reviewed here. The second volume, with lexical evidence, was completed only months before his recent death and is now being prepared for publication..

The main body of the book consists of three chapters. The first outlines the historical background of the Eurasiatic hypothesis, its similarities to and differences from the various versions of the Nostratic hypothesis, and discussion of other related topics (bilateral comparisons, the "Altaic problem," and the genetic positions of Japanese, Korean, Ainu, Kartvelian, and Etruscan). Greenberg points out that many of the taxonomic differences between his Eurasiatic and the earlier versions of Nostratic (Illich-Svitych, Dolgopolsky, Bomhard) have subsequently been reduced, so that there is increasing convergence among all the versions of Eurasiatic/Nostratic. For example, there is now general agreement that Afroasiatic "is a sister to, rather than a daughter of Nostratic" (p. 6). Greenberg also decries the exclusion of languages or families from a genetic hypothesis simply because they have not been adequately reconstructed, a practice that "leads to the positing of incomplete and erroneously defined families" (p. 10). In the next section of the chapter Greenberg effectively explains why arbitrarily restricted bilateral comparisons "cannot lead to a taxonomy of languages that reflects genuine linguistic history" (p. 11).

On the "Altaic problem," Greenberg concludes that the traditional Altaic family (Turkic + Mongolian + Tungusic) does indeed constitute a valid genetic node within Eurasiatic, and that Japanese, Korean, and Ainu together form another genetic subgroup. As Greenberg admits (pp. 9, 19, *et passim*), he stands alone against the Nostraticists in including Ainu in Eurasiatic. (My own research favors the inclusion of Ainu in Austric [e.g.; Bengtson 1996, Bengtson & Blažek 2000], though this problem is yet to be definitively resolved to the satisfaction of most linguists.) On the position of Kartvelian, Greenberg notes some pronominal resemblances to Eurasiatic, but suggests that Kartvelian is otherwise closer to Afroasiatic than to Eurasiatic. And regarding Etruscan, Greenberg had only recently come to "hesitate between two solutions: (1) Etruscan as a separate branch of Eurasiatic, [or] (2) Etruscan as a third branch of Indo-European," the first of which is tentatively adopted in the Eurasiatic taxonomy delineated on pp. 279-281, which I will briefly summarize as follows:

I †ETRUSCAN
II INDO-EUROPEAN
III URALIC-YUKAGHIR
IV ALTAIC [Turkic + Mongolian + Tungusic]
V KOREAN-JAPANESE-AINU
VI GILYAK [= Nivkh]
VII CHUKOTIAN [= Chukchi-Kamchatkan]
VIII ESKIMO-ALEUT

In chapter 2 Greenberg discusses four aspects of the comparative phonology of Eurasiatic. The first is an alternation of stops (*p, t, k, q*) with their homorganic nasals (*m, n, ŋ, ɴ*), "particularly conspicuous in Eskimo," and Greenberg finds possible traces in Uralic-Yukaghir, Altaic, and Chukotian. (Though Greenberg does not mention it, this kind of alternation also may provide an explanation for the problem of **bh* versus **m* in the Indo-European oblique cases; see pp. 139-147.) The second phonological topic involves a posited Eurasiatic vowel harmony system, best preserved in Tungusic, which has left traces in many other Eurasiatic languages. The third topic is the contrast between simple *r* (r_1) and palatal *r* (r_2), well known in Altaic, and Greenberg thinks this contrast is also present in Chukotian (possibly along with the analogous contrast between l_1 and l_2). Finally, Greenberg postulates that Proto-Indo-European (PIE) final long vowels may derive from Eurasiatic final velars (e.g., Eurasiatic **-ek* > **-eH*$_2$ > PIE **-ā*, or the like).

Chapter 3, "Grammatical Evidence for Eurasiatic," is by far the longest chapter of the book. The format is the same as in his earlier books on language classification: Greenberg lists grammatical morphemes, identified by a simple phonetic 'tag' (e.g., first-person M, demonstrative KU) and traces each formant through the Eurasiatic families where he finds it. Typically, Indo-European is explored first, followed by Uralic, and then the rest, in west-to-east order, though there are exceptions where the evidence is strongest in eastern Eurasiatic.

At first glance, many linguists might not be impressed by the list of grammatical morphemes (pp. 315-317: see Lehmann's article in this issue), particulary since some of the same morphemes occur in other families described by Greenberg. For example, third-person I, demonstrative T, and others are found in Amerind also (Greenberg 1987). Nevertheless a close reading reveals many gems of grammatical concordance that go far beyond a simple equivalence of monophonemic morphemes, particularly when Greenberg shows that strings of two or more separate morphemes have a wide distribution throughout the Eurasiatic family. Below I have extracted a few examples from Greenberg's narrative and arranged them in tabular form: The first example is a combination of second-person T + dual KI(N) that recurs in a form TIK / TEK in widely separated parts of the Eurasiatic domain (pp.71-73, 101-106). In some families the original dual meaning has changed to plural:

Indo-European: Armenian		*d*	*u*	*k'*	'you' (2^{nd} person plural)
Uralic:	PUralic	**-t*	*e*	*k*	'you' (2^{nd} person plural subject)
Chukotian:	Chukchi	*-t*	*ə*	*k*	'you' (2^{nd} person plural subject)
Eskimo-Aleut: Aleut		*-ð*	*i*	*x*	'you(r)' (2^{nd} person dual possessor, intransitive verb subject)
	Eskimo	*-t*	*i*	*k*	'you(r)' (2^{nd} person dual possessive, intransitive verb subject)

It should be noted that the Armenian plural *k'* (or *kh*) is isolated within Indo-European, and Indo-Europeanists have traditionally assumed it was a development of the plural ending **-s*.

The second example is analyzed by Greenberg as a combination of Eurasiatic interrogative K + absolutive NA (pp. 120-123, 217-224). In some languages the -NA element is no longer separable and the word has fused into a single morpheme with the original meaning 'who?'. As seen below, in some languages there has been a secondary development to a particle -KIN with an indefinite or generalizing force:

Indo-European: PIE		**k*^{w}	*e*	*-ne*	(indefinite) as in Sanskrit *kaś-cana* 'anyone'
Uralic:	Finnish	*k*	*e*	*n*	'who?' (archaic)
	Komi	*k*	*i*	*n*	'who?'
Yukaghir:	Tundra	*k*	*i*	*n*	'who?'
Altaic:	Mongolian	*k*	*e*	*n*	'who?'
Gilyak		*-g*	*i*	*n*	(indefinite) as in *aŋ-ha-gin* 'each one, anyone'
Chukotian:	Koryak	*-ŋ*	*ə*	*n*	(generalizing) as in *meki-ŋən* 'whoever'
Eskimo-Aleut: Aleut		*k*	*i*	*n*	'who?'
	Eskimo	*k*	*i*	*-na*	'who?'

The third, and in my opinion the most ingenious, example is the combination of Eurasiatic pronoun base GE + first-person M (pp. 77-81) that accounts for Indo-European **eĝ(h)om* 'I', as well as similar words scattered across Eurasia:

Indo-European: PIE			**e-ĝ(h)*	*o*	*m*	'I' (> Sanskrit *ahám*, Greek ἐγώ(ν), etc.)
Uralic:	Hungarian	en-	*g*	*e*	*m*	'me'
	Vogul	am-	*kk*	*e*	*m*	'I alone'
	Selkup		*š*	*i*	*m*	'I' (< **kim*)
Chukotian:	Chukchi	(-i/-e)-	*ɣ*	*ə*	*m*	'I'
	West Kamchadal		*k*	*ə*	*mma*	'I'
	†South Kamchadal		*k*	*i*	*m*	'I'

In addition to this first-person-singular paradigm, Greenberg posits a parallel second-person-singular pattern, which is not found in IE (Hungarian *té-ged* 'thee', Chukchi *(-i-/-e)-ɣət* 'thou', Eskimo **m-kət* 'I (act on) thee', etc.). Greenberg has even located a parallel to the Indo-European suppletive first-person singular in the far away, now extinct, South Kamchadal dialect (p. 79):

PIE:	**eĝ(h)om*	'I':	**me*	'me'	
†South Kamchadal	*kim*	'I'	:	*ma*	'me'

It should not be surprising that the suppletion persists (or persisted) only in widely separated languages. The leveling force of analogy, over millennia of linguistic change, almost assures the elimination of such grammatical oddities. Indo-European and Chukotian, in their peripheral positions in Eurasia, also agree in the form **tu* for 'thou' (pp. 72-73), which is **te* or **ti* elsewhere:

Indo-European:	PIE	**tu*	'thou' > Latin *tū*, Greek σύ, Gothic *þu*, etc.
Chukotian:	†NE Kamchadal	*tu*	'thou'
	†East Kamchadal	*tue*	'thou'
	Chukchi	*tu-ri*	'you' (pl.)

Greenberg notes some other sporadic retentions, for example the reduplicated word for 'give' (p. 236) in Indo-European (Greek δίδω–μι and Sanskrit *dadā*-mi 'I give', Russian *dadi*-m 'we give') and Yukaghir (Kolyma, Tundra *tadi* 'to give' [third-person receiver]).

Following chapter 3 we find an appendix detailing vowel alternations (*e ~ a, i ~ e, u ~ o)* in Ainu, a classification of Eurasiatic languages (summarized briefly above), and a bibliography (28 pp.) listing the numerous books and articles Greenberg consulted. The book includes three maps by Merritt Ruhlen, who also designed and typeset the volume and assisted Greenberg in many ways.

Greenberg has demonstrated the Eurasiatic family (superstock, macro-phylum) in the time-honored Indo-Europeanist tradition: by showing paradigmatic grammatical patterns for which genetic inheritance, not borrowing or chance resemblance, is the only plausible explanation. As Alfred Kroeber told Edward Sapir (regarding the pronominal prefixes of Algic), "The pronouns turn the trick, alone ..." (See Ruhlen 1994, p. 111.) While one could quibble about certain details (e.g., Armenian plural *-kh,* cited above), there is no doubt that Greenberg's grammatical evidence for Eurasiatic is a monumental achievement and a fitting capstone to his life's work as the supreme linguistic taxonomist of all time.

References:

Bengtson, John D. 1996. "Nihali and Ainu." *Mother Tongue* 2: 51-55.

Bengtson, John D., and Václav Blažek. 2000. "Lexical Parallels Between Ainu and Austric, and Their Implications." *Archiv Orientální* 68.2: 237-258.

Greenberg, Joseph H. 1987. *Language in the Americas.* Stanford: Stanford University Press.

Ruhlen, Merritt. 1994. *On the Origin of Languages: Studies in Linguistic Taxonomy.* Stanford: Stanford University Press.

African Languages: An Introduction, edited by Bernd Heine and Derek Nurse . Cambridge, UK: Cambridge University Press. 2000. ix + 396 pp.

Reviewed by John D. Bengtson

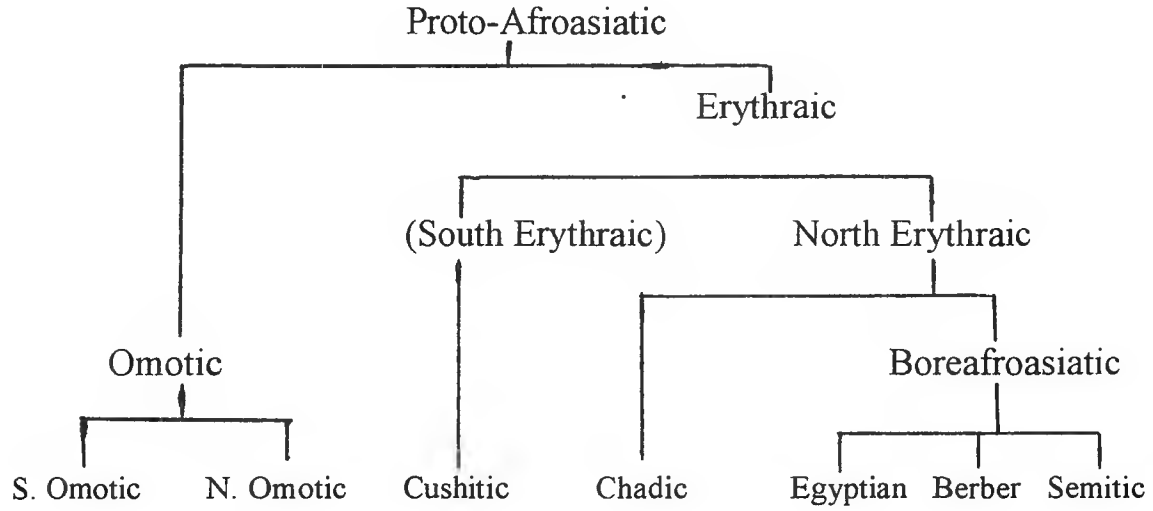
Whether one favors the "Out of Africa" hypothesis (that all biologically and culturally modern humans descend from a group that came out of Africa fairly recently, after 100,000 BP), or the "Multiregional Development" hypothesis (that modern humans evolved convergently in several regions of the world from older hominid lines), or something in between, almost everyone agrees that the human species ultimately originated in Africa. Thus, the origins and classification of African languages is no local concern, but one that should be of vital interest to anyone exploring the deep relationships among the world's languages. In *African Languages* Heine and Nurse cite the figure of 2,035 African languages. This is about 30% of the worldwide total, though Africa's population is only about 11%, and its land area about 20%. As phrased by Jared Diamond (1999, p. 377), "before the arrival of white colonists, Africa already harbored not just blacks but five of the world's six major divisions of humanity, and three of them are confined as natives to Africa. One-quarter of the world's languages are spoken only in Africa. No other continent approaches this human diversity." According to Heine and Nurse (see above), the percentage of languages is even higher: 30%. A similar proportion is found in Grimes (1996): 2011 of 6703 languages.

We now have a new resource, aimed primarily for undergraduate college students, but potentially of use by anyone interested in the subject. This book, edited by Bernd Heine (Universität zu Köln) and Derek Nurse (Memorial University of Newfoundland), is divided into three components. The first consists of articles addressing the four major language phyla of Africa: Niger Congo (by Kay Williamson and Roger Blench), Nilo-Saharan (by Lionel Bender), Afroasiatic (by Richard J. Hayward), and Khoisan (by Tom Güldemann and Rainer Vossen). The second is made up of articles discussing linguistic topics: Phonology (by G.N. Clements), Morphology (by Gerrit J. Dimmendaal), Syntax (by John R. Watters), and Typology (by Denis Creissels). The last section includes articles of more general application: Comparative Linguistics (by Paul Newman), Language and History (by Christopher Ehret), and Language and Society (by H. Ekkehard Wolff).

Given the nature of the readership of *Mother Tongue,* most of this review will be devoted to the classification of African languages, which inevitably involves a discussion of the revolutionary classification by Joseph H. Greenberg and the modifications proposed in the five decades since his initial classifications were published.

Almost everyone would agree with Hayward (p. 74), that "Afroasiatic (AA) is probably the least controversial of the four phyla of languages proposed by Greenberg for the African continent." Of course, as Hayward points out, the existence of this family was widely accepted long before Greenberg, when it was known as "Hamito-Semitic," a term that implied, especially in the 19th Century, a binary structure made up of Semitic and an ill-defined "Hamitic," a grab-bag of the African "poor relations" (p. 83) of Semitic. Greenberg's greatest contributions, as noted by Hayward (p. 85), were the recognition of Chadic as part of Afroasiatic and the exclusion of spurious "Hamitic" languages such as Maasai (now classified as Nilo-Saharan) and Fulfulde (= Fulani, Peul, now included in

Niger-Congo). Another important step was the recognition by Harold C. Fleming that "West Cushitic" was actually a distinct branch of Afroasiatic, now known as Omotic Christopher Ehret's classification of Afroasiatic (shown here and on p. 291, in his article on language and history), is a far cry from the early concept of Semitic as a primary branch opposed to the rest of the (macro-)family:

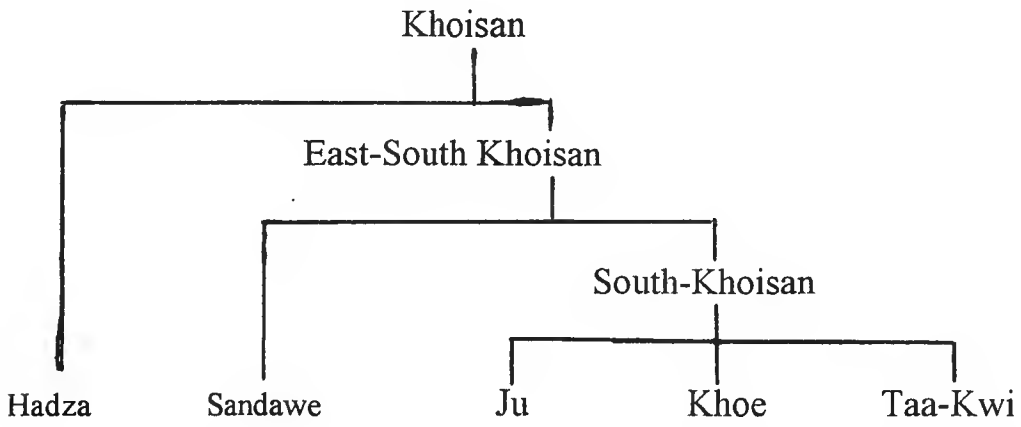


The "demotion" of Semitic to a subgroup of a subgroup of a subgroup of Afroasiatic recalls the demotion of Bantu by Greenberg (see pp. 15-18). (This classification was previously reported in MT by Fleming, 1995.)

Probably the next most accepted of Greenberg's four phyla is the largest phylum in the world with some 1,400 members (p. 11): Niger-Congo (NC, also known as Niger-Kordofanian), even though, as Williamson and Blench admit, "[n]o comprehensive reconstruction has yet been done for the phylum as a whole." They go on to maintain that Niger-Congo specialists nevertheless think that this phylum is indeed a genetic unity, based on shared morphological characteristics (noun class systems, verbal extensions) and shared basic vocabulary. Since Greenberg's original classification of Niger-Congo (1963), there have been several proposals to reshuffle the taxonomic ordering of subgroups, but the most important change has been Thilo Schadeberg's excision of Kadu (Greenberg's "Tumtum") from Niger-Congo and its reclassification as Nilo-Saharan (NS). This may be one of the few times Greenberg was misled by typological or areal similarities. (Cf. his earlier inclusion of Miao-Yao in Sino-Tibetan, a classification he later repudiated.)

The third phylum, in order of general acceptance, is either Khoisan or Nilo-Saharan. In the introduction (p. 5), the editors opine that the "least secure of the four is Khoisan," but Bender (p. 43) makes the claim that NS "is probably the least widely accepted" of Greenberg's four proposed phyla. Even so, Güldemann and Vossen express little confidence in the genetic unity of Khoisan, preferring to "use the term 'Khoisan' as a cover for all non-Bantu as well as non-Cushitic click languages of eastern and southern Africa" (p. 102). This negatively-defined concept recalls the former "Paleo-Asiatic" (or "Paleo-Siberian," still used in this book by Dimmendaal, p. 176), a catchall for East

Asian languages that were not Uralic, Altaic, or Sino-Tibetan. The hesitance of Güldemann and Vossen is echoed by the editors (p. 9): "It may well turn out that Khoisan could be more appropriately defined as a convergence area rather than as a genetic unit." On the other hand, as the authors acknowledge (p. 102), some other scholars do accept the genetic unity of Khoisan. One of these is Ehret, who, later in the book (pp. 289-290) identifies the Proto-Khoisan speakers with the Eastern African Microlithic tradition, ca. 17,000 BP. Ehret's classification of Khoisan is as follows:



Of well-known Khoisan ethnic groups, the !Xũũ (also known as !Kung "bushmen") belong to the Ju family, and the Nama (formerly "Hottentot") belong the Khoe family. This phyletic structure recalls that of the Na-Dene family (Ruhlen 1987: 197-200), in which single languages (in this case, Haida and Tlingit, corresponding structurally to Hadza and Sandawe, respectively) successively "peel off" from the proto-language, leaving a core family with multiple members (here, Eyak-Athapaskan, corresponding structurally to South-Khoisan).

As already mentioned, Bender, the author of the Nilo-Saharan section, considers NS the "least widely accepted" of the four Greenbergian phyla. Even so, this scholar, a well-known skeptic of remote relationships, expresses considerable confidence in the genetic unity of NS, based on "lexical and grammatical morphemes reconstructable to a common ancestor by the comparative method" (p. 60). More specifically, Bender (*ibid.*) refers to "seven major retentions and thirty-nine other retentions which serve to define Nilo-Saharan." So in spite of the "least widely accepted" dictum, Bender clearly affirms, in general, the NS phylum discovered by Greenberg, who, Bender says, "got it right for the most part and his African classification culminating in [*The Languages of Africa* (1963)] is a tremendous advance in African classification" (p. 54). Yet, a great deal remains to be done, as indicated by the radical differences between, *e.g.*, the NS classification schemes of Bender (p. 55) vs. Ehret (p. 274).

One may be (as I was) surprised at the rather wide acceptance of the idea that Niger-Congo and Nilo-Saharan together form an even deeper phylum (Gregersen's "Kongo-Saharan").[1] Bender (p. 57) also thinks that NC and NS are "part of a single larger phylum," but is leery of Blench's proposal (pp. 17, 57) of a "Niger-Saharan" in which all of NC is a subgroup (or subgroup of a subgroup, etc.) of NS. If true, this would be a demotion even more radical than those of Semitic or Bantu, and would result in a

continent with only three distinct major phyla. (As Trombetti said, *"distinto non vuol dire disconnesso."*)

Bender's affirmation of Greenberg's classification resounds throughout the book, *e.g.*, echoed by Williamson and Blench (p. 15): "Greenberg's work [on Niger-Congo] was initially controversial but was gradually accepted by most scholars"; by Hayward (p. 86): "on the basis of 'mass comparison' ... the canon of AA languages was established by Greenberg ... and has, in the present writer's view, come up with the right conclusions"; by Güldemann and Vossen (p. 101): "However controversial Greenberg's hypothesis and classification of macro-Khoisan may be, they surely mark a turning point in the roughly 200-year-old history of Khoisan linguistics insofar as they were established on purely linguistic grounds"; and by Newman (pp. 260, 271): "[Greenberg's] classification ... has served as the point of reference for Africanists for a generation. ... Greenberg's monumental contribution to African historical linguistics is now fifty years old: the field is clearly ready for a new leap forward." Greenberg's impact on African language classification is comparable to the impact of Newton on physics, or of Freud on psychology. In addition, we are constantly reminded of Greenberg's contributions in other areas of linguistics, *e.g.*: phonological universals (p. 130), and morphological typology (pp. 177, 244).

