In Hot Pursuit of Language in Prehistory

Essays in the four fields of anthropology

Edited by John D. Bengtson

John Benjamins Publishing Company

In Hot Pursuit of Language in Prehistory

In Hot Pursuit of Language in Prehistory

Essays in the four fields of anthropology

In honor of Harold Crane Fleming

Edited by
John D. Bengtson

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

In hot pursuit of language in prehistory : essays in the four fields of anthropology in honor of Harold Crane Fleming / edited by John D. Bengtson.
p. cm.
Includes bibliographical references and index.
1. Comparative linguistics. 2. Language and languages--Origin. 3. Human beings--Origin. I. Bengtson, John D.
P143.I36 2008
306.44--dc22 2008030839
ISBN 978 90 272 3252 6 (Hb; alk. paper)

© 2008 – John Benjamins B.V.
No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Foreword ix
Acknowledgments xiii
Photographs xv
Works of Harold Crane Fleming xix

Part I. African peoples

Geography, selected Afro-Asiatic families, and Y chromosome lineage variation: An exploration in linguistics and phylogeography 3
Shomarka Omar Keita

A dental anthropological hypothesis relating to the ethnogenesis, origin, and antiquity of the Afro-Asiatic language family: Peopling of the Eurafrican-South Asian triangle IV 17
Christy G. Turner II

African weeks 25
Daniel F. McCall

Part II. African languages – Synchronic studies

Gender distinction and affirmative copula clauses in Zargulla 39
Azeb Amha

Riddling in Gidole 49
Paul Black

Part III. African languages – Classification and prehistory

A lexicostatistical comparison of Omotic languages 57
Václav Blažek

The primary branches of Cushitic: Seriating the diagnostic sound change rules 149
Christopher Ehret

Erosion in Chadic 161
Herrmann Jungraithmayr

On Kunama *ukunkula* "elbow" and its proposed cognates in Nilo-Saharan languages 169
Philippe Bürgisser

The problem of pan-African roots 189
Roger M. Blench

Part IV. Languages of Eurasia, Oceania, and the Americas

Some thoughts on the Proto-Indo-European cardinal numbers 213
Allan R. Bomhard

Some Old World experience of linguistic dating 223
J.A. Janhunen

The languages of Northern Eurasia: Inference to the best explanation 241
John D. Bengtson

Slaying the Dragon across Eurasia 263
Michael Witzel

Trombetti: The forefather of Indo-Pacific 287
Jonathan Morris

Otomanguean loan words in Proto-Uto-Aztecan maize vocabulary? 309
Jane H. Hill

Historical interpretations of geographical distributions of Amerind subfamilies 321
Larry Lepionka

Part V. Human origins, Language origins, and Proto-Sapiens language

Current topics in human evolutionary genetics 343
Stephen L. Zegura

A wild 50,000-year ride 359
Philip Lieberman

Can Paleolithic stone artifacts serve as evidence for prehistoric language? 373
Ofer Bar-Yosef

The origin of language: Symbiosism and symbiomism 381
George van Driem

Some speculations on the evolution of language, and the language of evolution 401
Paul Whitehouse

The age of Mama and Papa 417
Alain Matthey de l'Etang & Pierre J. Bancel

The millennial persistence of Indo-European and Eurasiatic pronouns and the origin of nominals 439
Pierre J. Bancel & Alain Matthey de l'Etang

General index 465
Index of languages and language families 471
Index of scholars discussed 475

Foreword

Harold Crane Fleming (hereafter, "Hal") was born the 23rd day of December, 1926, in Winsted, Connecticut. Hal was the youngest of four children to parents of Scottish and New England Yankee stock. Hal was born into a wealthy home, but soon the Great Depression arrived and his father lost his fortune, his parents divorced, and Hal and a sister were supported by his mother, who worked as a "sales girl."

By his own account Hal "flunked out of kindergarten" because he hated to read. But this propensity was soon to change radically, because across the street from the family apartment in Winsted "was a good public library with circa 20,000 volumes. Where else could a lonely little boy with no money and no friends hang out after school? It took very little time to get over the hatred for reading because some kind of paradise had just opened up. The library was rich in history books, geography books, historical novels and stories. Thus was born the makings of an anthropologist with an interest in historical approaches. Winsted with its many ethnic groups living harmoniously together provided the other element, an interesting society or the last element needed to make an anthropologist."[1]

After attending elementary and secondary schools in Winsted Hal was drafted into the Navy in early 1945 and trained as a radio operator in the amphibious forces which were meant to invade Japan later on that year. Instead the atomic bomb intervened and saved Hal from what could well have been a violent death or injury. Also fortunate for him was a program called the GI Bill of Rights that paid for four years of college for military veterans.

Hal enrolled in Yale University (Yale College; Jonathon Edwards) and attained the Bachelor of Arts degree in 1951. Hal continued his post-graduate studies at Yale (1953–1963), passed the comprehensive examinations in anthropology (1955) and was admitted to candidacy for the Ph.D. (now = MPhil) the following year. By this time Hal was married, to Barbara Anthony, and had a child, Leslie, so it was necessary to support the family by working 20 hours a week for a land surveyor. This skill also served him later when he served as Chief of Party of a land survey group in Ethiopia (1958–1959).

---

1. This is excerpted from Hal's "Autobiography of a Lucky Man."

After field research, supported by the Ford Foundation, in Ethiopia, Kenya, Tanganyika, Uganda, Congo, and Rwanda (1957–1960), Hal was admitted to Graduate School at the University of Pittsburgh, passed comprehensive examinations in anthropology in 1964, and was granted the degree of