Greenberg's success in classification seems to be taken rather grudgingly by some, *e.g.*, Hayward (pp. 86-87): "Now [Greenberg's classification] was on the basis of 'mass comparison', rather than the comparative method ... [and] a methodology that does not invoke the rigour of the principle [of regular phonological correspondence] cannot make predictions, and so falls short of true theoretical status." One wonders how regular phonological correspondences could have been detected and tabulated *in advance* of knowing which languages to compare! This statement is typical of an all too common misconception among historical linguists: the idea that 'mass comparison' is somehow an inferior method of linguistic classification, and that Greenberg "should have" used 'the comparative method' instead.

Fortunately, *African Languages: An Introduction* provides a refreshing perspective that is by itself well worth the price of the book: Paul Newman's article on comparative linguistics. Wisely, he divides the subject into four logically sequential areas: "(1) classification and subclassification; (2) reconstruction; (3) establishment of sound laws; and (4) treatment of loan words."

Newman begins the section on classification (p. 260) by enumerating five of Greenberg's principles and guidelines of linguistic classification, the first of which is that "[l]anguage classification must be based on linguistic evidence alone and not on racial or cultural criteria." As Newman points out, it seems obvious to us today, but this was not so before Greenberg made the principle explicit. This principle is just as valid in areas of the world other than Africa, for example, Eurasia, where the cumulative efforts of paleolinguists have identified three major linguistic streams (the following chains in roughly north-to-south and west-to-east order):

(A) INDO-EUROPEAN	URALIC	ALTAIC	CHUKCHI-KAMCHAT-KAN	ESKIMO-ALEUT
KARTVELIAN	Yukaghir	Nivkh		
	Korean	Japanese		

(B) Basque	(NORTH-) CAUCASIAN	Burushaski	SINO-TIBETAN		YENISEIAN	NA-DENE
(C) Nihali	MUNDA-MON-KHMER		MIAO-YIAO	KADAI	AUSTRONESIAN	Ainu

Many readers of *Mother Tongue* will easily see that series (A) corresponds to Nostratic (and Greenberg's Eurasiatic), (B) is Dene-Caucasian, and (C) is Austric. One notes that while the proposed linguistic connections are primarily horizontal, cultural and racial similarities often run vertically. For example, Basque and (Western) Indo-European peoples are more similar, culturally and racially, to each other than either is to the peoples farther east who are closer linguistically. In the same way , the Sino-Tibetan peoples have more in common, racially and culturally, with the Miao-Yao and Kadai peoples than with the Caucasians, Yeniseians, etc., leading to the former "Indo-Chinese" family that united Sino-Tibetan, Miao-Yao, and Kadai. The application of the "linguistic evidence only" principle by Paul K. Benedict (see Ruhlen 1987: 144, 152) led to the separation of Kadai and Miao-Yao from Sino-Tibetan, but this distinction had as much to do with the second principle (p. 260), that "classification must be based on *specific* points of resemblance and not on the presence or absence of general features of a typological nature." To Benedict, the fact that Chinese, Hmong, and Thai were all tonal – a typological feature – was not enough to put them in the same family. His demand for *specific* points of resemblance (mainly basic vocabulary) led to the recognition that these three languages had distinct origins.

Newman (p. 261) reiterates the other principles of classification (the rule of transitivity, and the rule that vocabulary and grammar lead to the same results), but most important, Newman recognizes that "it is *not* necessary that the linguist 'prove' that the classification is certain by the presentation of conclusive evidence." The concept of 'proof' in genetic linguistics was discussed at length in the first issue of this journal by Greenberg himself (1995). There Greenberg (1995: 213) quoted yet another statement by Newman:

> The proof of genetic relationship does not depend on the demonstration of sound laws. Rather, the discovery of sound laws and the reconstruction of linguistic history normally emerge from careful comparison of languages known to be related.

Thus, Newman and Greenberg essentially agree on the logical sequence of (1) classification (carried out explicitly or implicitly by the application of mass comparison), followed by (2) reconstruction and establishment of sound laws (= recurrent phonological correspondences), and in fact this has been the process by which every generally accepted language family has been recognized and accepted by scholars, beginning with the discoveries of Finno-Ugric and Indo-European (Greenberg 1995: 208-209). In this light, it seems to me that Hayward's statement (quoted above) that Greenberg's "methodology ... falls short of true theoretical status" itself falls short of validity. Newman (p. 261) correctly concludes that "[t]he job of the comparative linguist is to provide the best explanation possible consistent with the facts." The fact that Greenberg's classification (with more or less minor adjustments) has held up for fifty years suggests that his methodology is indeed valid, and should be regarded as the standard methodology of linguistic classification.

Among defects of this book, one could mention the confusion of nomenclature, which of course did not originate with this book, but is a long tradition in African linguistics. A salient example is the language generally denoted as *Fulfulde* (Niger Congo, Atlantic) in this book (e.g., p. 21). Because of their nomadic tendencies, the Fulfulde live in several countries in western and northern Africa, from Senegal to Sudan, and are known under a multitude of names, including Ful, Fula, Fulani, Fulbe, Peul, Peuhl, Peulh, Pulaar, etc. (Its erroneous classification as a "Hamitic" language, corrected by Greenberg, was mentioned above.) This book partially solves the problem of nomenclature by cross-references in the index (e.g., Pulaar *see* Fulfulde), but the use of other reference books, such as Grimes (1996) is still necessary in sorting out the confusion. The cross-referencing is not pervasive: for example, on page 315 a "Bamanan" language is mentioned, but the student has no way of knowing that it is synonymous with the language otherwise known as "Bambara" (Niger-Congo, Mande) on page 20. The Ethiopian liturgical language *Geʕez* (p. 321) is also denoted as *Gi'iz, Ge'ez,* and *Gə'əz* elsewhere in the book. To his credit, Bender devotes a portion of his article (pp. 46-48) to the clarification of nomenclature.

I was puzzled by the absence of the Kordofanian (sub-)family on both maps (pp. 2 and 12) where one would expect to see it located. A possible explanation might be their recent displacement from the Nuba Mountains of Sudan, as mentioned on page 17, but even so, their historic location should have been acknowledged, as is done for the extinct Egyptian (p. 75).

Some misnomers and typographical errors: a contrast between "bilabial and labio-velar fricatives" is mentioned (p. 127), where surely the latter should be "labio-dental." In the first paragraph on page 49, cultural groups are described both as "pastoralists" and "pasturalists," where only the former is correct. [Knut] "Bergslund" (pp. 296, 351) is actually *Bergsland.* The Silt'e language is cited as appearing on page 214 (index, p. 386), but the correct page number is 224.

These minor defects aside, *African Languages* is a valuable new resource for students at all levels.

1. In one of my notebooks from around 1970, I put together the following "cognate sets" common to Greenberg's Congo-Kordofanian and Nilo-Saharan, which seemed to me to indicate a deeper super-phylum:

	Niger-Congo:	**Nilo-Saharan:**
'break':	Soninke *kara* Utoro *geri*	Gao *keyri,* Berta *gula,* Maasai *gɪl*
'buy':	Yoruba *ra* Katla *la*	Fur *ulu,* Nandi *al*
'dry':	Bantu **kakɥ*	Zerma *kogu,* Zagawa *koʔá,* Muguzi *k'waak'a,* Baka *gágá*

'fall':	Kpelle *to* Twi *tɔ*	Kanuri *du-*, Teso *ado*, Madi *dɛ*
'give':	Ewe, Tiv *na*	Zerma *no*, Didinga *nya*, Maba *ɲi-*, Mesalit *ɲú*, etc.
'go':	Fanti *kɔ́*	Zerma *koi*, Zagawa *ke-*, Kunama *ga-*, Tirma *oko*
'large':	Kpelle *bayi* Ebang *ipa*	Fur *apa*, Teda *bo*, Murle *äpi*
'small':	Kpelle *doŋ* Zande *toni*	Fur *itiŋ*, Debri wa-*tono*, Shilluk *ten*, Jur *θiin*
'tooth':	Fula *ɲi:*-re Tegali *ɲin*	Didinga *nigitat*, Nera *nihi*
'urine':	Balanta *sara* Dyan *siri*	Kanuri *collo*, Berta *sara*
'white':	Diola *fur*, etc.	Mabang *fafara*, etc.

Of the above comparisons, five ('break, buy, give, small, white') coincide with those of Gregersen (1972).

References:

Diamond, Jared. 1999. *Guns, Germs, and Steel: The Fates of Human Societies.* New York / London: W.W. Norton & Co.

Fleming, Harold C. 1995. "Editorial II: Afrasian Gets Better, Better, and Deeper." *Mother Tongue* 1: 233-234.

Greenberg, Joseph H. 1963. *The Languages of Africa.* Bloomington, IN: Indiana University Center in Anthropology, Folklore, and Linguistics.

---------- 1995. "The Concept of Proof in Genetic Linguistics." *Mother Tongue* 1: 207-216.

Gregersen, Edgar A. 1972. "Kongo-Saharan." *Journal of African Languages* 2.1: 69-89.

Grimes, Barbara F. (Ed.) 1996. *Ethnologue* (13th Edition). Dallas: Summer Institute of Linguistics.

Ruhlen, Merritt. 1987. *A Guide to the World's Languages. Volume 1: Classification.* Stanford: Stanford University Press.

The Nostratic Macrofamily and Linguistic Palaeontology, by Aharon Dolgopolsky. (Papers in the Prehistory of Languages, 1.) Cambridge, UK: The McDonald Institute for Archaeological Research, 1998. xxii + 116 pp.

Reviewed by Vitaly Shevoroshkin
University of Michigan

[Editor's Note: This and the following review have been minimally edited, preserving much of the distinctive style of Professor Shevoroshkin, familiar to readers of *Mother Tongue* (newsletter and journal). Many bibliographical notes refer to abbreviations, e.g., *APPJ, OSNJ*, which are explained in the list of references following each review.]

The book we are discussing here (generally referred to as *NM*) reflects materials prepared by Aharon Dolgopolsky (Dolg) for his mini-seminar on Nostratic paleontology (Cambridge, England, 1997). This seminar was sponsored by the McDonald Institute for archaeological research (director - Prof. Lord Colin Renfrew); the McDonald Institute (along with the US-based A. P. Sloan Foundation) also provided funds for publication of *Nostratic: Examining a Linguistic Macrofamily (NELM)*, for the Symposium on the Nostratic Macrofamily (Cambridge, 17th-18th July 1998) based on *NM*, for the publication of the symposium materials in the 1999 book *NELM*, and for the publication of Dolgopolsky's *Nostratic Dictionary (ND)*: with over 2300 entries, accomplished in the year 2000, forthcoming in Cambridge as well. At the moment, the work on compiling indexes is being done, which will require a few more months.

The main bulk of *NM* is represented by 124 entries chosen by the author from *ND* to illustrate his perception of how the speakers of Nostratic (N) may have lived many millennia ago. The book contains appropriate introductions (by Renfrew and Dolgopolsky) and comments by Dolg. on the 124 entries, which were divided into certain groups: Where and when (the N people lived); Hunter-Gatherers; Food; Technological Activities; Anatomy; Kinship; The Realm of the Supernatural. Dolg. tried to show in his book that the N people lived in pre-neolithic times somewhere in SW Asia (apparently in the Near East).

Each entry represents a reconstruction of a certain root/word of the N language (spoken many millennia ago) with its meaning; accordingly, each entry also contains genetically related words - as a rule, reconstructed roots supported by lexical evidence from appropriate languages - in daughter languages of N: Hamito-Semitic (HS = Afroasiatic), Kartvelian (K), Indo-European (IE) in the West, and Uralic (U), Altaic (A), Dravidian (D) in the East. (Each of these languages became the ancestor of a language family. There is no bibliography to indicate sources of the lexical forms used in the above entries (such bibliography is, naturally, present in the *Nostratic Dictionary*).

Most participants of the symposium considered *NM* a work whose main aim was to prove that N languages can be compared in the same way as IE or Uralic, and that a N proto-language can be reconstructed. This was not Dolgopolsky's aim, as it is clear from the title of his book; therefore *NM* does not contain any explicit proof of the validity of N. Naturally, *NM* contains many words which denoted animals and plants: such words easily become borrowings. All this led several colleagues to a conclusion that the material presented in *NM* doesn't prove that the N

theory works, and that N comparisons don't follow strict rules, comparable to those in the IE or U historical linguistics.

It is amazing that most participants of the symposium who were not Nostraticists seemed never to try and check if Dolgopolsky's phonetic charts work (these charts are present in both *NM* and *NELM*; see my review of *NELM* below). Checking the charts wouldn't take much time but would provide a quick and decisive answer to an important question: Does the Nostratic theory work?

Kamil Zvelebil, a prominent specialist in D languages, who is sympathetic to Nostraticists' work (see his paper in *NELM*: 361) writes as follows (*ibid.*: 364):

> "When I see the N consonant chart, the reconstructed phonemic inventory (*NM*, 101) and as I try to find my way through the 'transcription signs and other symbols' on pp. 11-16 of the work under review, I am reminded ... of what Emperor Joseph II said to Mozart in Forman's movie *Amadeus* ... : *'Too many notes!'* "

I certainly agree that the number of signs used by Dolg. is excessive (Illich-Svitych used much fewer - but all was clear. As for the uncertainty signs, both [ ] and ( ) would suffice: cf. Starostin's use of these signs in his paper in *NELM*). But as for the phonetic charts, almost all signs used there are both clear and necessary. Nevertheless, the charts weren't consulted by many of those who were busy writing their reviews of *NM*.

Note what Zvelebil asks on p. 362 of his paper where he discusses words designating footwear (IE **krep-*, Sem. **krp*, D **kerVppV-*. K.Z. approves of this set): "However, even this item, this reconstruction is not quite unproblematic (*why the initial glottalization?*)". He means glottalized *K'* (= *k'* or *q'*) in *NM* 100 **K'VRVHp'/pV*: the N word from which all three above forms have originated.

Zvelebil implies here that the initial *K'-* in the N reconstruction wrongly shows glottalization because none of the daughter languages shows it. But one of the fundamental rules of N comparison states that IE voiceless consonants regularly originate from glottalized N consonants (there are a few clear exceptions).

It would take precisely two minutes to find in the charts the following line, which contains initial *k-* in all three languages: IE, Sem., and D (cf. the above forms). This line shows that S, IE, D *k-* originate from N *k'-*

N	...	S	...	IE	...	D
**k'-*		**k', *k*		**k, *k̂, *k^w*		**k*

This line not only explains "why the initial glottalization" but it also shows that Dolg. could be more precise in his reconstruction of the initial N consonant in the root in question: this should be *k'-* and not an uncertain *K'-*, which means "either **k'-* or **q'-*." But *q'-* cannot be reconstructed here because *q'-* yields either *k'-* or *x-* in Sem., and not *k-* which we have:

N	S	IE	D
**q'-*	**k', *x*	**k, *k̂, *k^w*	**k*

We may go on checking if Dolg.'s N reconstructions, as presented in *NM*, really work: for that, we may choose any N sounds from the phonetic charts and check if these sounds are, indeed, reflected in words from daughter languages precisely in the way they are presented in the charts (if some languages are not represented in the charts, such as Cushitic or Chadic, or proto-Altaic, then the needed words can be found in the appropriate *NM* entries. I have also added one example from Dolg.'s

paper on N phonology in *Reconstructing Languages and Cultures (RLC)* [see bibl.]: it is indicated by *R*).

Let us start with some "unusual" N sounds that yield phonetically very different consonants in the daughter languages, such as intervocalic lateral obstruents *-ẑ-* (fricative) and *-ʒ̂-* (affricate).

According to Dolg's charts and *NM* data, both these sounds are reflected as *ẑ* in Uralic (including FU); as *ŝ* in Semitic; as *l* in Cushitic (see appropriate *NM* entries), IE and A (Turkic [T], Mongolic[M], Tungus-Manchu [TM] = Tungusic).

In the following table, six N reconstructions, containing the above consonants are presented: five from *NM* (50 'fur-bearing animal'; 60 'berry'; 99 'to skin, to [remove] bark'; 90 'thorn, hook (< tooth)'; 21 'tasty beverage', and one from *R* ('to glow, burn').

It is easy to conclude that all correspondences match. Even if we drop all questioned examples, the conclusion will be the same: Nostratic rules work properly:

N	(HS:) Sem.	(HS:) Cush.	IE	U: FU	A: T, M, TM
-ẑ-, -ʒ̂-	*ŝ*	*l*	*l*	*ẑ*	*l*
50 *bUyẑV*			*bhel-*	*poyẑV*	M *bulugan*
60 *moyẑV*	?? S *mŝ-mŝ*		?? *māl-*	FU *moẑV*	TM *mi/olV-*
99 *K'oẑV*	*k'ŝw*			*koẑV-*	M *qol-tu-*
90 *?eẑekU*	(?) **?iŝikk-*	ECu. *?ilik-*			*ēlku* :M *elgü-*
R *śüẑVxV*			*swel-*	*śüẑe*	TM *sulu-*
21 *mayʒ̂V*		ECu. *mal(?)-*	*mel-*	FU *mayẑV*	T *bal*

Among initial consonants in the above exx. we have N *K'*, which means "either *k'-* or *q'-*" (the latter is preserved only in Kartv.); it mostly remains in Sem. as *k'-* but becomes *k-/ǩ-/kʷ* in IE and *k-* in U.

The following table represents 2 segments of the charts which show what N *k'-* and *q'-* yield in the daughter languages, namely in (HS:) Semitic, Kartv. [sometimes represented by modern Georgian], IE, U [or only in FU], A (or only in Turkic):

N	(HS:) Sem.	K [=Georg.]	IE	U [=FU]	A (:Turk.)
k'-	*k', k*	*k'*	*k, ǩ, kʷ*	*k*	*k'* › T *k', ǩ*
q'-	*k', x*	*q'*	*k, ǩ, kʷ*	*k*	*k'* › T *k', ǩ*

As we see, there are three possible reflexes of N *k', q'* (both initial and non-initial) in IE. Usually, IE *k* appears in the N roots which contain the sequence *k'/q'a*; IE *ǩ* appears in roots which contain N *k'/q'* before a front vowel (*i, e, ä*); IE *kʷ* appeared in roots which contain N *k'/q'* before a labial vowel (*u, ü, o*) [For details concerning this distribution of N consonants in the daughter languages see V. Dybo in *RLC*].

The following table contains N words in *k'-, q'-* (or *K'-*, as explained above) and their reflexes in appropriate daughter languages (all data according to *NM*).

I am using the following *NM* entries: 73 '(large) fish'; 10 'ice (etc.)'; 61 'fruit of a leguminous plant'; 98 '(some) bark, skin'; 93 'bark'; 91 'tooth, claw; hook'; 23 'basket' [apparently N with *-ć-* as shown by Kartv.]; 99 'to skin, to [remove] bark' (we already had this word); 67 'intestines'.

We may now check sound correspondences, according to the phonetic charts.

[There are two problems which concern vowels: U **kala* shows *-a-* where we expect *-o-* (as in N **K'olV*); but U **kopa* shows *-o-* where we expect *-a-* (as in N **K'ap?E*). These problems still remain unresolved].

N	(HS:) Sem.	K (:Georg.)	IE	U (:FU)	A (:Turk.)
73 *K'olV*	S:Jb. *kāl*		*kʷo/al-*	*kala*	*k'olV*
10 *k'iruqa*	S *k'arVx*	G *k'rux-wa*	*ǩer(H)-*	FU *kirV*	T *k'ira-gu*
61 *K'ERV*	S *k'VrV-θ'-*		*ǩi-ǩer-*		
98 *k'oRup'V*	S *k'rp*	G *k'orp'-*			
93 *K'ap?/ɕE*		G *k'ep-*		*kopa*	T *k'āpuk*
91 *k'ak'wV*	S *kakk-*	*k'ak'w-*		*kokka*	
23 *k'o?ć/cV*	S *ka?as-*	*k'ec-*	*kʷes-yā*	FU *koća*	
99 *K'oẑV*	S *k'ŝw*			*koẑV-*	M *qol-tu-*
67 *q'u3V*		*q'wi3-l-*	*kews-/kus-*		

Compare now our words with the following table, which shows reflexes of the appropriate N non-initial consonants *l, r* and *R* (= *r* or *ŕ*), *q, p', p, k', c, ć, ẑ, 3* in our words in five languages (= language groups) represented in the above exx (I didn't use D: its material is scanty); all data match here as well:

N	Sem.	K (:Georg.)	IE	U (:FU)	A (:Turk.)
-l-	*l*	*l*	*l*	*l*	*l*
-r-	*r*	*r*	*r*	*r*	*r*
-ŕ-	*r*	*r*	*r*	*r*	T *ŕ*
-q-	*x*	*-q-* › *x*	*H*	-	-
-p'-	*p*	*p, p'*	*p*	*pp*	*p* (M *b*)
-p-	*p*	*p*	*p, b*	*p*	*p* (M *b*)
-k'-	*k'*	*k'*	*k, ǩ, kʷ-*	*kk*	T *k*
-c-	*s*	*ċ*	*s*	*ć*	
-ć-	*s*	*c*	*s*	*ć*	
-ẑ-	*ŝ*	*l* (?)	*l*	*ẑ*	
-3-	*z*	*3*	*s*	*ć* (?)	

Let us now take initial *k-* and *g-* in N with its reflexes in daughter languages: I am listing 5 N words below: *NM* #109 'a woman from the other moiety'; 110 'a man from the other moiety'; 83 'stone'; 31 '(follow the) track'; 106 'occiput; hind part'.

N	Sem. (etc.)	K	IE	U (:FU)	(A:) Turkic
k-	*k*	*k*	*g, ǵ, gʷ*	*k*	T *k̄*
g-	*g*	*g*	*gh, ǵh, gʷh*	*k*	T *k̄, k'*
109 *kälu/ü*	*kall-at-*	*kal-*	*glōw-*	*kälü*	T *k̄äl-i-n*
110 *küda*		**kwid-iš-* (?)		*küδü*	T *k̄üδä-*
83 *kiwVɦE*	Gz *kʷa-kʷəɦ*	*kwa-*		FU *kiwe*	
31 *goki*	Chad. *gk*			*koke*	T *KogV*
106 *gedi*	Cu:Kh. *gĭd*				T *KEδi-n*

As expected, initial consonants in the above words match those segments of Dolg.'s phonetic charts which reflect initial N *k-*, *g-* in HS, K, IE, U, and A (:Turkic) languages (now Altaic can be represented more accurately and simply, but we follow Dolg.'s table which is, in principle, correct). Note that IE **ǵlōw-* may have originated from ***ǵelow-*: *ǵ-* presupposes a sequence **ǵe-*, not **ǵa-* or **ǵo-*):

The above five words reflect N non-initial consonants *l*, *d* (twice), *w*, *ḫ*, *k*. According to Dolg.'s charts, all fits well in these, seemingly old, roots (correct a typo in the charts: IE *d* from N *d*; this should be IE *dh*, of course):

N	Sem., etc	K (:Georg.)	IE	U (:FU)	A: Turkic
-l-	*l*	*l*	*l*	*l*	*l*
-d-	*d*	*d*	*dh*	*δ*	*δ*
-w-	*w*	*w*	*w*	*w*	(*w* ›) *b*
-ḫ-	*ḫ*	-	*H*	-	
-k-	*k*	*k*	*g, ǵ, g*^{w}	*k*	T *g, k*

I would like to underline that N linguistics has existed in its present form (comparison of six language families and reconstruction of several hundred N roots) for four decades. The first publications in modern Nostratic [papers by Dolg. and Illich-Svitych (IS)] appeared in 1964. Nostraticists use the same strict methodology as, say, researchers working in IE historical linguistics. N linguistics is built upon a well-researched corpus of appropriate daughter languages. These daughter languages still undergo intensive research by historical linguists, including Nostraticists. Accordingly, existing roots may be somewhat changed during the ongoing studies; this is inevitable.

N linguistics deals both with stable lexicon and with grammar (both are reconstructed on the basis of N daughter languages).

Though Dolg. started his work in the field of Nostratic linguistics much earlier than IS, it is the latter who is considered by scholars, including Dolg. (see his foreword to *NM*), as the founder of modern Nostratics. Along with V. Dybo and R. Bulatova, Dolg. spent an enormous amount of time and energy in arranging and editing IS's posthumous work in the years 1966 - 1976 (in 1976 he emigrated to Haifa, Israel).

IS was a first-class expert in all 6 language families he considered genetically related; some of his work was published (Dogopolsky didn't consider D a N language in his early research). IS started to work on the reconstruction of N grammar; N lexicon was meant to form some kind of a supplement to this grammar. He was also reconstructing roots (whenever it was necessary) on the level of families (A, K, D, etc.) before turning to proto-N lexical reconstruction. (Other scholars, Dolg. among them, didn't reconstruct N roots at that time). IS possessed a rare intuition which allowed him to immediately recognize the needed item in a broad and seemingly chaotic language inventory.

A well-known Uralicist, Eugene Helimski, wrote in the English version of his long paper *Illich-Svitych's Work and the Development of Nostratic Research Abroad* (the Russian original appeared in Moscow in 1986), 1988:

> "As early as 1965, the patriarch of IE-U, U-A, and U-Yukaghir comparison, B. Collinder (Sweden), saw "decisive success in the area of IE-U-A linguistic comparison" in IS's paper on the genesis of IE gutturals. ... In the opinion of the prominent Altaist, N. Poppe (USA), the work of IS is the 'most extensive, and without a doubt the most well-founded and convincing, work on N problems.' ... 'The results achieved, the convincing survey of the comparative and historical phonetics of the N languages, and an etymological dictionary ... are scholarly facts which no one who studies the

comparison of languages beyond the bound of one family can disregard,' writes R. Eckert (Germany). ... Danish Indo-Europeanist J. Rasmussen connects his own appraisal of IS's first publications with the hope that pre-IE will finally be open to normal comparative reconstruction which will come to supplant speculative glottogonic constructions.' "

R. L. Trask writes as follows in his *Historical Linguistics* (1996: 390):

> "How successfully have the proponents of the super-families made their case? There is no doubt at all that N represents ... the most plausible of all these proposals. ... [T]he defenders of N, almost uniquely among the proponents of super-families, have at least attempted to stick rigorously to conventional historical methodology and conventional standards of evidence. Rejecting all appeals to casual resemblances, no matter how numerous, they accept as data only the best available reconstructions for items in the six proto-languages which they regard as daughter of PN ...; and, most importantly, they insist upon the identification of systematic correspondences and of regular phonological developments in all branches. Even so, they still managed to come up with some hundreds of PN roots and affixes. Their work clearly deserves to be taken seriously."

Now I would like to return to the discussion of the life and time of the people who spoke N in prehistory. As early as the 1960's, Dolg. started to publish popular papers on Nostratic that contained more or less the same conclusions about speakers of the N language as his 1998 book. Naturally, at that time the number of illustrative exx. was not as large. Dolg. believed the speakers of N were hunters and gatherers, partly because there were no N terms describing agriculture or cattle-breeding, although there were designations of edible cereals, berries, fruits, and certain tools such as fishing nets, etc.

The question about the age of the N language was first discussed in print as early as 1963 by A. Lamprecht and M. Čejka. As E. Helimsky writes (*ibid.*):

> "[T]hey made an attempt to apply glottochronology to the dating of the 'IE-A' epoch ... Their conclusion was that it is possible to date this epoch 'to the 10th-9th m. BC,' that is, to the time that directly followed the last glacial period."

Both IS and V. Dybo came to the same conclusion four decades ago, and S. Starostin (St) now has a similar perception. In St.'s paper published recently in *HLL*, proto-IE is regarded as being 6 to 7 millennia (m.) old, - not 9 m. as C. Renfrew seems to think. Accordingly, Dolg.'s dating of N (about 17 m.) doesn't seem realistic.

St. thinks that both IE and K are some 6 m. old; U and A – 8 m.; D (which first split from N) – 9m. As for HS (12 to 11 m. old), it was not a daughter, but a sister of N, the third sister being Sino-Caucasian (SC = Dene-Caucasian): see St. in *NELM*.

Note that D seems to show [according to G. Starostin in *MLŽ* vol. 1] remnants of palatal and voiced stops; this connects D with Western languages: IE-K-HS. We may compare this situation with what we have in proto-Anatolian *vs.* "narrow/Western IE": The pAnat. language (which rather early split from proto-IE) shows reflexes of all three IE stops of the type *k*, *ḱ*, *kʷ*, whereas Western IE languages show reflexes either of *k*, *ḱ* [having changed *kʷ* to *k*] or of *k*, *kʷ* [having changed *ḱ* to *k*].

Apparently at some point in time, at least three proto-proto-languages (N, HS, Sino-Caucasian) were spoken in the Near East before they split into daughter languages.

If there are indeed five, and not six, N daughter languages, the reconstruction of N will provide somewhat simplified phonetic and lexico-grammatical systems (which may be quite welcome – the existing reconstruction of at least 50 N consonants

seems too complicated. On the other hand, SC/ North Caucasian (NC) phonetics show even more sounds). Besides, connections between HS (now a separate phylum) and the reduced N phylum (now K-IE-U-A-D) makes much more suspicious those cognates which show only two roots: HS and, say, only K (or only IE; or only U; or only A; or only D). Indeed, now such pairs of roots cannot even be considered legitimate cognates: at least two daughter languages are required to produce a N reconstruction – which then can be compared with HS (or SC, for that matter). So, if we have a HS root (representing a phylum), we need at least a pair of cognates on the level of N daughter families, – to form a legitimate representation of the N phylum. So, we need HS *vs.* (at least) K and IE; or K and U; or IE and U, etc.

Dolg. provides 30 sets (## 1-30) to support his theory that the homeland of the N people was located in SW Asia in the pre-Neolithic. This seems correct, but as for words with the meaning 'fig tree' (#1), 'hyena' (#2), 'leopard' (##3-4), 'monkey' (#6), 'saline earth, desert' (#11), they seem either not to exist (see J-1, J-2, J-6 below) or to have a somewhat different meaning (not 'saline earth, desert' but 'soil, clay': G-11; not 'leopard': E-4; note also G-36 about **gurHa*: not 'antelope'), or might be borrowings (H-3, H-24); cf. also Vovin (*NELM*). - Dolg.'s #54 'fig tree' may have designated some kind of berries (C-54). - All this seems to suggest that the climate at that time was milder (as scholars have frequently mentioned).

I disagree with Dolg. when he says (p. 26) that N "has words for harvesting (in defiance of the famous maxim)": #15 **qaRp/p'V* didn't mean 'to harvest' (Hitt. shows 'gather; a heap', related to 'reap' in IE); it rather meant 'gather': see B-15.

Words ## 5, 8, 20, 23, and 29 seem to represent correct reconstructions (see A). There are minor problems with ## 10, 12, 27, and 30 (see B). Additional words can be listed for sets ## 13 and 19 (see C). Sets ## 4 'leopard', 9 'hoar-frost', 21 'tasty beverage' and require substantial restructuring (see E). Set #14 'body of water' seems to represent only a part of a reconstruction that includes additional words from several languages; the meaning of the N root may have been 'water': see F-14.

Two different roots/words (either on the N level or on the level of a daughter language) may be identified in each of the following sets: ## 11 ('soil, clay' *vs.* 'stony/shallow place'); 22 ('wicker, wattle' *vs.* 'to mix'); 25 ('bend, bow' *vs.* 'to bind; rope'); 28 ('sinew; to tie' *vs.* 'arrow'). Three roots can be identified in the set 26 ('bow, arrow' *vs.* 'bow' *vs.* 'lead, pull'): see G.

It is possible that reconstructions ## 3, 17, 18, 24 represent borrowings. Reconstructions ## 1, 2, 6, 16 seem incorrect.

* * *

Sets ## 31-62 include words which show, as Dolg. states on pp. 38 and 50, that the N people were hunters-gatherers. From the above sets, eight (all referring to hunting) seem quite correct: ## 31 '(follow the) track'; 34 'hit (the target)'; 35 'miss one's aim'; animals' designations: 41, 47, 48, 50, 52 (see A). To the HS-D set # 54 an A root can be added (see C). – Neither HS nor IE words seem to match U-A in # 60 (some berries): see D. A restructuring is needed in the set 51 'squirrel' (see E).

Three reconstructions seem to represent two different roots each: ## 33 ('sharp edge, spear' *vs.* 'spear, arrow'), 36 ('deer, wild animal' *vs.* 'male'); 62 ('root, sinew' *vs.* 'tendon'). Reconstr. # 42 can be split into three roots ('wild sheep, goats' *vs.* "wound; hunt, kill' *vs.* 'herd of horses'): see G.

It is not excluded that some of the sets 39, 40, 43, 45, 46, 49, 53, 57, 61(words for plants and animals) represent borrowing: see H.

Roots ## 32, 56, 58 seem not to exist: see J.

* * *

Dolg. lists 17 words (## 63-79) that reflect ways of food preparation and eating habits of the N people. – Sets ## 70 ('soft parts of animal's body'), 73 ('large fish'), and 77 ('hard-roe') represent correct reconstructions: see A. Reconstructions ## 69 ('liver') and 74 ('fish') may require some adjustments: see B. A few words may be added to sets ## 66 ('meat'), 71 ('egg'), and 75 (some fish): see C. A word shall be eliminated from each of the sets ## 67 ('intestines') and 76 (some fish): see D.

Each of the following two reconstructions can be split in two or three words: ## 65 ("food made of ground cereals' *vs.* 'flour') and 68 ('brain, marrow' *vs.* 'to smear' *vs.* 'cheek; beak'): see G.

Sets ## 78 ('hard-roe, span') and 79 ('honey') may represent borrowings: see H. Reconstruction # 72 ('egg') doesn't exist: see J.

* * *

Dolg. selects 22 words (## 80-101) to illustrate "technological activities" of the N people: twisting, boring/drilling, barking/flaying/peeling, rubbing, etc. Dolg. considers different kinds of stone as tools; we shall add 'sharp bone' (mentioned sub # 101). As material, wood, rods, tendons, thorns, and bark were used.

I disagree with Dolg. when he says (p. 65) that "teeth and claws were used as hooks"; rather some natural hooks were designated with the words which normally meant 'tooth/teeth' and 'claw(s)'.

The following words from the above list seem to be properly reconstructed: ## 80 ('flint-stone, knife'), 82-3 ('stone'), 90 ('thorn, hook'), 93 ('bark'), 95 ('hide, skin'), 96 ('skin, fell'), 101 ('sharp bone/tool): see A. A Dravidian word can be added to the root # 97 ('skin, film, bark'): see C.

A word shall be eliminated from each of the following reconstructions: ## 84 ('stone'), 85 ('stalk, stick'), 86 ('tree trunk'), 91 ('tooth, fang, hook'), 98 ('bark, skin'), 100 ('piece of leather'): see D.

Reconst. # 89 ('vein, sinew') needs a substantial restructuring: see E.

Reconst. # 81 ('flint; cut with a flint') shall be merged with words from several languages which mean 'bore, drill, split, chisel': see F.

Each of the following reconstructions shall be split in two or three: ## 87 ('stalk' *vs.* 'sprout, shoot' *vs.* 'ground, floor'); 88 'pole' *vs.* 'arrow'); 92 ('bark' *vs.* 'split, peel'); 94 ('bark' *vs.* 'bark, shell' *vs.* 'bark, crust, peel'); 99 ('to skin, peel off; bark' *vs.* 'to skin'): see G.

There are no sets which could be considered as borrowings or as failures.

* * *

Sets ## 102-8 show that the N people had some knowledge of anatomy (primarily of animals, not humans, which is quite understandable: animal parts were used for food, etc.).

Sets ## 102 ('bile'), 103 ('spleen'), 108 ('jugular vertebra, nape') represent correct reconstructions: see A. Reconstr. # 107 ('back of the knee, armpit') needs some restructuring: see B. Some words can be added to the sets ## 104 ('spleen') and 106 ('occiput, hind part'): see C. A word shall be dropped from # 100: see D.

Reconstr. # 105 ('sinciput, crown of the head') needs serious changes: see E.

There are no erroneous reconstructions in the above word group.

* * *

Sets ## 109-20 represent kinship terms.

The reconstructions ## 109 ('woman from the other moiety'), 110 ('man from the other moiety'), 111 ('relative from the other moiety'), and 117 ('mother') seem quite correct: see A. Reconstr. # 115 ('head of a family') requires a minor change: see B.

Words can be added to the sets ## 116 ('mother') and 118 ('father'): see C.

The following reconstructions require splitting in two parts each: ## 112 ('relative from the other moiety' *vs.* 'wife's sibling'), 114 ('master, elder relative' *vs.* words which originate from **edi-*, # 115); 119 ('child, son' *vs.* 'child, boy; servant'): see G.

Sets ## 113 ('woman from the other moiety') and 120 ('member of a clan') may represent borrowings: see H.

* * *

Dolg.'s last word group (## 121-4) represents "the realm of the supernatural." We have here a seemingly correct reconstruction # 124 ('tell, pronounce magic words'): see A. An Altaic term can be added to # 123 ('exercise magic power): see C.

A semantically dubious word shall be removed from the set # 121 ('make magic, cast spells'): see D. Set # 122 shall be divided into three roots: 'burn (sacrifices)', "deceit, cunning', and 'bless, sacrifice': see G.

* * *

The following lists show different types of evaluation of *NM* data: from those that seem not to contain any problems (list A) to those which may amount to false reconstructions (list J). This evaluation is based on St.'s paper in *NELM* but without his rigid rules for accepting or discarding a reconstruction. Data of other *NELM* authors were also taken in consideration. I find especially interesting those sets that can be split into two or three etymologies (as shown by St.).

A. Sets that seem to be correct, or to require only minor changes (Relatively minor additions in ## 5, 23, 45, 48, 64, 73, 80, 95, 96, 103; dropping off a language [= a borrowing or an irregular form]: ## 20, 23, 90, 111).

5 **?oŕ[u]* 'antelope, deer' (HS-A-D). - St.: 140 (adds A: T); Vov.: 368 (supportive).
8 **čal[U]gV* 'snow, hoar-frost' (HS-U-A [Vov.: 373 maintains that A **č-* doesn't turn to **t-* in T, but this is not *t-*, this is a lenis *ť* matching *d-* in AED phon. charts; still, Vov. objects to the rule A *č-* › T *d-* in APPJ and AED]). - St. 140; Voigt: 320; G.T.
20 **ħalbV* 'white' (HS [drop ECu.: Somali ‹Arab.]-IE-D); St.: 142; Ap.: 309; Voigt: 321.
23 **k'o?ć/cV* 'basket' (HS [drop Eg.: a borrowing from S] -K-IE-U-A-D?). Add to A (**k'ač'a*): T **Kača* (a vessel). - St.: 142; Voigt: 318 and 320; Heg.: 264; Zv.: 363; G.T.
29 **p/p'ešqE ~ *p/p'eqšE* 'spear' (HS-U). - St.: 143. [Invalid if HS is a phylum].
31 **goki* 'track' (HS-U-A). - St.: 143; Ap.: 309; Vov.: 377.
34 **t'ap'V* 'hit (the target)' (HS-IE-U-A-D). - St.: 144; Vov.: 368.
35 **ment'V* 'miss one's aim' (HS-IE-U-A [this latter is dubious]). - St.: 144.
41 **č[a]w(V)RV* 'bull, calf' (HS-IE [not *tauros*]-A). - St.: 145; Vov.: 383 (some doubts).
47 **ç̌ir[i]* '(male, young) hoofed mammal' (HS-K-IE-D). - St.: 146.
48 **p'oK'ü* 'pack, wild cattle' (IE-A; HS: ECh: Ndam only); St.:146; Vov.:369 adds TM.
50 **bVyẑV* 'fur-bearing animal' (IE-U-A-D). - St.: 146; Vov.: 377.
52 **k'un/ńV* 'small carnivore' (HS-K-IE-A). - St.: 146.
64 **?äPHi* 'bake' (HS-IE-A-D). A **ep'e* (+TM **epe-* 'cake' +OJap.); St.: 148; Zv.: 363.
70 **ń[a]K'u* 'soft parts of animal's body' (HS?-IE-U-A). - St.: 148; note Vov.: 382.
73 **K'olV* 'large fish' (HS-K-IE-U-A-D). Add. A: Kor.-Jap., K: Svan. - St.: 149.
77 **t'üRV* 'hard-roe' (U-A). - St.: 149.
80 **č̣[ü]rV* 'flint-stone, knife' (HS-A-D). Add A: T **čar* 'whetstone'. - St.: 150.
82 **til/e(?a)l̂o* 'stone' (K-A-D). - St.: 150.
83 **kiw(V)ñE* 'stone' (HS?-U-K). - St.: 150.
90 **?eẑekU* 'thorn, hook' (HS-A). Plausible match: HS: Cu. **?iŝik^{w}-*, ECu. **ilik-* (etc.) vs A **ēlku* (Vov.: 371). - St.: 151 missed the above, but he justly discarded S **ŝikk-*.
93 **K'a[p?/ç̌E]* 'bark' (HS-K-U-A). - St.: 93.

95 *t'o(w)q'a* 'hide, skin' (HS-K-IE-U-A-D). TM =*tüki-* (‹A *t'uk'i*; + T, M); St.: 151.
96 *t'al(u)ya* 'skin, fell' (HS-U-A-D). Add T, M to A (*t'alo*). - St.: 151; cf. Vov.: 378.
101 *p'ix/ɣyA* 'sharp bone, sharp tool' [**ne** zdes'] (HS?-K-IE-U-A). - St.: 152.
102 *pišV* 'bile' (IE-U-D). - St.: 152.
103 *[t']äxl/ḷA* 'spleen' (HS-K-A). Add TM and Kor. to A (*tiāle*). - St.: 152.
108 *ńiK'a* 'jugular vertebra, nape' (IE-U-A). - St.: 153; Vov.: 372.
109 *kālu/ü* 'woman from the other moiety' (HS-K-IE-U-A-D). - St.: 153; Vov.: 372.
110 *küda* 'man from the other moiety' (HS-U-A). - St.: 153; Vov.: 372.

111 *śeʒA* 'relative from the other moiety' (HS [drop Eg.: G.T.] -K-U). - St.: 153.

117 *?[ä]yV* 'mother' (HS-IE?-U-D). - St.: 117 [*?a-* = old prefix in 116-8?].
124 *tulV* 'tell; pronouns magic words' (HS-IE-U). - St.: 124.

B. Reconstructions that may require substantial changes (A few may be borrowings and rather belong to the list H).

10 *k'ir(u)qa* 'ice, hoar-frost' (HS-K-IE-U-A [but T *k'ira-gu* may be aborrowing from Mong. *kira-ɣu(n)*: Vov.: 379]). - St.: 140-1; Voigt: 318.
12 *tälwV* 'cold season, rain' (U-A; IE *del-* uncertain). Add OJap. to A. - St.: 141.
15 *qaRp/p'V* 'to harvest' (HS-IE-A; IE). - IE: Hitt. rather *harp-* 'gather; a heap' (: Gr. *erép-to* -'reap', Lat. *rapiō*, Lith. *-rëpiu*), not the borrowing *harpi-*; A: corrections (St.: 141).
27 *ɬoŋK'a* 'bend' (HS [drop unrelated Eg.] -IE-U-A). D **toŋk-* 'bow, bend; dangle' may belong here (as per IS; phonetically unclear). - St.: 143; G.T..

30 *t'ul(i)[ǫ]V* 'catch with a net (etc.)' (HS [drop unrelated Eg.] -K-U-A-D?); semantic problems in Altaic. - St.: 143; Vov.: 381; G.T.
69 *mag(i)za* 'liver'; rather **mayzV* (HS **mayz-*, U **maksa* ‹ **mas-ka* ?). - St. 148.
74 *doTgiHU* 'fish' (HS-IE-U-A): some phonetic irregularities: Vov.: 374; **MM**: O

107 *ǫo/atK'E* 'back of the knee, armpit' (HS [drop unrelated Eg.] -A-IE); rather **Ho/aK'E* (as indicated by IE **ok-* and A **ok'-*); phon. irreg. in Semitic. - St.: 153; G.T.
115 **?ediNV* 'head of a family' (HS-A); rather * *?edi* (+ suff.). - St.: 154.; Vov.: 375.

C. Adding a root, representing a N daughter language, to a reconstruction (Note also additions to daughter languages: ## 66, 106, 104; dropping: 13, 54):

13 **yamV* 'body of water' (HS [drop Eg. ‹ S] -U). Add D **am-*. - St. 141; G.T.

19 **mälge* 'breast' (HS?-IE-U). Add K **mlʒe* 'milk' (MR) ? - St.: 142 (sub # 21).

54 **ʒ/ʒugbV* 'a fig tree' [scarcily original meaning] (HS [drop unrelated and phonetically different Eg.] -D). Add A **ʒiugV* 'berry' from # 53. - St.: 146; G.T.

66 **[ç]omśa* 'meat' (HS [add Cu. **HVmS-/*HmVS-*] -IE-U). Add D **ūñ(c)-*. - St.: 148.
71 **muńa* 'egg, testicle(s)' (HS?-IE-U-D). Add A: TM, M. - St.: 148; Zv.: 363.
75 **mEn/ṅi* (some fish) (IE?-U-D). - Add A **mańu-kV*. - St.: 149.
97 **K'aĺ[ü]* 'skin, film, bark' (IE-U-A). Add D **kaḷ-* 'to skin' (from # 99). - St.: 152.
104 **l[ä/e]p'A* 'spleen' (HS-U-A [add M]). Add IE: Germanic **libara* 'liver' and Arm. *leard* (no relation to IE **yekʷr-*). Altaic root: **lip'a* or **liap'a* (: U **läleppV*). - St.: 152.
106 **[g]edi* 'occiput, hind part' (HS-K?-A [add OJap.]). Add IE **ǵhed-*. - St.: 153.
116 **?emV* 'mother' (HS-U-A-D). Add IE? (**mā-* in **mā-tēr*). - St.: 154.
118 **?aba* ~ **?ap'a* 'father' (HS-A-D). Add IE? (**pə-* in **pə-tēr*). - St.: 154
123 **ŝot'V* 'exercise magic power' (HS-K-U-D). Add A **sVtV* 'to curse'. - St.: 155.

D. Reducing the number of N daughter-languages, used in a set, by dropping a root/word (or two, as in # 60).

7 **šüŋU* 'snow' (A-D). Drop HS (uncertain Eg. word); St.: 140; Vov.: 368; Heg.: 262.
37 **?El/ḷi* 'deer' (K-IE-A-D). Drop Yukagir (not a U language). - St.: 145; Vov.: 369.
38 **boča* 'young deer' (HS?-K-U). Drop A: TM: Neg. *bočan* (etc.) ‹ Manchu **bugu-čan* (borrowed from Mong. *buγu* 'deer'). - But Vov.: 377 supports A, contra St.: 145.
60 **m[o](y)ĝV* (some berries) (U-A). Drop HS (dubious Arab.), IE (different).- St.: 147.
67 **q'[u]ʒV* 'intestines' (K-IE-D). Drop A (no phonetic match in TM). - Vov.: 376.
76 **p/p'ayV* (some fish) (U?-D?). - IE **peisk-* may be borrowed from NC. - St.: 149.
84 **boruç/γV* 'stone' (IE-U-D?). Drop HS (different meaning). - St.: 150; Zv.: 363.
85 **ćUḷV* 'stalk, stick' (HS-K-U-D). Drop A: TM: Solon (doutf.) . - St.: 150; Vov. 379
86 **k'oʒçV* 'tree trunk' (HS-K). - Different: A: M. - St.: 150; cf. Vov.: 379.
91 **k'[a]k(w)V* 'tooth, fang, hook' (K?-IE?-U-D). No match in TM; St.: 151; Vov. 376.
98 **k'oRup'V* 'bark, skin' (HS-IE). Drop dubious A: M. - St.: 152; Ap.: 310; Vov.: 378.
100 **K'VRVHp'/pV* 'piece of leather (etc.)' (IE-D). Drop HS (phon. irregular); cf. # 98 (it may represent the same root). - St.: 152.
121 **?arba* 'make magic, cast spells' (A-U). Drop HS (dubious semantics). - St.: 153.

E. Removing root(s)/word(s) from a set that do not belong, simultaneously adding other root(s)/word(s) (Note a radical restructuring in # 9).

4 **SiwVŋgE* 'leopard' [this is not the meaning of the N root] (IE **singh-os* : A [add **zibke* ~ **zipge* › T **jebke-n* 'wolverine']; D **civVŋki* 'hyena, tiger-wolf' or the like). Drop HS. - St.: 140; Vov.: 381.

9 **č[a]RçV* 'hoar-frost' [meaning and spelling inadequate]. Drop words meaning 'hoar-frost' (isol. Arab.; Georg. dial. *čxar-* [cf. NC]). - Restore N [**ć'ara* =] OSNJ # 47 **ĆarV* 'hard crust' by keeping A: T, M, as well as by adding TM: Ewk. *čerī* (all to A **č'era* 'h. c.'); K: Chan *cara* 'hard terrane'; IE **sker-* 'crust, hard cover'; U **ćarV* 'film; to harden'; D **carV* 'coarse, brushy' (thus: K?-IE-U-A-D). - St.: 140.

21 **mayʒV* 'tasty beverage' (HS-IE-A-U). Transfer K **mlʒe* [Laz *mža* 'milk'] to # 19 **mälge* [MR]? - Add D **maṭṭ-* 'honey, sweet juice' [OSNJ # 276]; add TM to A.- St.: 142.

51**?/hUrV* 'squirrel' (IE-U-D). Drop HS (an unsecure Akkad. form). - Add A **Uri-k'V* 'ground squirrel' (T **örke*, TM **urike*). - St.: 146.

89 **ʒiryu/ü* 'vein, sinew' [rather ʒ: Vov.: 371] (K-IE?-A). Drop HS (a dubious Iraqw word) .- Add: D **cīr-* 'root' (Mat.: 341); HS: Class. Ethiop. *śərw* -, - St. 151; Voigt: 322.

105 **t'eqmE* 'sinciput, crown of head' (HS [only Cu.: Awngi *dūmī* 'top' may be valid] -K-IE). Drop A: Mong. *teme-sü* 'edges of a net', etc.; add a new A root: T **tumak* 'hat', Mong. *tomi-la-* 'chief', *tumur-liγ* 'hat', TM **tumŋu-* 'sinciput', OJap. *tumuri / tuburi* 'head, top'. - St.: 152-3; Vov.: 371-2.

F. Merging several roots in one (thus reconstructing semantic archaisms):

14 **moRE* 'body of water'; rather 'water, moisture' (OSNJ #294 **märä*). Merge: HS **mar-* 'drop, rain' and **mir-* 'river'; K: Svan *mare* 'cloud' and Megr. 'lake'; add TM **mu* ‹**mu[r]u* 'water', MKor. *mir*, Ojap. *mi-du* to M. **mören* 'river' (MMong. also 'sea'), T *-mur* (all ‹ A **müri* 'water'); add D **maṛai-* 'rain, cloud'. - St.: 141; Vov.: 368; Heg.: 263.

81**buRV* 'flint > cut with a flint' (HS-A [only TM] + "??? IE **bher-*"); restructure with additions as follows: HS; IE **bherə-* 'cut with a sharp instr.' (quite reliable); A: TM **bura* 'flint', to which add: Ewk. *burbe* 'pierce (etc.)'; T **buragu* 'drill'; Mong. *burgui-* 'wire (for cleaning pipes)'; OJap. *por-* 'bore, engrave': all to A **burV* 'pierce, bore through (etc.)'. Merge all with U **pura* '(to) drill'; D **pōṛ-* 'split, chisel, bore'. - St.: 150 [dropping D: Tamil *pōr* 'hollow/cavity', Kannada *pōr* 'hole' etc. (‹ N **p'urV* dig, hole') from *OSNJ* # 21; adding D **pṓr-*; A: Ewk., T, Mong., OJap.]. - Cf. Vov.: 377.

G. Splitting a reconstructed root in two or three different roots/words:

11 **Sah(i)bV* 'saline earth, desert'. Split into: 1) N **śab[ç]V* 'soil, clay' (U **ś/sojwa*, D. **cav-a ~ cuv-a*); 2) A **sajV* 'stony or shallow place'. - St.: 141; Zv.: 362; cf. Vov.: 376.

22 **k'adV* 'wicker, wattle'. Split into: 1) N **k'adV* (as above: HS-K-IE-D); 2) A **k/k'at'V* 'to mix' ([add to T **kat-*:] M: Mong. *qudqu-*; OJap. *kata-*). - St.: 142; Zv.: 362.

25 **ç/γaŕK'[u]* 'sinew'. Split into: 1) N: K **γrek'(w)-* 'bend, bow' (K-IE-D?): IE **ark^{w}-* 'bow, curve'; D **eṛVt-* 'bow' (‹ **erkVt-*? [phonetically plausible]); 2) HS: Arab. *çirq-* 'root, sinew'; A **ark'a* 'to bind; rope'. - St.: 142-3; Heg. 264; cf. Vov.: 382.

28 **ńoγ/ç lE* 'sinew; to tie'. Split into HS: S **naçl-* '(shoe) strap' and U **ńōle* 'arrow' (+ A: MT: Ewk. *ńul-ga* 'arrow, iron arrowhead' ?). - Cf. St.: 143.; Heg.: 264.

26 **yaŋ[y]V* 'sinew, tendon'. Split into: 1) A: T **jāń* 'bow' (to TM **ǯeje-n* 'sharp point', OJap. *ja* 'arrow', all to A **ǯēja*); 2) U **joŋ(k)se* 'bow'; 3) N: U **jäntä* 'sinew, tendon' (‹**jäntV-* 'stretch, strain'); A: T **jēt-* 'lead, pull'; D **ēnt-* 'stretch (arms)'. - St.: 143.

33 **šubyV* 'spike, spear; to pierce'. Split into: 1) N **šub[i]* 'sharp edge, spear' (K?-A [HS: only Arab.]); 2) N: U: FU **šuye* 'spear, spike'; A: TM **šüje*, Mong. *sojuγa* (add T **sojagu*, TM: Ewk. *čije*, OJap. *soja* 'arrow', MKor. *sāi* 'straw'). - St.: 144.

36 **gurHa* 'antelope' [inadequate form and meaning]. Split into: 1) N **guyrV* 'deer, wild animal' (HS-IE-U-A-D): IE **ǵhwēr-* 'w. a.'; A **gurl ~ *gorl* 'deer, w. a.': M *gŏrüγe* 'w. a.', TM **gur-*, MKor. *korani* 'deer'; D **kūr-* 'deer, antelope'; 2) A **ŋurV* 'male (etc.)': OTurk. *uri-* 'male child, son'; TM **ŋur* 'male (of small wild animals)'; Mong. *gura(n)* 'roe-buck, male wild goat' (not 'antelope'). - St.: 144.

42 **γ/ɋawV* 'wild sheep/goats'. Split into: 1) N **γ/ɋ[o]wV* (as above: HS-IE); 2) N: IE **əwā* 'wound, hurt', A **āba* (**āwa*) 'hunt, kill' (T **āb*, M **aba*, TM **wā-*); 3) TM **abdu(n)* (‹A **at-bu(n)*) 'herd of horses' [not 'cattle, flock'], etc., to **at-* 'horse'. - St.: 145.

62 **m[u]rk'V* 'root, sinew'. Split into: 1) N 'root, sinew' (as above: HS-K-IE-D); 2) A: Ewk. *muni* 'tendon' (‹ A **mujŋe* 'horn, cartilage, tendon' › T **büŋüŕ* 'horn', also Mong. and MKor. words). - St.: 147; Vov.: 379.

65 **qUbźV* 'food made of ground cereals'. Split into: 1) N (as above: HS-K); 2) TM **upa* 'flour' (‹ A **op'V* 'powder' › T **opa*, M *oγo*). - St.: 148.

68 **?ayŋo* 'marrow, brain'. Split into: 1) N.: HS **Hang*^{w}- 'brain'; U **ayŋo* 'brain, marrow'; 2) IE **ong*^{w}- 'to smear'; 3) T **eŋ* 'cheek' (‹A **eŋgV* › Manchu *eŋge* 'beak'). - St.: 148; Ap.: 309.

87 **kańV(-bV)* 'stalk, trunk'. Split into: 1) N: IE **gen(ə)bh-* 'peg, stalk'; D **kāmp-* 'stem, stalk'; 2) HS **kann-* : D **kańń-* 'sprou, shoot'; 3) [N **kAnt'a:*] U **kanta* 'ground, foundation' [and A **kent'a* 'floor, threshhold']. - St.: 150; Zv. 363.

88 **ʒuRV* 'pole, long piece' (HS-K-IE). - Split into: 1) N (as above: HS-K-IE); 2) Mong. *ʒoruγa* 'arrow' (‹ A **ńöra* 'blade, edge' › TM **ńuru* 'arrow') . - St.: 150.

92 **toŕV* 'bark'. Split into: 1) N (as above: HS-A); 2) IE **der-* 'split, peel, flay' (from N **teri* 'to tear, burst'). - St.: 151; Vov.: 371 [change here '82' to 92] .

94 **K'ayerV* 'bark, film' [inadequate spelling]. Split into: 1) A **k'E[ŕä]*, **K'aŕ* (as OSNJ # 217; note HS: S *k'rm*); add K: Georg. *k'r-ol-* (from # 97 **K'aĺ[ü]*); 2) U: FU **kōrV* (genetically related to A. **k'iuru* 'bark, shell'); 3) K **kerk-* 'bark, crust; peel' . - St.: 151; Vov.: 371.

99 **K'oźV* 'to skin'. Split into: 1) N (as above: U-A [HS: only Arab.]); A: Mong. *qoltu-* 'bark; peel off' (from A **k'ulV* 'bark, scales'; add TM **xolda-ksa* 'bark; board'; add T); 2) D **kaḷ-* 'to skin' (from # 97 **K'aĺ[ü]* 'skin, film, bark';). - St.: 152.; Vov.: 378.

112 **[ħ/γV]wäń/nV* 'relative from the other moiety'. Split into: 1) N (as above: HS-U-D); 2) TM **bene-* 'wife's sibling' (TM **bener* ‹ **bere-n*; cf. Mong *beri*). St.: 153; Vov.: 378.

114 **Hić/cxV* 'father, head of the family'. Split into: 1) N: IE **esHo-* 'master'; U **iće* (add A.: T **ečі/ü* 'ancestor, elder relative'; TM **ačV*; drop dubious HS: Ge'ez); 2) A: forms which originate from **edi-* ‹ N **?edi* (# 115). - St.: 154; Vov.: 374.

119 **[?/h]oq'l/ḽV* 'child'. Split into: 1) N (HS; A); A: **uka(lV)* 'child, son': T **ogul* 'son' (add Mong., MKor.); 2) N [**q'uHlV*]: K: Svan *q'law-* 'child, boy'; add to N: A **kūlV* 'slave, servant'; D **kūli* 'hired laborer, servant'. - St.: 154; Voigt: 322.

122 **ʕ[a]lV* 'burn (sacrifices)'. Split into: 1) N (as above: HS-K-IE); 2) A **āli* 'deceit, cunning [dubious semant.]; 3) A: T **alka-* 'bless, sacrifice' [‹**p'iolko* › TM **pulgu*, M **hörgil*]). - St.: 155.

H. Genetic relationship or borrowing? (Note [mostly suspicious] isoglosses: HS-D ## 3, 18; HS-IE: 61; HS-A: 45; HS-IE-D: 49, 79; K-IE: 57; HS-K-IE: 17, 43, 113; K-D: 40).

3 **?[ü]ŕVwV* 'large feline' (HS-D); A: T is different. - St.: 139-40; Vov.: 379.

17 **ɋalV* 'cereals' (K? [HS: only Arab.; IE: loans in Ht., Gr., Lat.]). - St. 17-8; Shev.: 85.

18 **xänt'V* 'kernel, grain' (HS-D [very weak link]); drop IE (not a metath.). - Shev. 86.
24 **p'/pat[a]* 'basket, box' (HS-IE-U-D). - St.: 142 accepts this set, but see Shev.: 84.
39 **buK'a* 'bovine(s)' (T. **būka* 'bull' rather to A **mūk'u* 'male'; IE doubtf.). - St.: 145.
40 **čoma* 'aurochs, wild bovine' (K [only in a Georg. dial.] -D). - St.: 145 (no comm).
43 **diq'a* 'goat' (HS [K, IE: irregular phonetics, borrowings from NC?]). - St.: 145.
45 **bukEγ/ʕV* 'billy goat, ram' (HS?-A [**pUkV*; add TM: Ewk.]); drop IE. - Is there a match between A *p-* (T *b-*, M *b-* and *h-*, TM *p-*) and HS: Sem. *b-*? - Cf. St.: 146.
46 **ʕVp'VrV* 'wild boar' (IE [HS: only Arab.]). - St.: 146.
49 **gadi* or **gati* 'young goat' (HS-D [IE dobtful]). Irreg. phonet. - St.: 146; Voigt: 321.
53 **dik'V* 'edible cereals/fruit' (HS?-K?-D); non-exist.: A **diK-ktä* 'edible berries' (existing words borrowed from M **edege*; this latter - from T **jigde* [< A **ʒiugV*]). - St.: 146.
57 **LV(w)ʒV* (some nut) (K-IE). Drop HS: S **lawδ/z-* (from an IE source?). - G.T.
61 **K'ERV* 'fruit of a leguminous plant' (HS-IE [unusual root structure: *ki-ker-*]).
78 **[k']ür(w)V* 'hard-roe, span' (A-K?-IE? [borrow. in A, K]). Add T, M. to A; St.: 78.
79 **madu* 'honey ' (unclear HS; IE may be a borrowing; D **maṭṭ-* to # 21). - St.: 149.
113 **n/ńu/üśV* 'woman of the other moiety' (HS; borrowing in IE and K?]). - St.: 153.
120 **?arV* 'member of a clan' in: HS: S **?arya-* id., Cu.: Beja *?araw-* 'friend' : IE **aryo-* 'member of a tribe', Hitt. *arawa-* 'free' : U: Ugric **arV* or **arwa* 'relative - member of a clan', etc. (Mostly borrowings? Cf. St.: 120).

I. Reconstructed roots that possibly never existed.

1 **?ibrE* 'fig tree' (HS-D?). - Minimal data (Ap.: 308). Not to Dr. *ir-* _ *ir-* (Zv.: 363).
2 **ĉ[i]bVγV* 'hyena' (HS [D and A: Ewk. belong to # 4, see part E]). - St.: 139; .
6 **maŋ[g]V* 'monkey' (HS: ECh.: Mubi *móŋgò* doesn't match A: TM: Manchu *mońo* (< TM **bońa*). D **maŋk-* < **malVk-* < ST **mlŭk* ?). - St.: 140; Vov.: 373.
16 **ʒükV* 'edible cereals' (U [HS: only Arab.; Eg. word is different]). - St.: 141; G.T.
32 **[d]eʕSV* 'follow the tracks' (HS is totally different; Georg. doesn't match IE; M is rather *de-s-* [thus unrelated]). - St.:143; cf. Vov.: 380.
56 **K'uSV* 'nut' (actually, two A roots: **kušu*; **k'usa*; IE: irreg. phonetics). - St. 147.
58 **but'V* 'pistachio' (HS [A: T only, irreg. phonetics]). - St. 147; Voigt: 322; Vov.: 374.
72 **ʕ[a/o]ħ/xi* 'egg' (HS: only dial. Arab.; OJap. non-exist.]). - St.: 148-9; Vov.: 380.

* * *

Despite many corrections to N sets as represented in *NM*, the vast majority of his reconstructions shall be preserved: either in the form as given by Dolg., or in a changed form, with added or dropped data, sometimes strongly restructured (many reconstructions require splitting in two or even three sets/roots; this sometimes increases the number of N reconstructions).

Only a few reconstructions should be eliminated. It is much more difficult to deal with sets that potentially may represent borrowings (especially if, formally, they match N phonetic charts).

Now I would like to present some latest achievements of N research on the N palatal consonant **ń* (there are only a few words with *ń* in *NM*).

N initial *ń-* becomes IE *y-*; HS (: S) and K *n-*; but it remains *ń-* in U and A; D shows *ń- / n- / ñ-*.

This rule was formulated by IS in the early 60's (see *Mat.* [translated into English by M. Kaiser]; *OSNJ 2*; cf. Engl. translation of some data in *ELM*).

Recent research (cf. data in *Altaic Etymological Dictionary*) has confirmed the correctness of IS's rule, and many more roots of this kind have been reconstructed.

In the table below, sets 1', 4', 7', 15' refer to IS's reconstructions. Sets 1 through 15 are presented in A. Dybo's recent paper (*NENN*; she didn't use HS; N roots in her exx. may occasionally show length; they may end in consonants).

15 roots presented in the table below are as follows: 1) **ńamV* 'squeeze' (also 'grasp'); 2) **ńarkU* 'deer' (or sim.); 3) **ńāk'V* 'rite; medicine' (IE also 'healthy'; A 'soul, feeling'); 4) **ńacV* 'wet, drops' ('fog' in K); 5) **ńVrV* 'fish'; 6) **ńa(y)rV* 'remove hair' (U 'hairless fur/leather'; 7) **ńāyŕV* [rather **ńaçiŕV*; cf. HS: S **nçr* 'youth' and K: Georg. *noγr-* below] 'spring (time), sprout' (K: Georg. 'new grass'; A 'young, spring, summer'; IE 'year'); 8) 'wet, raw'; 9) **ńew-l-* 'young' (U 'calf'); 10) *ńeŋV* 'brother's wife' (U 'woman', etc.); 11) **ńENT-* 'south; noon' (A 'south; warm season'; D 'day'); 12) **ńēlV* 'ground, dirt' (IE 'silt'); 13) **ńimC-* 'fat (of different kind)' (U 'milt'; D 'core, marrow'); 14) **ńo(y)NV* 'reed, rush' (etc.); 15) **ńUlkV* 'remove hair' (U 'to skin'; D 'to pinch').

	N	S | K	IE	U	A (›T)	D
	ń-	*n*	*y*	*ń*	*ń* (› T *j*)	*ń, ñ, n*
1'	*ńamV*		*yem-*	*ńo/amV*	*ńo/umV*	*ñamV*
1	*ńamV*		*yem-*	*ńoma-*	*ńiumi*	*ńem-*
2	*ńarkU*		*york-*		*ńargu*	
3	*ńāk'V*		*yēk-*		*ńák'i*	*nēk-*
4'	*ńäč'/ʒV*	*nd̥-*		*~ńäč/ćV*		
4	*ńacV*	| *nis-l-*	*yes-*	*ńačkV*		
5	*ńVrV*				*ńiro*	*nāri*
6	*ńa(y)rV*			*ńarV*	*ńiår-ke*	
7'	*ńaçrV*	*nçr* | *noγr-*	*yeĥr-*	*ńōre*	*ńārV*	*ñār̲-*
7	*ńāyŕV*	K: G. *noγr-*	*yār-*	*ńVr(k)V*	*niā́ŕa* › T *jāŕ*	*ñār̲*
8	*ńā́ylV*		*yēlo-*		*niāli*	*nāl̲-*
9	*ńew-l-*		*ye/ow-en-*	*ńeplV*	*ńabĺu*	
10	*ńeŋV*		*yenH-ter*	*ńiŋä*	*ńeŋu*	
11	*ńENT-*				*ńéŋńi*	*ńēinḍ-*
12	*ńēlV*		*īlu* (*ī* ‹ **yē*)			
13	*ńimC-*		*īnst-* (*ī* ‹**yi*)		*ńimči-*	*nenʒ-*
14	*ńo(y)NV*		*yo(y)n-*		*n[u]ńa*	*nāṇal*
15'	*ńülV-*			*ńül-ke-*	*ńü/öli-*	*nul̲l̲-*
15	*ńUl-kV*			*ńül-ke*	*ńōlo*	*nul̲-*

Several roots which showed N *ń-* in IS's reconstruction have been restructured since U *ń-* in these cases is secondary; all these roots show in N either initial *n-* or (seldom) *ŋ-* followed by a diphthong of the type -Vy- (*ń-*, *ŋ-* + *-ay-*, *-oy-*, etc.).

In the tabe below, exx. numbered as 16', 17' (etc.) show IS's reconstructions; exx. numbered as 16, 17 (etc.) are from A. Dybo's *NENN*:

16) **nayRV* 'man, male'; 17) **ŋAik-* 'fur-bearing animal'; 18) **noymV* 'soft, weak'; 19) **noyŕV* 'soak, drawn' (IE 'submerge, dive'; U 'swamp'; D 'hide in a hole'); 20) **no(y)rV* 'flow' (A 'lake'; D 'water'); 21) **noyl-* 'saliva, mucus' (A 'snot, tear'); 22) **nVyd-rk-* 'fist'; 23) **nVyk'-* 'shake [D], winnow [IE], knead [A]' (U 'pull out, snatch').

N	HS | K	IE	U	A	D
nVy-	*n-* | *n-*	*n-, an-*	*ń-*	*n(i)V-, ńi(V)-*	*n-, ń, ñ-*

	ŋVy-			*ń-*	*ŋiV-*	*n-*
16'	*NayRV*	CCu *nierā*	*ner-*		*niarV*	
16	*nayRV*		*ăner-*	*ńōre*	*ńiȧŕi*	
17	*ŋAik-*			*ńukV-śV*	*ŋiāk'u*	*nak-*
18'	*ńämV*			*ńälimV-kV*	*ńälimV*	
18	*noymV*			*ńämV*	*niùma*	*nam-*
19'	*ńorV*			*ńorV*	*ńōru*	
19	*noyŕV*		*ne(u)r-*	*ńorV*	*niŭŕe*	*nuṛ*
20'	*ńohrV*	S *nhr*				*ñir*
20	*no(y)rV*				*najrV*	*ńir*
21'	*ńoHlV*			*ńol-ke*	*ńal(a)-*	*ñōlV*
21	*noyl-*			*ńol-ke*	*niale-*› T *jāš*	*ñoḷ-*
22	*nVyd-rk-*			*ńurka*	*niádurgi*	
23	*nVyk'-*		*neik-*	*ńükV-*	*n(i)ok'e*	*nukV*

I would like to underline that neither the restructuring of *NM* roots (primarily proposed by St.; see lists B through J above), nor the recent additions to N reconstructions, have substantially changed rules of correspondences as established by IS and refined later by Dolg.

The N theory certainly works: We could see it both when we were checking Dolg.'s reconstructions against his phonetic charts and when we considered many N reconstructions as recently proposed by the Moscow school Nostraticists.

The N research will certainly intensify in the nearest future since we now have two fundamental etymological dictionaries: the *Altaic Etymological Dictionary* by S. Starostin, A. Dybo and O. Mudrak, and the *Nostratic Dictionary* by A. Dolgopolsky.

Let us see what S. Starostin, the most severe critic of the NM data, says about *NM* in general (St.: 137, *NELM*):

> "Most of the lexical material that [Dolg.] presents is valid and reflects, to my opinion, a deep genetic unity of the languages involved... The book that we are discussing is the best proof of the validity of traditional comparative method and its applicability to distantly related genetic units."

References

AED = *Altaic Etymological Dictionary*, by S.A. Starostin, A. Dybo, O. Mudrak.

APPJ = Starostin, S.A. *Altajskaja problema i proisxoždenie japonskogo jazyka.* Moscow: Nauka, 1991.

ELM = *Explorations in Language Macrofamilies* (ed. V. Shevoroshkin). Bochum: Brockmeyer, 1989.

HLL = *Historical Linguistics and Lexicostatistics* (ed. by V. Shevoroshkin and P.Sidwell). Melbourne: AHL, 1999.

JNPP = Mudrak, O. "Jukagiry i nivxi (problema paleoaziatov)." *PIDR*, pp. 133-48.

Mat. = "V.M. Illič-Svityč's Early Reconstructions of Nostratic (Translated and re-arranged alphabetically by M.Kaiser)". *RLC*, pp. 125-74.

MLŽ = *Moskovskij lingvističeskij žurnal* (ed. by S.A. Starostin). Moscow: RGGU, 1995.

ND = *Nostratic Dictionary*, by Aharon Dolgopolsky. Forthcoming.

NENN= Dybo, A. "Nostratičeskie ėtimologii s načalnymi nosovymi." *PIDR*, pp. 31-8.

NNN = Reshetnikov, K. "Neskol'ko novyx nostraticheskix ėtimologij." *PIDR*,

pp. 163-70.
OS = Orël, V. and O. Stolbova. *Hamito-Semitic Etymological Dictionary.* Leiden – New York – Köln: Brill, 1995.
OSNJ = Illič-Svityč, V. *Opyt sravnenija ostratičeskix jazykov.* Moscow: Nauka: v. 1, 1971; v. 2, 1976; 1st pt. of v. 3, 1984.
PIDR = *Problemy izučeija dal'nego rodstva jazykov na rubeže tret'jego tys'ačeletija.* Moscow: RGGU, 2000.
RLC = *Reconstructing Languages and Cultures* (ed. by V. Shevoroshkin). Bochum: Brockmeyer, 1989.

Nostratic: Examining a Linguistic Macrofamily, edited by Colin Renfrew and Daniel Nettle. (Papers in the Prehistory of Languages, 2.) Cambridge, UK: The McDonald Institute for Archaeological Research, 1999. vii + 419 pp.

Reviewed by Vitaly Shevoroshkin
University of Michigan

[Editor's Note: This and the preceding review have been minimally edited, preserving much of the distinctive style of Professor Shevoroshkin, familiar to readers of *Mother Tongue* (newsletter and journal). Many bibliographical notes refer to abbreviations, e.g., *APPJ, OSNJ,* which are explained in the list of references following each review.]

This book (*NELM*) contains twenty papers (arranged as chapters) presented to the Symposium on the Nostratic Macrofamily (Cambridge, July 1988). Participants in this symposium have mainly discussed Aharon Dolgopolsky's book *The Nostratic Macrofamily and Linguistic Palaeontology* (*NM*: ibid., 1998): see my review above.

According to V. Illich-Svitych (IS) and Dolgopolsky (Dolg.), the Nostratic (N) macrofamily / phylum consisted of HS (Hamito-Semitic = Afro-Asiatic), K [Kartvelian], IE [Indo-European], U [Uralic], A [Altaic], and D [Dravidian]. According to A. Militarev and S. Starostin [St.], HS was not a daughter, but a sister of N (this latter included K, IE, U, A, and D).

In his *Introduction to Nostratic* (*NELM*: 3-18), Lord Colin Renfrew, the host of the above symposium, characterized the Nostratic hypothesis as "one of the most interesting and challenging in the field of modern historical linguistics" (p. 3; see my above review of *NM*).

Renfrew seems to agree with theories that the Natufian culture is closely associated with both Nostratic and Afro-Asiatic (= HS) people. Thus, the N proto-language probably was not as old as some think; it was spoken in the same area where HS was spoken some 11-12 millennia (m.) ago (as well as Sino-Caucasian (SC), according to St.; see below). Such chronology may be possible only if HS was not a daughter but a sister of N (see below, remarks on Dolg.'s and St.'s papers).

If N was spoken just 12 m. ago, the age of IE should not be very high either: probably only 6 to 7 m. (such dating is confirmed by some reliable data: see St. in *HLL*). All this would mean that Indo-Europeans were not the people who developed the 9-m. old urban culture of Asia Minor: rather it was developed by the ancestors of Hurri-Caucasians [Dolg.'s term], or North-Caucasians. These latter seem to have been the people from whom the Indo-Europeans borrowed many more culture-related words than they borrowed from the Semites: see St. in *Drevnij vostok*; Moscow, Nauka, 1988, and IS in *Problemy IE jazykoznanija*, ibid., 1964).

* * *

Aharon Dolgopolsky devotes his paper *The Nostratic macrofamily: a short introduction* (pp. 19-44) entirely to linguistics. He underlines (p. 19) the identity of some most stable words in Nostratic languages: N **mi* 'I' (present in IE-U-A); **?eśo* 'stay' ('be') (IE **es-*; also HS-U-K); N **?itä* 'eat' (HS-IE-M[ongolc]: a branch of A); **bari* 'take' (HS-IE-U-A-D); **wetV* 'water' (HS-IE-U-A-D); **nim?V* '(to) name' (HS-IE-U-A), etc.

Appropriate words in daughter languages (first reconstructed in the early 60's by IS, later modified by Dolg. where necessary) match each other according to strict rules: one can use phonetic tables (presented on pp. 21-24 of *NELM*) to "predict" what reflexes each sound of the given N word (see table of 50 N consonants on p. 20) has in its reflex in a N daughter-language. Naturally, the above phonetic tables correspond to the forms which are now present in Dolg.'s *Nostratic Dictionary*.

There is a table of 34 N cognate sets discussed by Dolg. in the part 6 of his paper (*Place of Hamito-Semitic*: pp. 30-32) of which I present belw (added: few examples (exx.) from *OSNJ, APPJ, NENN* (pp. 35 and 37]). - Note: Cu = Cushitic; Eg = Egyptian; S = Semitic; T = Turkic; TM = Tungus-Manchu [= Tungusic]).

N	HS	K	IE	U	A	D
**mi* I	ECu **-m*	**me/i*	**mē*	**mi*	**bi-/*men-*	
**mä* we		**m-* us	**me-s* we	**mä/e* we		
**ŋV* 1st p.		**naj* we	**n-* us		**ŋa* 1st p.	**nām* we
**nV* 2nd p.					**ńa* thou	**njĭn* thou
**t'ü/i* thou	**t-*		**tu* ~ **te*	**ti*		
**śü/ï* thou		**se/i*, **še-*	**-s* verbal		T, TM **si*	
**mi* what	**m-*	**min* who	**mo-*	**mi*	T **mi*	? **mi*
**yV* which	? S **?ayy-*	**ja-* who	**ye/o-*	**yo-*	**ya*	**yă*
**t'ä* that+	**t* f.; inan.		**to-*	**tä-*	**t'ä-*	**-t* inanim.
**nV* no(t)	**(V)n*	**nu* don't!	**ne*	**ne*	**anV*	
**-tV* pl. suff.	S **-āt*	**-ta*		**-t*	**-t*	
**?itä* eat	ECu **?it-*		**ed-*		**ite-*	
**k'ab/p'V* take, grasp	Cu **k'ab-*		**kap-*	**kappV-*	**k'ab/pa-*	**kap(p)-*
**K'a[p?/ʕE]* bark	Cu **k'app*	Georg. *k'ep'-* sheet ...		**kopa* bark	**k/k'āp'ɨ* bark	

Using this material Dolg. wanted to underline that Joseph H. Greenberg's "Eurasiatic," which includes only three major N languages (IE, U, A), is very close to other major N languages (HS, K, D). Dolg. showed (p. 29) that M. Ruhlen had misinterpreted some important linguistic data which led him to a false evaluation of the degree of relatedness between "Eurasiatic" and other N languages: The main IE word for 'eat' is **ed-*, not **tap-* (this latter appeared only in Tokharian; it should be removed from Greenberg & Ruhlen's list of stable IE-U-A words). IE **tek-* 'touch' doesn't mean 'arrive', but precisely 'touch', it matches well the HS root **tk'*. The ancient root for 'what' is not **yV* (which means 'which') but **mi*; this latter appears not only in IE, U, A, but also in other N languages: HS, K, probably D. If correctly compiled and enlarged, a list of stable words in IE-U-A would generally match HS, K, and D data.

We may add that a relative scantity of the K material can be explained by the fact that only a half of K etymologies are regularly used in comparative research (cf. a disproportianately low number of entries in Klimov's 1964 Kartvelian etymological dictionary).

Dolg. also objected to Starostin's relatively recent interpretation of HS as a sister of N. Here Dolg.'s argumentation (pp. 32-4) that HS shows a very large number (namely, about 1600) of the 2300 N roots which represent the above *Nostratic*

Dictionary seems to backfire (many Arabic forms are highly suspicious.): As St. bluntly puts it (*NELM*: 155-6):

> "The fact that HS roots are so abundant within Dolg.'s material is explained by a - probably involuntary - violation of the 'rules of the game'. The huge number of HS languages and, unfortunately, a not very good state of comparative HS linguistics allows to find matches for almost any root - in some HS subbranch or even in a single isolated HS language."

Starostin objected (as also V. Dybo did) to Dolg.'s unrealistic dating of the N proto-language as being up to 17 m. old. N is rather as old as HS, i.e., 12 m. or so (same age as Sino-Caucasian, possibly a sister of N and HS [St.: 156]).

The list of N grammatical elements in Dolg.'s paper is especially interesting (p. 27). Beside personal and interrogative pronouns we find here deictic particles **?i*, **?a*, **?u* which indicate the degree of proximity to the speaker(s); demonstrative pronouns for active / animate (**sE*) and inanimate (**t'V*) objects; particles indicating collectivity (*Ha* [IE **-ā*], **lA*); collectivity-plurality, duality, etc. Most of these elements are present in IS's N dictionary and in his word lists; grammatical words appear as well as full-meaning words (cf. Index of N words in IS's works [partly translated into English and published] compiled by M. Kaiser; see *ELM*: 128-161).

* * *

What follows now is my comment on several papers in *NELM*, first of all, on those written by anti-Nostraticists.

Alan S. Kaye, *The current state of Nostratic linguistics* (327-58).
A.K. made some proper remarks, for instance: "the N hypothesis is today largely associated with Russian scholarship on the subject, to the point where even American fellow Nostraticist A. Bomhard has often felt like an outsider (p. 329); "either Bomhard or Dolg. is wrong in certain proto-phonemes and their sound correspondences" (p. 330); "if it can be proven that N resemblances are indeed random, then N should disintegrate at once" (p. 332)

A.K. doesn't accept the N theory, though he (as it is clear enough from my exx. below) has never even tried to find out if the N reconstruction is valid or not. It seems, A.K. simply doesn't understand what roots can be considered genetically related on N level and what evidence Nostraticists provide in support of such relationship.

P. 346: on NM # 17 **ʕaLV* 'cereals'. Here A.K. objects to Dolg.'s comparison of Arab. *ʕallat* 'cereals' with Hitt. *halki-* [not *halki*] 'grain (etc.)' as far-fetched, and concludes: "Why not also compare to English *millet*, which is closer to the Arabic form than the Hittite word? Is this possible further evidence for proto-World?".

This might be considered a joke by most readers, but it is not. This trick truly shows A.K.'s perception of Dolg.'s [and other Nostraticists'] work.

For A.K., as well for some of his colleagues, *NM* [and other work on N] is just one of many mass-comparison compilations. This may even be the reason why A.K. misspells so many examples from *NM*; he never tries to check N cognates against phonetic charts in *NM* or *NELM*. He "knows" that all this N enterprise is a fake, so why bother?

A.K. can not understand that there is a <u>difference</u> between N voiced uvular stop *ɢ* (as in *NM* # 17 above) and N voiced velar stop *g* (as in N **mälge* 'breast', see next). I will show below that A.K. makes such elementary mistakes frequently.

P. 346: on *NM* # 19 **mälɢe* [correct: **mälge*. - V.S.] 'breast'. A.K. writes here that the root *mlg* "is Greenberg & Ruhlen's proto-World root ... It is surprising that Dolg. does not mention the alleged wider connections". A.K. clearly does not understand that Dolg.'s *Nostratic Dictionary* (the source of *NM* data) is a fundamental work, based on strict limitations: the author is concerned only with the comparison of N daughter languages and with reconstruction of proto-Nostratic. It is <u>not</u> a popular book on distant relationship of languages.

As I have mentioned above, A.K. is unable to distinguish between N *ɢ* and *g*. He uses only *ɢ* where Dolg. provides either *ɢ* or *g*. If A.K. had consulted phonetic charts at least once he would see that both *g*-like consonants actually represent two different subsystems: the former is a voiced uvular, a pendant to the voiceless *q*, whereas the latter is a voiced velar, cf. voiceless velar *k*. These N voiced stops are reflected in Semitic, Eg., Kartv., and IE as follows (cf. *NM*: 22):

N	*g-*	S	*g*	Eg.	*g, ʒ*	K	*g*	IE *gh, ǵh, gʷh*
N	*-g-*	S	*g*	Eg.	*g, ʒ*	K	*g*	IE *gh, ǵh, gʷh*
N	*ɢ-*	S	*γ*	Eg.	*ʕ?*	K	*γ*	IE *x, [ǰ?], xʷ*
N	*-ɢ-*	S	*γ*	Eg.	*H*	K	*γ*	IE *X*

P. 349: on *NM* # 95 **t'o[w]qa* [correct: **t'o[w]q'a*. - V.S.] 'hide, skin'. A.K. seems not to have the slightest idea of what he is talking about when he says:

> "Bole *ḍiši* ‹ pAA **dihus* [correct: **diḥus*. - V.S.] and pCh[adic] [correct: pWCh] **diHus* 'hide, skin', not **ḍk* (Orël & Stolbova 1995, 160). Further, this is related to Arabic *daḫasa* 'to cut; skin' (Orël & Stolbova 1995)."

The N root in question cannot show initial **d* in Semitic: see phonetic chart, *NM*: 22 (HS **t'*- becomes S **t-'*, **t-*; this rule is very important for revealing sound correspondences between N and HS, and further between HS and S).

The N root in question shows **q'* (not **q*); N **q'* can not turn *h*, *H* in HS languages, inc luding Semitic. Again, check the charts.

So, A.K. (who didn't check the above charts) cites an entirely different HS root: S **diḥus*, WCh. **diHus-* (etc.) as originating from N **t'o[w]q'a*. - As for Dolg., he cites the legitimate reflection of N **t'o[w]q'a* in Chadic: both **ḍk* and the more archaic form **t'k*. - Dolg. cites also the precise IE parallel: **twak-os* (n. sg.). - Starostin mentions the exact A match: **t'uk'i* (St.: 151). - D **tokk-* (*NM*) shows the expected intervocalic reflex *-kk-* of N. **-q'-*.

P. 349: on *NM* # 99 **K'ožV* [correct: **K'oẑV*. - V.S.] 'to skin, to [remove] bark'. After stating that Arab. *qašā* is not listed in Wehr's dictionary (a fact which is irrelevant for our discussion) A.K. declares Dolg.'s comparison [of HS and U data] with Class. Mong. *qoltu/asum* as "far-fetched".

If A.K. had spent just one minute checking phonetic charts we could spare the discussion of this entry today: indeed, N **K'oŝ-* matches Mong. **qol-* precisely (< A **k'UĺV*; St.: 152). HS correspondences, though only from Arabic, fit as well.

It seems, A.K. doesn't understand the difference betw. *ŝ* and *ž* (hence his misspelling and the succeeding, grossly inadequate, comments on *NM* # 99): consonant *ŝ* is a lateral fricative which easily becomes *ĺ* or *l* in many languages (see the beginning of my *NM* review), whereas *ž* is a voiced counterpart to *š*.

P. 345-6: on NM # 13 **yamV* 'body of water'. A.K. tries hard (but to no avail) to discredit this N reconstruction. He says that Arab. *yamm-* may be a borrowing, which it is not (see OS: 536). He says that OS reconstruct pAfAs. (=HS) **ham-*, including **ham-* in Chadic; I would like to know what **ham-* 'water' has to do with **yam-* 'body of water', except that both roots co-exist. Note that A.K. does not even mention here that Orël and Stolbova do reconstruct HS **yam-* 'water, sea'; this word has reflexes both in S **yamm-* and CChad. **yami-* (OS: 536).

This confirms Dolg.'s reconstuction of S **yamm-* (= **jăm* 'sea, lake' < N **jamV* 'water': IS in *OSNJ* # 144).

As for Chadic, A.K. says that Dolg.'s reconstruction of pCh. **[HV]yVm-* "is incorrect (p. 25). This is **yam-* (Bomhard-Kerns 1994, 471), following IS (1971, 279-80)". This is a misrepresentation of facts. Nostraticists imply in many reconstructions that there were ancient affixes (prefixes or suffixes) in the given stem, without formally identifying them. Accordingly, Dolg.'s reconstruction is, actually, **[HV-]yVm-* (the root is **yVm*), - hence Tera *?yim* (< **?V-yim*), etc.

At the end, A.K. mentions a conversation with P. Newman who reported that pCh. 'water' is **amV*. This opinion does not affect the above reconstruction which is built on many Semitic and Central Chadic words showing **y-*. The reconstruction with **y-* appears in IS, OS, ND, being well-founded in all these sources. - Reconstruction of the HS word for a body of water is well supported by Uralic: cf. Samoyed. **yäm* 'sea, large river', etc. (Dolg. in *NM* # 13; cf. IS in *OSNJ* # 144).

P. 349: on *NM* # 90 **?eŝekU* 'thorn, hook' < 'tooth'. After citing the above root along with Dolg.'s Arab. example *šikk-at*, etc., A.K. exclaims: "How can all of the aforementioned relate with pT **il-* 'hang on (smth.)' ?".

A.K. deliberately leaves out other, much better preserved, related forms: Cu. **?iŝikʷ-*, ECu. **?ilk-*; A **ēlku* 'hook', etc. (St.: 151 cites A verb **ōlke-* 'hang on a hook'). These are legitimate cognates which follow all appropriate rules of phonetic correspondences. - Naturally, Vovin unconditionally supported Dolg.'s Altaic reconstruction: "N **?eŝekU* ... seems to be solidly supported by M **elgü-* 'hang on (smth.), hang on a hook' [etc.]" (Vov.: 371).

As for S **ŝikk-*, Arab. *šikk-* (etc.), it seems, indeed, better to drop this part of the set; we still have enough HS and A material to preserve the root in question on the N level (St.: 150 discards Semitic, but ignores Cushitic data; cf. also OS: 103 HS **?ilik-* 'tooth' [mostly Cushitic; they compare a related root: HS **?alVk-* 'bite, chew']; both authors prefer to regard related forms with *ŝ* in SCu. as secondary, though they might be archaic, as Dolg.'s interpretation suggests).

P. 350: on *NM* # 109 **kälu/ü* [correct: **kälu/ü*. - V.S.]. A.K. asks here: "Is it not strange that this root survives only in Semitic and not in the rest of Afroasiatic ?". If A.K. had consulted OS (to which he frequently makes references) he would find there the Chadic counterpart: OS: 310 WCh. **kalya* 'woman', etc. (sub HS **kal-* 'female in-law').

But, in principle, absence of a root in one or more of several groups of a language family is not strange at all. In our case, K **kal-* is only present in

Georgian ('maid; woman; daughter'); IE **ǵlōw-* is only present in Gr. ('husband's sister', etc.), Lat., Phryg. ('brother's wife'), Slav. ('husband's sister'): only 4 out of 12 groups of IE languages.

P. 347: on *NM* # 41 **č[a]w(V)RV* or **čuRV* 'bull, calf'. A.K. again uses his "method" of citing two, phonetically most different, words from two daughter languages, and concluding that these words can not be related:

> "I do not believe many historical linguists would agree with the cognates: Arab. *θawr* 'bull' and Engl. *steer*. The latter < pIE **stā-* 'stand', whereas the Arabic has nothing at all to do with that root (*The American Heritage College Dictionary* 1993, 1330; the IE etymologies are penned by Harvard University's well-known Indo-Europeanist, Calvert Watkins)".

If A.K. had looked up the appropriate sound correspondences in the charts he would see that Semitic *θ-* matches IE *st-* precisely (both S **θawar* and IE **stewr- / *stowr-* originate from N **čawRV*); other sounds fit neatly as well. [Note that IE **taur-os* is a borrowing from Semit. (IS et al.; see sub *NM* # 41)]. Cf. *NM*: 23, 43-4:

N	S	IE	Tg
č-	*θ*	*(s)t*	*C*
č[a]w(V)RV or *čuRV*	*θawar*	*stewr-*	*Cur-*

As for Watkins, he not only provided an obsolete etymology for *steer* and related IE words; he also managed to etymologize this root twice: connecting it both to IE **stā-* 'stand' and to IE **teu-* 'swell'; some etymology!

P. 348: on *NM* # 74 **doTǥiHU* [correct: **doTgiHU*. - V.S.] 'fish'. In his typical way, choosing phonetically very different words, A.K. writes here: "It is difficult to see the connection between Hebr. *dāγ* and Stand. Jap. *üo*. Could the latter be a Chinese borrowing from *ü* ... ?".

He should have consulted the Altaic part of Dolg.'s entry just a little more attentively to avoid embarassment; he would see how A **dőglki* 'fish' became pJap. **(d)i̇wuá*, then OJap. *iwo*, and finaly *üo* in the dialect of Tokyo (cf. archaic *i̇wo* in Kagoshima, *i̇yu* in Shuri, etc.). It is even easier to understand the changes in HS.

Note also what A.K. is proposing, after having discarded the Jap. form (which, actually, is one of many words that evolved in a very similar way): He proposes a comparison with a Chinese word just because it is similar to Japanese. Haven't we enough of look-alikes?

P. 351: on *NM* # 123. A.K. seems not to notice that the N root he cites here is *not* the root he is discussing (this latter is **ŝot'V*).

P. 347: on *NM* # 46. A.K. correctly indicates that this set (which only includes Arab. *ǧi/ufr* [*f* < S **p*] and West IE **apr-os* 'wild boar') may signal a borrowing. But when he adds: "The same Semitic root means 'dust'" he shows that his perception about what homonyms are might be not fully adequate.

P. 347: on *NM* # 64 **?äPHi* 'to bake'. A.K. writes here: "How is Pero *ápò* 'bake' reconciled with pWCh. **ħas-* 'roast, bake' (Orël & Stolbova 1995, 275)?". My only explanation of this strange statement is as follows: A.K. is losing his perception about synonyms. Another example seems to support my explanation:

On the same page, A.K.'s comments on *NM* # 54 **ʒ/ʒugbV* 'a fig tree': "The CCh. *acúwa* 'fig tree' cited by Dolg. is *not* the pCCh. **tiyin-* cited by Orël and Stolbova". Here again, A.K. seems not to realize that he is dealing with synonyms.

P. 347: on *NM* # 55. A.K. writes here: "The Arabic 'terebinth tree' is *but'm ~ but'um* (Lane 1863.I, 219), and *not* the forms cited". But this set is *not* about 'terebinth tree' (see *NM* # 58 **but'V* for this latter): it is about some fruit **b[i]ŕ[uw]q'a*.

P. 348: on *NM* # 71. There are several indications in A.K.'s paper that he did not even see IS's dictionary, though he refers to it several times. Indeed, when mentioning Bomhard and Kerns's reference to IS's reconstruction of a N word for 'egg' (**muña*, 2nd vol. of *OSNJ*), A.K. writes: "... = Illič-Svityč (1971.II, 72-3)" instead of indicating the correct year: 1976. Besides, he wrongly cites here IS's reconstruction (using Bomhard and Kerns's notation where both authors refer to IS.' **muña*). A.K. refers to the tiny first part of IS's third volume as to vol. III; those who *use* the dictionary never refer to it this way.

Here are the correct publication data: The first volume of *OSNJ* was published in 1971; the second in 1976; the first part (only 136 pp.) of the third vol. in 1984.

Only a few critical remarks made by A.K., out of very many, may be considered as serious, unfortunately, all of them were also made by *other* participants of the Cambridge 1998 symposium.

On many occasions, A.K. indicates that this or that Arab. word (cited by Dolg.) is lacking in certain Arab. dictionaries (see A.K.'s remarks to *NM* ## 9, 33, 52, 66, 99, 121, 123, 124.). Apparently, all needed references will be present in ND.

Further, A.K. provides many corrections to spelling and meaning of Arab. words, which is commendable, but this is *not* the expected serious criticism.

A.K.'s very modest criticism, combined with his profound lack of knowledge in linguistic matters, as well as with his obvious indifference to the most important thing - the revealing of sound correspondences between N languages, - scarcelly give him a moral right to statements such as this one (his conclusion: p. 351):

> "It is far more reasonable to assume that either linguistic borrowing through language contact or pure chance accounts for the similarities among the languages discussed in *NM*. Thus the Nostratic hypothesis must remain just that – a hypothesis, and am extremely unlikely one to boot!"

I hope I made it clear how profound the difference is between those who are reconstructing Nostratic, or Altaic, or other proto-languages of a very substantial time-depth (including both grammar and lexicon), and the Nostratic-haters, practically without any knowledge about the data they are discussing and discarding, who are determined to destroy the above reconstructions, no matter what the cost. The latter are destroyers, but also losers; the former are creators, and they are winners; they were, and they are here, for the future. It is certainly not not by chance that several recently established linguistic laws (Dybo's law, Helimski's law) were authored by Nostraticists.

Lyle Campbell, *Nostratic and linguistic paleontology in methodological perspective* (pp. 179-230).

In his opening statement, L.C. (*not* a specialist in any N language) writes about *NM* authored by Dolg. (a *prominent* specialist in HS, A, and other N languages), as well as about the N theory in general, as follows (*NELM*: 179):

> "Nearly all of Dolg.'s 124 N lexical items exhibit serious problems from the point of methodology, meaning the hypothesis of genetic relationship among the language families involved is not adequately supported and provides no adequate foundation for inferences about prehistory."

For a non-specialist, the only way to show that the N methodology doesn't work for the vast majority of cognate sets proposed (124 in our case) is to check the data and to prove that sound correspondences between languages, as presented in cumulative comparison tables (both *NM* and *NELM*), do not materialize (which is, the cognates in any given set don't match each other phonetically). Otherwise, a criticism will remain precisely what it is: a phony accusation, a demagoguery.

Let us see what L.C. himself has to say about checking the existing data (p. 187):

> "I didn't have time (?! - V.S.] to attempt to check the fit of the sounds in the [15] forms presented with the sound correspondences of Dolg.'s charts ... I have little doubt that a thorough investigation would find many more violations".

R. Voigt, not a Nostraticist but an expert in Semitology, formulates his opinion about *NM* as follows (Voigt: 317-8):

> "Contrary to Greenberg's methodology, A.Dolg. adheres to the method of comparative historical linguitics which has been practised by Indo-Europeanists for a long time by using the genealogical tree model and – what is more important – by relying on recurrent sound correspondences ... "

K. Zvelebil, a well-known Dravidologist, has this to say about Dolg.'s D material in *NM* (see Zv.: 361):

> "When it comes to the reconstructible N lexical stock, I have (from the point of a Dravidianist) classified these reconstructions as (a) convincing, (b) ingenious but problematic / in need of additional comments, (c) weak and unconvincing. I have to admit that the first two groups represent a majority of cases ..."

On many occasions (see, for instance, pp. 180-3), L.C. stresses that Nostraticists are widely divided both in classifying languages as branches of N and in actual results of N reconstructions (including both comparative phonetics and concrete roots).

First of all, we have to distinguish between those who belong to the Moscow school of comparative-historical linguistics (Russia and Eastern Europe; also American Nostraticists who support this school: M. Kaiser, A. Manaster Ramer, P. Michalove) and those who don't.

The Moscow school Nostraticists do not differ much in classification of N languages, as well as in N reconstruction. The main argument here is about the status of HS (Afro-Asiatic): Dolg. still includes it into N, whereas St. considers HS as a sister of N. All agree that K, IE, U, A, D are N languages. It is clear now that Yukaghir is not a U language; it is rather closely related to Nivkh (Gilyak): see O.Mudrak in *JNPP*.

No expert doubts today that Korean and Japanese are A languages (cf. *APPJ* and *AED*) whereas Ainu is not even a N language (cf. Vovin's reconstruction of pAinu). Esk-Aleutian is identified as N (probably A) language. Most agree that Chukchee-Kamchatkan (and possibly Nivkh and Yukaghir) are N languages.

There is a profound difference between Moscow school Nostraticists and those who follow A. Bomhard, both in the recoonstruction of phonetic correspondences in N languages and in reconstruction of most N roots (see below).

L.C. (p. 190) defends (naturally) a theory (broadly supported by anti-Altaicists and anti-Nostraticists) about the archaic character of Turkic **s*, **š* as opposed to *r*, *l* in

other Altaic languages, where appropriate words are said either to show a secondary phonetic development or to be borrowings. L.C. asserts (p. 191, implying that neither Altaicists nor Nostraticists can reconstuct phonological systems properly, and that their reconstructions are phonetically both unnatural and implausible):

> "[T]hose with a grasp of phonological systems and phonetic plausibility all postulate a change of *s > z > r and *š > ž > l, where the steps in the change are seen as incremental, intimately interrelated, and natural. No one with a sense of phonology postulates the reverse, the unnatural and implausible change of *r > z/s and *l > ž/š, which are almost unknown in languages elsewhere".

L.C. has profoundly distorted here the approach to the matter by both Altaicists and Nostraticists who deal with <u>two</u> series of proto-A and proto-N phonemes in non-initial position: *r and *l [or $*r_1$ and l_1] (which stay as such in three main daughter languages: T, M, TM) vs *ŕ and *ĺ [or r_2 and l_2] (which merged with M and TM *r and *l, accordingly, but stay as *ŕ and *ĺ in T, becoming r, l in Chuvash, but z, š (etc.) in most other T languages).

Anti-Altaicists consider z and š to be archaic consonants which become r, l in Old Bulgar (ancestor of Chuvash). From here, they think, the appropriate words with r, l reached M and TM languages as loans.

A rich, excellent, lexical material which confirms the correctness of the interpretations presented by Dolg. in *NM*, is present both in *APPJ* and in *AED*. We may now add several new A-U sets from *R* = K. Reshetnikov's study (*NNN* in the bibl.); they show, as expected, the same rule: A *ŕ, ĺ > T *ŕ, ĺ (becoming r, l in Chuvash, but becoming z, š in many other T languages); accordingly, we find *ŕ, *ĺ preserved both in M and TM:

R 8: A **siŕU* > T **siŕ-* 'trickle; thaw, melt' (OTurk., Turk. *siz-*; *-r-* in Chuv.); M and TM **sir-*. - Cf. U **še/ärV* > Komi *šor* 'brook', Udmurt *šur* 'brook, river', etc.
R 12: A **uĺi/e* > T **oĺun* 'shoulder bone' (OTurk *osun* +); *-l-* in TM. - U **wolka* 'shoulder'.
R 13: A **ziuĺa* > T **jíĺ* 'spine',etc. (Yakut *sis*); M **sili* 'back of the head'. - Fin. *selka* 'back'

On many occasions, L.C. maintains (usually re-iterating anti-Nostraticists' arguments) that certain words of N daughter languages are loans where, in reality, they are not: they clearly belong to the basic lexicon (cf. U **wete*, or **wet-*, 'water', etc.).

On p. 195 he writes that Finno-Ugric words of the type Fin. *kalvo* 'film, membrane, wall-eye' (cf. related Eston. *kale, kalu* 'wall-eye') "have been considered to involve a loanword from early proto-Germ. **kalƀan* 'sack-like end of a seine, womb'". In reality, Fin. *kalvo* and many other U words originate from U **kaĺwV* 'film, thin skin' which matches the Germanic form neither semantically nor otherwise. U **kaĺwV* naturally originates from N **K'aĺ[ü]* 'skin, film, bark'; this root is also present in A: TM **xalu-* 'pellicle', etc. (see systematic data sub *NM* # 97).

L.C. (pp. 195-6) proposes to eliminate *NM* sets 109 **kälu/ü* 'female relative-in-law' and 110 *küda* 'male relative-in-law' simply because "similar forms for affinal kin may easily represent old borrowings." The reality is quite different, though.

These two words, along with other kinship terms, are certainly genuine N words, present in *many* N languages, thus being no borrowings (**kalu/ü* is reflected in all six languages); these words strictly follow appropriate rules of sound correspondences in the N languages.

These words allow us to penetrate deep into the structure of a society which existed some 12 m. ago. It is interesting in this respect that, in U, the reflex of N **küda* shows an unexpected auslaut *-ü* – apparently, under influence of its pendant N = U **kälü* (*OSNJ* 1: 303). - See thoughtful comments by Dolg. in *NM*: 84-5.

Next three exx. show reflexes (in all six languages) of N kinship terms which start in **k-*: 1) **kälu/ü* 'female relative-in-law, bride' ‹ 'a woman of the other exogamous moiety' (*NM* # 109, *OSNJ* # 162, with IS's comments; 2) **küda* 'male relative-in-law' ‹ 'a man of the other exogamous moiety' (*NM* # 110, *OSNJ* # 174, with IS's comments); 3) **küni* 'wife, woman' (*OSNJ* # 178, with IS's comments).

HS	K	IE	U	A	D
S **kall-at-*	OGeorg. *kal-*	**ǵlōw-*	**kälü*	**kälin*	ND **kall[i]*
d.-in-l., bride	y. wo., maid	bro. wife, etc	female-in-l.	female-in-l.	female-in-l.
	? **kwiṣ-al-*		**küδü*	T **güδä-*	
	wife's si.'s husb.		hu.'s/wi.'s bro.	y. si.'s husb.	
**k(w)n/*knw*		**g^{w}en-*		**küni*	
one of wives		woman, wife		one of wives	

Many pages in L.C.'s paper (204-11) are dedicated to the dismantling of the Altaic theory. All this writing, unfortunately, didn't persuade Altaicists to abandon their studies; as a result, an *Altaic Etymological Dictionary (AED)* is accomplished, containing over 2600 roots.

There is no doubt that some of L.C.'s criticism is valid, but this is not his own criticism, and it is relatively benign (nobody has yet managed to provide a really destructive anti-N criticism). He simply puts together whatever "damaging" material he managed to obtain from other, usually not very knowledgeable, sources.

An idea to try and examine Dolg.'s material without prejudice, most certainly, never crossed his mind.

April McMahon, Marisa Lohr, Robert McMahon, *Family trees and favorite daughters* (pp. 269-85).

The authors provide some computational models designed to access biological relatedness to linguistic data. There is a feeling that the main aim of the authors is (as it frequently happens nowadays) to undermine the N research.

On p. 283 the authors seem to agree to the preposterous claim by Ringe that a mathematical comparison of 205 of IS's N roots to various probability curves "shows the N results are indistinguishable from chance resemblances under the simplest model he applies".

And the authors are making the next step, siding with Ringe's elementary falsification of linguistic data: "It would seem, then, that not all work on N is based on the comparative method", clearly implying that IS's N dictionary is a work which is *not* based on comparative-historical methodology. How have I to reply to this, not just rejecting these lies but providing something to think about?

The best way seems to cite here V. Ivanov, a prominent comparativist, who was IS's teacher, and with whom IS frequently discussed his work on reconstructing N grammar and stable lexics. The next excerpt comes from V. Ivanov's review of the vol. 1 [1971] of IS's dictionary (*OSNJ*). It was published in Russian in 1974, and later appeared in English (in a book whch I usually don't recommend because of its many typos [origin of which remains a mystery]: *Typology, Relationship and Time*, Karoma, Ann Arbor 1976, edited by VS and T. Markey: 51-6):

> "The main distinction between IS's work and other ... attempts at a broader comparison ... lies in the exceptional precision of his methodology. This can be seen from his scrupulous selection of material ... which is ranged in well-conceived systems of phonological correspondences ... and from the exhaustiveness of his preliminary investigation of data from within each of the language families that are compared ... The extent of IS's insights into the details of each of the [historico-comparative] grammars is

quite exceptional, so much so that the specialists in each of these families can find many new data in the entries of the dictionary.

IS's work demonstrates the explanatory power of N theory by showing that a large number of facts which remained inexplicable within the framework of a given language family can, in fact, be explained from the larger N perspective. ...

One of the better examples might be Sl[avic] **ber-* whose meaning 'to take' happens to be an archaism (*OSNJ* 1: 176) and not, as was generally assumed, an innovation. – N theory demonstrates that such semantic values as 'spring, young plants, young' for Sl. **jar-* are archaic (cf. correspondences in other N languages) wheras the value 'year' (as in many IE languages) is an innovation (*OSNJ* 1: 37, IV).

... IS's dictionary enables the Indo-Eropeanists ... to search for new facts in languages that were considered to have been thoroughly investigated. Suddenly, certain phenomena move to the foreground, phenomena which, when viewed within a framework of a given family alone, seemed rather trivial. There are many examples of such phenomena; we will cite one of them from the [domain] of grammatical relations. IS discussed (p. 206) the formant **-či-* in frequentative and iterative verbs. Its use in verbal suffixes with the significance "polypersonality" in A ... undoubtedly requires comparison with Hitt. *-šk-* in its capacity as an affix denoting polyobjectivity. This is important for ... an understanding of the typology of the semantic development of this formant. ...

Upon reading IS's work, one is left with a sensation ... of exceptional aesthetic perfection, a perfection achieved by having discovered certain principal correlations fundamentally in tune with the objects observed. ...

These disparate examples are cited only by way of illustration to reveal the perspectives which IS saw in his work, a work that outstripped the development of linguistics for decades to come."

Robert Coleman, *Reflections of a distant prospect of Nostratic* (pp. 113-26).

R.C.'s attitude to N seems generally positive (p. 126), but on several occasions, R.C.'s conclusions seem to be based on misunderstanding:

On p. 118 he writes about *NM* # 100 *k'VRVHp'/pV* [actually, with *K'* but this doesn't matter. - V.S.] 'piece of leather': "[H]ere we have ... an initial ejective velar, which is not at all guaranteed by the cited reflexes except for PIE, and then only if we adopt the controversial Gamkrelidze–Ivanov version of the occlusive system". By the way, Zvelebil highly appreciates the above N etymology which shows precise match in HS, IE and D.

It seems that R.C. connects Gamkrelidze's "IE ejectives" *k'* (etc.) with those of N; but Gamkrelidze's *k'* does *not* match N *k'* (this latter is matched in Dolg.'s chart by IE *k* [sic!]). As for Gamkrelidze's *k'*, it is traditional IE *g* which corresponds to N *k* as it is fully clear from the chart. (Only in Bomhard's system traditional IE *g*, his *k'*, matches N *k'*: this leads to many false etymologies: see below). Gamkrelidze's system has absolutely nothing to do with N reconstruction as reflected in *NM*.

On p. 118 R.C. writes: "The change of *č-* [in *NM* # 41 **č[a]w[V]rV* 'bull'] needs supporting evidence, and the appeal to *tauro-* beside **steuro-* 'steer' merely contributes to the mobile *s-* problem, as well as reviving an etymological connection nowadays generally rejected." But Dolg. was far from such "reviving"; any attentive reader will see that Dolg. *separates* **steuro-* (which originates from N) from *tauro-* which is a *borrowing* from Semitic (as underlined by IS in 1964). As for "supportive evidence," Dolg. had no intention to provide it because he didn't write a manual on N phonetics. His phonetic charts are sufficient, and if additional lexical exx. are needed, a reader can take any publication on N (there are enough works in English) and check words in N *č-* or *č'-*: the phonetic match in IE is *st-*.

On p. 117 R.C. writes: "The derivation of PIE **medhu-* 'honey' from PN **madu* (*NM* 79) is incompatible with the table reflexes (*NM* p. 21), which indicates PN initial *d* › PIE *dh*, but medial *d* › *d*. A systematic historical phonology would

eliminate this inconsistency, though one suspects at the price of removing an important family attestation". All this pseudo-scientific chat is for nothing (and Nostraticists are really not that stupid, one should now this by now): there simply is a typo in Dolg.'s phonetic chart: IE *d* printed instead of *dh*: regrettable but not tragic.

R.L. Trask, *Why should a language have any relatives?* (pp. 157-76).
L.T. considers questionable the possibility of reconstructing very old languages (of the N type) since in the past there mostly were isolates and tiny families (with two to four languages) around: thousands of such units. This was "the almost inevitable outcome of many millennia of a foraging experience" (pp. 160-1).

Existing reconstructions of IE, U, A (in *AED*), NC, ST, even of some ancient American languages, show that the past was somewhat different: there always were languages on which a reliable reconstruction could be based. And now we have a N reconstruction, solidly based on language evidence, whatever the opponents may say. L.T. himself spoke very positively about the N research (in his book on historical linguistics).

L.T. considers 50 N consonants far too many (p. 164-5), and it is not excluded that this number will later be somewhat reduced. But the phonetic system of SC (which, according to St., was a sister of N) was, apparently, even richer.

On pp. 166-9 L.T. makes objections to N reconstruction which start with *m-*: in two cases (*NM* ## 6 'monkey' and 40 'aurochs, wild bovine') L.T. seems to be right, proposing to eliminate these sets (cf. lists J and H in my *NM* review). L.T. is right in objecting to Arab. and IE entries in # 60 (some berries; see list D, as above). L.T. is right in implying that two different roots are represented by # 62 ('root, sinew', etc.: see list G, as above). L.T. is right in objecting to M data in the set # 105 ('sinciput', etc.: M should be dropped (list E; note A addition).

In other cases, N reconstructions seem to hold:

NM ## 13 ('body of water') and 19 ('breast') are now supplemented by D and K evidence, accordingly (list C, as above); 14 ('body of water') has a new support from many languages (list F); 21 ('tasty beverage') is considerably restructured (list E); 66 ('meat') is a very solid reconstr., now supplemented by (Cush. and D: list C, as above); 69 ('liver') is corrected: list B; 71 ('egg, testicle') is well supplemented by A data, and dubious HS forms may be dropped (list C);

As for semantic diversity in the set # 35 ('miss one's aim'), it is quite comparable with the situation in IE.

Gyula Décsy, *Beyond Nostratic in time and space* (pp. 127-35).
G.D. presents here a typical (for him) global conception. There are several points in his paper which are totally unacceptable, such as: "Dolg.'s collection [*NM* and *ND*] are mainly gamblems: accidental similarities" (p. 128; again: why not try and check *NM* data against phonetic charts). – "A serious problem in Dolg.'s N word collection is the transcription" (p. 129; it might be difficult, but it is *not* a problem: it works quite well). – "Dolg. assumes 7 vowels for PN ... and 50 consonants, among them a large number of affricates. Affricates and, e.g., *r/l* are new everywhere" (p. 129; note that both affricates and *r/l*-sounds are well preserved in recent reconstructions of such ancient languages as HS [see OS] or NC; there are also many vowels and consonants). – "Palatality correlation (*nj, cj*, etc. ...) is phonemically completely improbable for the time 15,000 BC" (p. 129; proof, please).

Bernard Comrie, *Nostratic language and culture: some methodological reflections* (pp. 243-56). As typical for non-Nostraticists, dealing with N studies, B.C. presents Nostraticists of the Moscow school as routinely violating phonetic rules; see my

discussion of *NM* for examples which prove otherwise. There are some violations in known publications, but they are usually explicitely discussed. Recently, some new phonetic laws were discovered (cf. St., G. Starostin., A. Dybo, O. Stolbova, K. Reshetnikov, O. Mudrak, et al., in *PIDR*) which make reconstructions even more precise both on the N level and on the level of daughter languages.

B.C. objects to the reconstruction (by IS and Dolg.) of N kinship system; this criticism doesn't seem justified, though some points are interesting.

Allan R. Bomhard, *Review of Dolgopolsky's The Nostratic Macrofamily and Linguistic Palaeontology* (pp. 47-74).

A.B. has his own conception of phonology to which the Moscow school Nostraticists strongly object (there were appropriate reviews of A.B.'s work, etc.)

Accordingly, where A.B. writes in his review of *NM* "this etymology must be rejected" it reflects his own perception, not the linguistic reality.

Here is just one example of strong differences between A.B. and Moscow Nostraticists, based on different approaches to N phonetic correspondences:

On p. 65 A.B. writes about *NM* # 109 **kälu/ü* 'woman of the other moiety' (which is one of IS's best reconstructions, excellent phonetically, well preserved in many languages, including IE: **ǵlōw-*) as follows:

> "The initial **ǵ-* in the PIE is irregular – I would expect **k-* instead. This may mean that PIE form is a false cognate. ... I attempted to show that PIE **ǵ(ₑ)lōw-* 'husband's sister' is to be derived from the same root found in Greek *gála* 'milk' ... all ultimately from an unattested **ǵel-* ... 'to suckle, to nourish."

I think this one example is enough to illustrate A.B.'s methodology.

Christopher Ehret, *Nostratic – or proto-Human?* (pp. 93-112).

On p. 95 C.E. writes: "Because we lack as yet the promised etymological dictionary of A.Dolg., the N data has been taken from the careful work of A. Bomhard and J.C. Kerns (1994)": a statement to which I predictably object. This is one of the reasons I am unable to judge how reliable C.E.'s "Nostratic" – Nilo-Saharan comparisons are (pp. 96- 108).

* * *

Now I would like to briefly review papers by specialists in the following language families: Dravidian (K. Zvelebil); Altaic (A. Vovin and D. Sinor), HS = Afroasiatic (D. Appleyard, R. Voigt).

Kamil Zvelebil, *The Dravidian perspective* (359-66).

K.Z. states (p. 359) that a dismissive tone of those who oppose Nostratics "can hardly be maintained in face of the word-list offered by Dolg., even from the Dravidianist's point of view ...". When we look at the sound-shapes and the meanings of a number of items, including a few grammatical features, we have to admit that *something other than chance* is operating".

K.Z. seems somewhat misinformed, though, when he says (ibid.) that "the grammatical structure of the languages in question has so far been scarcely considered." There is a solid amount of grammatical data reconstructed by IS (see many such entries in *OSNJ*, 1 and 2), which are supplemented by many more in *ND*. Of course, N grammar is not yet as thoroughly researched as the N lexicon.

Cf. yet another statement by K.Z. (p. 360) which sounds obsolete: "[M]ost linguists consider these two languages [i.e., Kor. and Jap. - V.S.] as having no known external affiliation; to classify them as 'Altaic' is hardly acceptable". - I can refer the readers, first of all, to Starostin's *APPJ* (1991) where the A character of

both Kor. and Jap. was brilliantly demonstrated; and now we also have the *Altaic Etymological Dictionary* (by S. Starostin, A. Dybo, O. Mudrak) which contains over 2600 entries. – Both Kor. and Jap. regularly appear as A languages in the studies by many prominent Altaicists published over several years (cf. A. Dybo's book on body-part words in A, studies by V. Dybo, O. Mudrak, A. Vovin, K. Reshetnikov, and so on; cf. recent materials in *PIDR*).

K.Z. (p. 362) considers as problematic the reconstruction of initial glottalization in NM # 100 **K'VRVHp'/pV* 'piece of leather'. The answer, of course, is automatic: Since the IE cognate in this set shows **k-*, the N anlaut shall be either **k'-* or **q'-* (see Dolg.'s phonetic charts both in *NM* and in *NELM*). We don't have Kartvelian data among the cognates (Kartv. distinguishes between **k'* and **q'*), therefore we must reconstruct **K'* which means: "either **k'* or **q'*".

Alexander Vovin, *Altaic evidence for Nostratic* (pp. 367-86).
A.V., who accepts the N theory, objects to Dolg.'s location of the N homeland in a Southern area. In any case, he locates the A homeland rather far in the North East. A.V. provides several lists of A forms used in *NM*: accepting some, rejecting or restructuring others. In general, his criticism of *NM* seems too harsh.

I presenti here a comparative list of evaluations of certain A forms by Vov. and St. in their respective papers in *NELM:*

P. 373: on # 8 **čaĺ[U]gV* 'snow; hoar-frost'. - A.V. objects to St.'s reconstruction **d-* where A has **č-* (from N **č-*): "Starostin 1991 [= *APPJ*: 13-4] argued for the **d* as a reflex of pA **č*, but most of his examples are not persuasive". I disagree; cf. a few exx. from St.'s list (A **č-* › T, M **d-* [but cf. M **č[i]-*], TM *ʒ-*, Kor. *č-*):

T **dōl-* 'be filled'		TM **ʒalu-* 'be filled'	Kr *čǎrǎ* 'sufficient'
T **di/ǖr-* 'to string'	WrM *dürü* 'stick in'		MKr *čür-* 'to string'
T **dāl(-ak)* 'willow'	WrM *dola/uɣana* =	TM **ʒali-kta* 'hawthorn', also 'willow'	
T **dāl-* 'spleen'	WrM *deli-gün* 'spl.'		MKr. *čira/ä* 'spleen'
T **dür-* 'straight'			MKr. *čirı-* 'st. ahead'

In the *AED*, this set is as follows: A *č-* › T *d-* : M *d-/č[i]* : TM *ʒ-* : Jap. *t-* : Kor. *č-*.

P. 375-6: on # 59 **mar(y)V* 'berries'. A.V. considers T **bürü* based only on Azeri *müri* 'strawberry' whereas St. discards *müri* as a borrowing from Lezghian. Following IS (*OSNJ* # 282 **marja* 'berry'), St. reconstructs T **bürü-lgen* (well preserved in Tatar, Bashkyr, Kumyk, Chuvash; also borrowed into M languages).

P. 377: on # 21 **mayʒV* 'tasty beverage'. - A.V. writes here: "T **bal* 'honey' ... does not have parallels in other N languages". But cf. St.: 142 who adds TM **mala* 'sesame oil'. (St. follows IS in adding D **maṭṭ-* 'honey, sweet juice' to the N root).

P. 377: on # 80 **č'[ü]rV* 'flintstone, knife'. - A.V. considers this N root as supported only by TM data, but St.: 150 adds T **čar* 'whetstone'.

P. 377-8: # 81 **burV* 'flint; cut/curve with a flint'. A.V. maintains that this root is supported only by TM, whereas St.: 150 adds T, M, OJap., as well as some more

TM data. - This is because St. expands the meaning of the A root, bringing it nearer to N (not just a tool, but also appropriate action): see my preceding review.

P. 378: on # 96 **tal[U]ya* [correct: **t'-*. - V.S.] 'skin', etc. A.V. considers TM **talu* the only representative of A; St.: 96-7 adds T **tu/ol-gak* and (?) Mong *tulum* (‹ A **t'alo*).

P. 378: on # 99. A.V. writes here: "Only M **qoltu-sun* 'bark (of a tree)' [rather M **qoltu-* 'bark; peel off'] supports N **K'oẑV* 'to skin'", but St.: 152 adds TM **xolda-ksa* 'bark; board' and T: (?) Osm. *kuš* 'a hairless spot (on horse's skin)'.

P. 378: on # 122 **ʕ[a]lV* 'burn (sacrifices)'. A.V. considers A: T data valid: **āl-* 'device, deceit (etc.)'; **alqa-* 'bless, praise (etc.)' (he discards Kor.). St.: 155 considers both T roots as belonging to different A roots: cf. my review of *NM*.

P. 379: on N # 85. Both A.K. and St.: 150 discard A form as dubious and isolated.
P. 379: on # 86. Both A.K. and St.: 150 discard A form though by different reasons.
P. 382: on # 53. Both A.K. and St.: 146 consider A root as non-existent.

P. 379: on # 62 **m[u]rk'V* 'root, sinew'. A.V. states here that TM: Ewk. *muŋi* 'tendon' belongs to *NM* 62. But St.: 147 rejects this connection, maintaining that the Ewenki word originates from a different root: A **mujŋe* 'horn, cartilage, tendon'.

P. 381: on # 42 **γ/ɢawV* 'wild sheep, goats'. A.V. doesn't see that there is a substantial difference between N uvular *ɢ* (Starostin's transcription: *G*; this stop is lost in Altaic) as in the root in question, or in **ɢo/atK'E* (cited by A.V. on p. 380), and N velar *g* (turned to *g*, *k* in A languages) as in N **gurHa* (cited on p. 380), or in N **SiwVŋgE* (cited on p. 381). A.V. transcribes both N stops as *ɢ*.

Still, A.V. correctly discards A **āba* 'hunt' as related to *NM* 42; similar: St.: 145.

P. 382: on # 63 **moll̩V* 'pound, smash; gnaw'. A.V. thinks that the M **mölʒi-* 'gnaw' doesn't provide support for the N root. But St. adds MKor. *mɨr-* 'to bite' and other words from A languages in support of *NM* 63 and the appropriate A root.

P. 377: on # 38 **boča* 'young deer'. A.V. writes that only TM **buča* (some deer) supports the N root, whereas St. discards TM forms as borrowings from Manchu **bugu-čan* (› *bučin*), with **bugu* borrowed from Mong.

P. 377: on # 78 **[k']ür(w)V* 'hard-roe, span'. A.V. explains Azeri *kürü* as the only A parallel, but St. considers it a borrowing from Lezgh. *kür* ‹ pLezgh. **kʷir*.

P. 378: on # 112. A.V. accepts TM **bene-* 'wife's sibling' as inherited from A whereas St.: 153 considers TM **bener* as a metathesis of **bere-n*, a different root.

Now, I would like to quote Vovin's general conclusion about *NM* (p. 367):

> "It seems to me that ... *NM* was not the best basis for the discussion of the validity of the N hypothesis. Dolg.'s goal in the book is to reconstruct N homeland and habitat and not to prove the hypothesis itself. In order to discuss the hypothesis itself, in my opinion, it would be better to wait for Dolg.'s forthcoming *Nostratic Dictionary*, as it will undoubtedly include some of the strongest N etymologies that, for obvious reason,

were not provided in this book, since it deals with linguistic palaeontology. ... It seems that the N hypothesis is still viable ..."

Denis Sinor, *Some thoughts of the Nostratic theory and its historical implications.* (387-400). D.S. makes several critical remarks which practically coincide with those made by Vov. and St. On other occasions, D.S. considers some forms cited in *NM* as borrowings.

When he says (p. 393) that "most of the ... forms listed in *NM* rest on a handful of haphazardly collected words", he certainly exaggerates. Ad when he writes that "the genetic relationship of Turkic, Mongol, and Tunguz languages is subject to grave doubts" he is simply wrong. I really recommend to all those who want to find out more on the Altaic relationship to look at least through *APPJ* (to say nothing about important Altaic studies by K. Menges, N. Poppe, V. Illich-Svitych, V. Cincius, V. Dybo, A. Dybo, O. Mudrak, A. Vovin, K. Reshetnikov), and now also *AED.*

David Appleyard, *Afroasiatic and the Nostratic hypothesis* (pp. 289-314).
D.A. is against including Afroasiatic in a "Nostratic" super-phylum (p. 289). – He criticizes Dolg. for including single language citations as representative of whole families (p. 293). – He presents his own view and that of other scholars (I. Diakonoff, A. Dolgopolsky, V. Blažek) on different sub-systems of HS pronouns on pp. 293-300). He maintains that Afroasiatic evidence for N pronominal system is either easily contestable or highly suspect (p. 300), which seems too pessimistic.

He considers as borrowings many Afroasiatic forms used in *NM.* (pp. 304-11). Generally, he describes Afroasiatic entries in *NM* as "scant and weak" in comparison with IE and U material (p. 311). He accepts a posibility of IE and K, as well as IE and U genetic relationship, but he is skeptical about a genetic link between IE and Afroasiatic (p. 312).

Rainer Voigt, *On Semitohamitic comparison* (pp. 315-326).
On p. 317 R.V. states that Dolg. follows the comparative-historical methodology "by relying on recurrent sound correspondences". R.V. specifically discusses Dolg.'s studies in HS (=SH) languages on pp. 318-20. On pp. 321 R.V. compares three Sem. and Chad. reconstructions as presented by Dolg. and Orël-Stolbova (OS) and concudes that Dolg.'s reconstructions are more plausible than those by OS.

On p. 322 R.V. analyses HS material in *NM* entries ## 13, 58, 89, 111, 118, 119. proposing some corrections.

He concludes his short but very informative paper as follows (p. 322):

"There are many quite convincing comparisons that go far beyond SH [=HS] and IE. A great deal of them must be attributed to loans. Especially the words for animals (as 'hyena', 'lion', 'leopard', 'monkey', 'bull', 'calf'), cereals, 'basket, vessel', weapons (as 'spear') and 'piece of leather (used esp. as footwear)' are to be considered as loanwords ... which sometimes show a great diffusion throughout the Afroeurasian continent".

* * *

Let me now discuss, very quickly, papers authored by Nostraticists.

The most important paper in *NELM* is that by S.A. Starostin *Subgrouping of Nostratic: comments on A. Dolgopolsky's The Nostratic Macrofamily and Linguistic Paleontology* (pp. 137-56).

S.S. (pp. 138-9) applies a rigorous comparative methodology, dividing families under comparison in 3 groups (I am not discussing here SC languages):

I. Kartvelian. This small and compact family is represented by 4 languages, each being able to represent the whole family (if there is no K reconstr.).

II. Uralic, Dravidian. Each of these large families is characterized by a dichotomic split: U consists of FU and Samoyedic; D consists of ND and SCD. A root present in one branch only, – for instance, FU, – has a good chance to represent Common Uralic.

III. IE, HS, Altaic. These are very large families, each having many branches. The probability of a common root being preserved by only one branch is small. If there are many roots isolated within one branch, these roots may be explained as later loans (though some roots of this type may represent the appropriate common language).

St. regards as reflected in a family of type III "only when it is represented by at least two subbranches (as Slavic and Germanic within IE, or T and TM within A). – Such restriction is not necessary in families of type I or II, so "a root may be withdrawn from the comparison only if it can be shown to be irregular in phonology, dubious in semantics or borrowed from some other known source."

St. analyzes all data represented by *NM* and shows that there are only 47 HS roots which follow the above rules vs 63 IE, 61 A, 60 U 56 D. This is one of the reasons he considers HS not a daughter but a sister of N.

A similar parallel analysis was conducted for SC languages (NC, ST and Yeniseian); when comparing possible cognates between N and SC langauges, St. obtains 47 NC roots which have genetic parallels in N languages. So he concludes that both HS and NC are sisters of N. (Cf. St.'s paper in *ELM* where over 200 N - SC cognates are listed with appropriate comments).

In his analysis St. restructures many roots from *NM*: cf. my review of *NM* where St.'s data are broadly used.

Peter Michalove, Alexis Manaster Ramer, *The use of reconstructed forms in Nostratic studies* (pp. 231-42).

Both scholars consider two N reconstructions which designate fish and propose several interesting corrections.

On p. 240-1 they conclude their research as follows (correctly showing some important perspectives of the N research):

> "That Dolg.'s findings differ in many respects from IS's work of the 1960's, and that our views differ somewhat from Dolg.'s is inevitable; it would be hard to imagine that research in the various languages associated with N in the past generation did not affect current work in Nostratic. It is significant in this regard that IS did not limit his work to N, but made important contributions to some of the daughter languages, such as IE and A, as Dolg. has to Afroasiatic. Just as all living languages are continuously in a process of change, so all viable reconstructions are continuously subject to revision. This dynamic does not invalidate work on higher-level reconstructions; it is a healthy process that, in the long run, will refine and deepen our understanding of the history of the various attested languages."

Irén Hegedűs, *Linguistic palaeontology: for and against* (pp. 257-68).

On p. 263 I.H. makes an important remark, concerning the reconstructed meaning of M (an A language) root derived from N # 14 **moRE* 'body of water'. I.H. concludes that this meaning did not include a sub-meaning 'sea'. She thinks that the meaning 'sea' may have been present in Early N, and was preserved in HS and IE who lived near the sea in the Near East, but was changed to 'lake' in Kartvelian (after K people migrated into a mountaineous region [but didn't yet come to the sea].Accordingly, the meaning 'sea' was not preasent in Altaic languages which very early migrated to continental areas (cf. aso Vovin's paper in *NELM*).

On pp. 257; 265; 266 I.H. writes:

> "The primary task of those involved in the evaluation of *NM* ... is to examine the validity of Dolg.'s palaeontological conclusion rather than the feasibility of the N hypothesis itself. ... For eliminating inconsistencies it is crucial that linguists working on reconstructing the N protolanguage receive constructive criticism and work with a self-critical attitude. ... Morphological reconstruction rather than lexical reconstruction could be the cutting edge in further research because it seems to provide qualitative proof rather than merely accumulate evidence for N quantitatively. Most critics of N linguistics would tame their antagonistic attitude if N shaped up grammatically. ... Attempts in this direction have been made but little attention has been paid to them. ... The N hypothesis ... has grown out of material correspondences observable in languages of Eurasia and its supporters work in accordance with the principles and established practice of historical linguistic comparison and reconstruction. The process of proving this hypothesis ... has already provoked meaningful discussions, it attracts more and more attention and is gaining a more objective critical attitude whether we argue for or against the N hypothesis".

I.H. is one of vey few Nostraticists who work with morphology, – an area which was considered highly important by IS but which still is not properly represented in N studies. In this connection one may be reminded of a very interesting work by B. Čop whose area of research was IE and U as genetically related languages. – A prominent Uralicist and Nostraticist, E.Helimsky, wrote in his paper on N studies outside of USSR/Russia (see above) as follows:

> "The most substantial contribution ... came from the Yugoslavian Indo-Europeanist Bojan Čop of Ljubljana ... Largely following Collinder, as IS also did, Čop ... supported the hypothesis of genetic kinship between IE and U languages, and postulated the existence of their common ancestor, the Indo-Uralic proto-language, which in its turn constitutes, on a line with proto-Altaic, and, possibly, with several other languages, a branch of Eurasian macrofamily (Eurasian = N). ...
>
> Čop didn't alter his general approach, although his acquaintance with *OSNJ* found expression not only in a high esteem of this book in the pages of his own works, but also in using a large number of etymologies offered by IS and in systematic referring (occasionally also critically) to the latter's results. On the whole, *OSNJ* and the works by Čop may be considered as, to a certain degree, independent achievements in linguistic thought, and both the material and the conclusions of the two authors can be compared."

E.H. also underlines the importance of K. Menges (a pioneer of N studies who worked on morphology of several N languages):

> "In 1968 K. Menges, in the concluding chapters of his book, *The Turkic Languages and Peoples*, pointed to the kinship of the A languages with the U and D (closer) and also with the IE, Jap., Chukchi-Kamchatkan, HS (more distant), thus acknowledging the existence of the N macrofamily as it was singled out in the studies of Dolg. The adherence to the idea of N kinship is confirmed also by all later publications by Menges, which contain systematic references to IS and make a wide usage of etymologies, reconstructions, and phonetic correspondences drown from *OSNJ* (... the problems of D-A comparison, especially in the domain of morphology, discussed on the wide N background; ... the discussion of the results of R.Miller in Jap.-A comparison, with stressing that it would be pointless to isolate this comparison from Common N problems; ... the application of the data of N to the problems of Tungus [=TM] reconstruction)."

It is time to start paying more attention to N morphology, especially now when the N lexicon is well presented in Dolg.'s *ND*. It was the reconstruction of N morphology which became IS's primary goal when he devoted his knowledge and skills to N research.

Vitaly Shevoroshkin, *Nostratic languages: internal and external relationship* (pp. 75-91).

Phonetic correspondences between N languages are discussed on pp. 76-80. There exists a possibility that IE system of stops was not *T - D - Dh* (traditional reconstruction) but *Th - T(:) - D* which makes it identical to Altaic - and possibly to N if N *T'* were re-interpreted as *T'* (St.).

Hittite *h-*, *-hh-* represents one of IE laryngeals which were not "vowel-coloring": It is simpler to assume that N sequences of the type *qa-*, *xa-* became IE *xa-*; it is much more complicated to assume that N *qa-*, *xa-* first became IE *He-*, and then they became *(H)a-* because of the vowel-coloring nature of this IE *H-*.

The IE labiolaryngeal x^w (or H^w, for that matter) became *h(u)w-*, *hu-* in Hittite and Luwian.

It seems that N uvulars *q'*, *q* existed (contrary to St.): precise parallels can be found in Kartvelian [and HS], as well as in SC / NC languages.

Phonetic correspondences between N and SC / DC (Dene-Caucasian, such as Salishan) languages are discussed on pp. 81-4; cf a sequel to this in the *Festschrift for W.W. Schuhmacher* (edited by P. Sidwell in Melbourne).

Several sets from *NM* which might represent incorrect reconstructions are discussed on pp. 84-9.

The conclusion (p. 90) contains the following statement:

> "*NM* is especially important since it combines linguistic and ethnological data. The book supplies us with a better undertanding of the life of our remote ancestors."

---

* * *

I would like to conclude this review with a quotation from a 1988 paper by William H. Baxter, a prominent American linguist:

> "... I am optimistic that more distant language relationships can be established than are now widely recognized; specifically, I find the evidence for an Indo-Uralic relationship, and for an Altaic family (including Korean and Japanese) rather persuasive. I also take the N hypothesis quite seriously, and have great respect for the work of many of its advocates."

References

AED = *Altaic Etymological Dictionary,* by S.A. Starostin, A. Dybo, O. Mudrak.

APPJ = Starostin, S.A. *Altajskaja problema i proisxoždenie japonskogo jazyka.* Moscow: Nauka, 1991.

ELM = *Explorations in Language Macrofamilies* (ed. V. Shevoroshkin). Bochum: Brockmeyer, 1989.

HLL = *Historical Linguistics and Lexicostatistics* (ed. by V. Shevoroshkin and P.Sidwell). Melbourne: AHL, 1999.

JNPP = Mudrak, O. "Jukagiry i nivxi (problema paleoaziatov)." *PIDR*, pp. 133-48.

Mat. = "V.M. Illič-Svityč's Early Reconstructions of Nostratic (Translated and re-arranged alphabetically by M.Kaiser)". *RLC*, pp. 125-74.

MLŽ = *Moskovskij lingvističeskij žurnal* (ed. by S.A. Starostin). Moscow: RGGU, 1995.

ND = *Nostratic Dictionary,* by Aharon Dolgopolsky. Forthcoming.

NENN= Dybo, A. "Nostratičeskie ėtimologii s načalnymi nosovymi." *PIDR*, pp. 31-8.

NNN = Reshetnikov, K. "Neskol'ko novyx nostraticheskix ėtimologij." *PIDR*, pp. 163-70.

OS = Orël, V. and O. Stolbova. *Hamito-Semitic Etymological Dictionary*. Leiden – New York – Köln: Brill, 1995.
OSNJ = Illič-Svityč, V. *Opyt sravnenija ostratičeskix jazykov*. Moscow: Nauka: v. 1, 1971; v. 2, 1976; 1st pt. of v. 3, 1984.
PIDR = *Problemy izučeija dal'nego rodstva jazykov na rubeže tret'jego tys'ačeletija*. Moscow: RGGU, 2000.
RLC = *Reconstructing Languages and Cultures* (ed. by V. Shevoroshkin). Bochum: Brockmeyer, 1989.

Numerals: Comparative - Etymological Analyses of Numeral Systems and Their Implications (Saharan, Nubian, Egyptian, Berber, Kartvelian, Uralic, Altaic and Indo-European Languages), by Václav Blažek. Brno: Masaryk University, 1999. vii + 337 pp.

Reviewed by John D. Bengtson

This comprehensive study of numeral words around the world is the work of veteran Long Ranger Václav Blažek, whose scholarly life, as explained in his "acknowledgement" (page i), has alternated between linguistics and mathematics. He taught mathematics and physics for nine years before completing his doctorate in linguistics and joining the staff of Masaryk University in Brno, Czech Republic.

This book is divided into three major parts: (a) "Non-Indo-European Numerical Systems (Saharan, Nubian, Egyptian, Berber, Kartvelian, Uralic, Altaic)"; (b) "Indo-European Numerals"; and (c) "Patterns of Creating Numerals" (with examples from languages all over the world). Several of the chapters (listed on p. 337) have been published separately in journals and books. (For example, see Roger Wescott's review of Blažek's "Indo-European 'Seven'" in *Mother Tongue* IV, pp. 147-150.)

In the first section of the book, "Non-Indo-European Numerical Systems," the author comprehensively reviews the numeral words in seven language families. Here I will briefly describe the first chapter ("Saharan Numerals," dedicated to the memory of Karel Petráček) to give an idea of Blažek's thorough methodology. First, Blažek reviews the history of classification of the Saharan language family, which he accepts as part of the "vast Nilo-Saharan macro-phylum," noting that Petráček did not accept the Nilo-Saharan hypothesis. Classifications by Greenberg, Bender, and Ehret are briefly outlined. Blažek then mentions that there is as yet no complete comparative historical phonology of the Saharan languages, and acknowledges previous studies of Saharan numerals (F. Müller, Kluge, Petráček). Numeral words from the Saharan languages are listed in tabular form, followed by a "comparative-etymological analysis" of each etymon, with citations of possible external cognates in other Nilo-Saharan languages, and occasional references to areal or deep genetic parallels in other language families. Blažek ends each chapter with conclusions regarding the most widespread numeral etyma of the family concerned, and a bibliography (in the case of Saharan, almost three full pages). The same format is followed in treating the other six non-Indo-European language families.

The second major part of the book, "Indo-European Numerals," has a separate chapter for each numerical category ('one' through 'ten', 'hundred', and 'thousand'). In the case of 'one', three IE roots, **oy-* 'one', **sem-* 'one', and **per-/*pro-* 'first', are catalogued in detail in all the major branches of the IE family (Indo-Iranian, Anatolian, etc.), including scantily attested and extinct branches such as Phrygian (for other numerals, also Illyrian, Thracian, Dacian, [ancient] Macedonian, Lusitanian, Venetic, and

Messapic). Then Blažek analyzes the IE reconstructions and their proposed etymologies, however likely or unlikely. (E.g., for IE **kʷet-wōr* 'four' no fewer than eleven different hypotheses are discussed!) Blažek suggests external parallels in non-IE languages (e.g., IE **sem-* 'one' with Altaic **somi̯V* 'one [of a pair]; single'), draws conclusions, and provides references (for the chapter on IE 'one', almost four full pages).

In the third and smallest section of the book, "Patterns of Creating Numerals," Blažek gives examples from all over the world illustrating systems of numeral formation. For example, Telefol (New Guinea) is used to exemplify "transparent semantic motivation," its numerals analyzable as 'little finger of the left hand (1), ring finger of the left hand (2),' etc. Jawony (Australia) illustrates a binary system (3 = 2 + 1, 4 = 2 + 2, etc.). Yukaghir (Siberia) and Yuma (North America) exemplify ternary systems (6 = 3x2), etc. Blažek concludes this section (p. 336) with the statement: "The creation of numerals confirms more than any other human activity that **man is a measure of himself.**" Blažek's book is a worthy successor to the earlier compendia by August Pott, Alfredo Trombetti, and Theodor Kluge.

Die Burushaski-Sprache von Hunza und Nager, by Hermann Berger. (Neuindische Studien herausgegeben von Hermann Berger, Heidrun Brückner und Lothar Lutze , Band 13.) Wiesbaden: Harrassowitz Verlag, 1998. Teil I: Grammatik, xxiii + 271 pp.; Teil II: Texte mit Übersetzungen, x + 273 pp.; Teil III: Wörterbuch, 646 pp.

Reviewed by John D. Bengtson

The mysterious Burushaski language of the high mountainous Northern Areas of Pakistan has long fascinated linguists and anthropologists, because of the odd typological features of the language, and its supposed isolation from all other language families. At last, a definitive masterwork has been completed by the foremost present-day scholar of Burushaski, superseding the six-decade-old work by Colonel D.L.R. Lorimer (1935-1938). Hermann Berger's *magnum opus* is the result of his finely honed research during and between expeditions to the Hunza, Nager, and Yasin valleys over a period of 35 years.

Note that these three volumes primarily concern the Hunza and Nager dialects, which stand together (despite minor differences) against the more strongly divergent dialect of Yasin (= Lorimer's "Werchikwar," Zarubin's "Veršikskij"). Berger (1974) some time ago updated the description of the latter dialect. The following table illustrates some of the similarities and differences between the three dialects, which may involve phonology, lexicon, or grammar:

Meaning	**Hunza**	**Nager**	**Yasin**
'man'	*hir*	*hir*	*hir*
'(small) bird'	*ch̤in*	*ch̤in*	*c̣en*
'nostril'	*-múltur*	*-múltur*	*-múśpuśiŋ*
'barefoot'	*ćhu*	*ćhu*	*hultáṣ*
'a little bit'	*phíwan*	*kámaṇ*	*kamrénaṇ*
'firewood'	*γaśíl*	*ćúuni*	*gádar*
'(it) is' (y class)	*bilá*	*dilá*	*duá*

(See also Varma, 1941; Anderson, forthcoming a.)

The first volume of *Die Burushaski-Sprache* is a complete grammar, consisting of *Lautlehre* (phonology), *Formenlehre* (morphology), *Syntax,* and *Wortbildung* (word formation). Here I will mention only a few salient points. Regarding phonology, Berger describes Burushaski (Bur.) as having five vowel phonemes and thirty-four consonant phonemes. This in itself is a significant improvement over the book by Lorimer, who knew nothing about *phonemes*, but transcribed the sounds of Bur. as he heard them, without any inkling of the phonemic system underlying them. (For example, Lorimer variously recorded the word for 'open' as *phʌṭʌŋ, pfʌṭʌŋ,* or *fʌṭʌŋ,* where Berger uniformly has the phonemic *phaṭáṅ* = *phaṭáŋ*.) The Bur. consonants are typical of the greater Indian linguistic area in having the retroflex series [ṭh, ṭ, ḍ, ṣ, c̣h, c̣, j̣, ẏ], but are

atypical in having a uvular series [qh, q, ġ = γ]. The most peculiar sound in Bur. is the "dotted y" [ẏ], described by Berger (vol. 1, p. 22) as a "stimmhafter retroflexer Sibilant mit gleichzeitiger palatal-dorsaler Engebildung" [voiced retroflex sibilant with simultaneous palatal-dorsal narrowing]. Morgenstierne (1945, p. 68) described the same sound as "a fricative *r*, pronounced with the tongue in the retroflex ('cerebral') position." Anderson (forthcoming, b) writes of [ẏ] as "a curious sound whose phonetic realizations vary from a retroflex, spirantized glide, to a retroflex velarized spirant." The sound is found only in the Hunza and Nager varieties of Bur., not in Yasin, and then only in medial or final position (*e.g., giẏált* 'ladle', *bépaẏ* 'yak'). According to Morgenstierne a similar sound is also heard in Ḍumāki, an Indo-Aryan language spoken by a caste of musicians and blacksmiths who live in symbiosis with the Burusho.

Among grammatical peculiarities of Bur., Berger describes in detail its noun classification, a four-class system typologically similar to those of some East Caucasian languages, Yeniseian languages (see *Mother Tongue* IV), and even some African (Niger-Congo) languages. Bur. also has a rich system of noun inflection, including a variety of agglutinative case endings, and several plural endings. The pronominal system is intricate, with, for example, suppletive pronoun stems in the first and second person singular. The verbal declension is exceedingly complex, with a maximal template of four prefix positions preceding the verb stem and six suffix positions following it. (See also Anderson, forthcoming a.) Berger describes all of this clearly and succinctly.

The second volume consists of Burushaski texts, 41 in the Hunza dialect and 26 in the Nager dialect. Many of the texts describe spirits or supernatural beings, for example, the *bilás* 'man-eating demon', *biṭán* 'shaman, soothsayer', *ḍáŋlaṭhas* 'man-eating female demon', *meélgus/mayálgus* 'sparkling female spirit, seen in meteors', *parí* 'fairy', *phut* 'gnomish nature spirit', etc. Other texts include military tales and hunting stories. These newly published texts form a valuable complement to those transcribed earlier by Biddulph, Frémont, Hunzai, Lorimer, Tiffou, Varma, and others. (See the catalogs by Bashir, 2000, and Tiffou, 2000).

Berger's third and largest volume is a comprehensive dictionary of the Hunza and Nager dialects. Each word is cited with dialectal designations (hz., ng.) as necessary, with its Yasin equivalent, if any, and frequently with etymological references to other languages. In the past, before more recent contacts with Islamic Pakistan, the most prolific source of loanwords was Indo-Aryan, for example Bur. *yoór* '(water-)mill', ultimately from Old Indic (Sanskrit) *yantrá-*, or Bur. *aśtán* 'groom (of a horse)' < Shina *aśtóon* < Skt. **aśva-sthānin-* (Morgenstierne 1945: 92). Berger provides references to Turner's (1966) Indo-Aryan comparative dictionary, e.g. "T 10412," where he finds borrowing from Indo-Aryan likely. Older and younger layers of Indo-Aryan loanwords can be distinguished, for example, by the fact that in the older layer Sanskrit *bh, dh* become Bur. *ph, th,* respectively (Skt. *bhūta-* 'being' > Bur. *phut* 'nature spirit'; Skt. *dhūmāyana-* > Bur. Hunza *thómal* ~ Nager *thóman* 'fragrant juniper smoke'), while in recent loans *bh, dh* become *b, d,* respectively (Skt. *bhaga-* 'share, portion' > Bur. *báago,* Yasin *bágu;* Skt. *dhūmá-* 'smoke' > Bur. *dumáṣ* 'cloud', etc. Berger, personal communication, 1996). In some cases we find competition between an older Indic loanword, e.g., Bur. *saás* '1000' (< Skt. *sahásra-*) and its more recent Persian equivalent, mediated through Urdu, Bur. *hazaár* '1000' (< Urdu < Pers. *hazār*), etc. In isolated cases,

borrowings from Urdu or other languages can supplant basic native words (e.g., *náam* instead of the native *-ík* 'name'). Bur. has borrowed words from many surrounding languages, for example Tibetan or Tibetic: Bur. *bras* '(uncooked) rice', *zo* 'hybrid of yak and cow'. (A Tibetic language, Balti, adjoins the Burusho to the east.) Even some languages at a farther remove have contributed, e.g. Bur. *íran* 'cream', *qaćír* ~ *γaćir* 'mule, hinny' < Turkic. Since the British *raj,* many English loanwords are also heard, e.g. *ráfal* rifle', and *biskóṭ* ~ *miskóṭ* 'biscuit'.

It is now generally recognized that Burushaski has given as well as received, contributing words to neighboring languages, especially the Indo-Aryan Shina and Khowar. These words typically lack Indo-European etymologies, while their presence in Caucasian and/or Basque testifies to Euskaro-Caucasian origin (see below). For example, Bur. *bur* '(single) hair', *-ś-puraŋ* 'mane', *-l-pur* 'eyelid' > Khowar *phur* 'hair', Shina *buuri* 'mane'(cf. Basque *buru* 'head', *bepuru* 'eyebrow'); Bur. *sisín-* 'to be clear', *sisínum* 'clear' (water), 'slender' (person), 'soft' (voice) > Shina *sisíno* 'clear' (of water) (cf. Caucasian: Chechen *c'ena* 'clean, pure', etc.); Bur. *sukúin* 'kinsman' > Shina *uskúun* 'close relative', etc. (Morgenstierne 1946: 94-95). It is also likely that the territory of the Burushaski language once extended much farther than at present, and may have contributed some words to early Indic as its speakers passed through the Hindu-Kush-Karakorum passes on their way to the Panjab. Witzel (1999: 3-5) gives examples such as Vedic *kīlāla-* 'biestings, a sweet drink' (cf. Bur. *kiláaẏ* 'curds of biestings'), Skt. *karpāsa* 'cotton plant' (cf. Bur. *gupás,* Yasin *γupás* 'cotton'), Vedic *sindhu* 'river' (cf. Bur. *sínda,* Yasin *sénde* 'river'). (See also Anderson 1999; Tikkanen 1988.)

Besides the major part of the dictionary volume (Burushaski-German), Berger also provides a glossary of Bur. proper names, a reverse glossary (German-Burushaski), a botanical index (e.g., Panicum miliaceum = *baẏ* [' a kind of millet']), and a section of special lists by subject matter (e.g., trees, child language, diseases, insult words, food and drink, etc.). Berger again follows the lead of Lorimer in providing these special lists, which will be of great help to anyone researching specific semantic fields of the Bur. lexicon.

Regarding possible external relations of Burushaski, Berger makes no commitment in this book. (Vol. 1, p. 3: "Ein genealogischer Zusammenhang mit einer anderen Sprache oder Sprachfamilie der Welt konnte bisher nicht nachgewiesen werden.") However, in some of his earlier writings (1956, 1959) Berger proposed a fair number of comparisons with Basque, which he had studied under the great vasconist, Luis Michelena (Berger, 1959, notes 3 and 4). Berger (1959, p. 26, note 34) also discovered a recurrent phonological correspondence between Burushaski initial *t-* and Basque initial *l-,* which I think is correct (Bengtson 2001, ms.). While some of Berger's Burushaski-Basque comparisons may prove to be invalid, there are several which I think are valid or at least promising, and I have adopted them in my own work (Bengtson 1995, pp. 98-99; Bengtson 2001; Blažek & Bengtson 1995). Blažek and I, along with Greenberg & Ruhlen (1992), Peiros (1988), and Starostin (1996), think that Burushaski is most likely to be connected with Euskaro-Caucasian (= Macro-Caucasian = Basque + North Caucasian + Burushaski), and, at a deeper level, with Dene-Caucasian (= Sino-Caucasian). But much more needs to be done to make this hypothesis convincing to most linguists.

This masterwork by Berger has been long awaited by everyone interested in the Burushaski language, and it fulfills our expectations.

References:

Anderson, Gregory D.S. 1999. "M. Witzel's 'South Asian Substrate Languages' from a Burushaski Perspective." *Mother Tongue* (Special Issue, October 1999).

---------- forthcoming a. "Burushaski Morphology." In *Morphologies of Asia and Africa,* ed. by Alan Kaye. Winona Lake, IN: Eisenbrauns.

---------- forthcoming b. "Burushaski." In *Language Islands: Isolates and Microfamilies of Eurasia,* ed. by D.A. Abondolo. London: Curzon Press.

Bashir, Elena. 2000. "A Thematic Survey of Burushaski Research." *History of Language* 6.1: 1-14.

Bengtson, John D. 1995. "Basque: An orphan forever? A response to Trask." *Mother Tongue* 1: 84-103.

---------- 1997. "Ein Vergleich von Buruschaski und Nordkaukasisch." *Georgica* 20: 88-94.

---------- 1999. "A Comparison of Basque and (North) Caucasian Basic Vocabulary." *Mother Tongue* 5: 41-57.

---------- 2000. Review of Čašule 1998. *History of Language* 6.1: 22-26.

---------- 2001. "Genetic and Cultural Linguistic Links between Burushaski and the Caucasian Languages and Basque." (Paper presented at the 3rd Harvard Round Table on Ethnogenesis of South and Central Asia, Harvard University, May 13, 2001.)

---------- (ms.) "The Dene-Caucasian Lateral Affricates."

Berger, Hermann. 1956. "Mittelmeerische Kulturpflanzennamen aus dem Burušaski." *Münchener Studien zur Sprachwissenschaft* 9: 4-33.

---------- 1959. "Die Burušaski Lehnwörter in der Zigeunersprache." *Indo-Iranian Journal* 3.1: 17-43.

---------- 1974. *Das Yasin-Burushaski (Werchikwar).* (= Neuindische Studien herausgegeben von Hermann Berger, Lothar Lutze, und Günther Sontheimer, Band 3.) Wiesbaden: Otto Harrassowitz.

Blažek, Václav, and John D. Bengtson. 1995. "Lexica Dene-Caucasica." *Central Asiatic Journal* 39.1: 11-50; 39.2: 161-164.

Čašule, Ilija. 1998. *Basic Burushaski Etymologies: The Indo-European and Paleo-Balkanic Affinities of Burushaski.* (LINCOM Etymological Studies 01.) Munich:LINCOM Europa.

Greenberg, Joseph H., and Merritt Ruhlen. 1992. "Linguistic Origins of Native Americans." *Scientific American* 267.5: 94-99. (November 1992)

Lorimer, D.L.R. 1935-1938. *The Burushaski Language* (3 vols.). Oslo: Instituttet for Sammenlignende Kulturforskning.

Morgenstierne, Georg. 1945. "Notes on Burushaski Phonology." *Norsk Tidsskrift for Sprogvidenskap* 13: 61-95.

Peiros, Ilia. [Пейрос, И.]. 1988. "Синокавказкская теория и бурушаский язык." [The Sino-Caucasian theory and the Burushaski language.] *Проблемы изучения сравнительниого-исторического языкознания и лингвистической историй о востоке и юге азий*, pp. 214-227. Moscow: Institute of Far Eastern Studies, Academy of Sciences, USSR.

Starostin, Sergei A. 1996. "Comments on the Basque-Dene-Caucasian Comparisons." *Mother Tongue* 2: 101-109.

Tiffou, Étienne. 2000. "Current Research in Burushaski: A Survey." *History of Language* 6.1: 15-20.

Tikkanen, Bertil. 1988. "On Burushaski and other ancient substrata in northwest South Asia." *Studia Orientalia* 64: 303-325.

Turner, R.L. 1966. *A Comparative Dictionary of the Indo-Aryan Languages.* London: Oxford University Press.

Varma, Siddheshwar. 1941. "Studies in Burushaski Dialectology." *Journal of the Royal Asiatic Society of Bengal, Letters* 7: 133-173.

Witzel, Michael. 1999. "Early Sources for South Asian Substrate Languages." *Mother Tongue* (Special Issue, October 1999): 1-70.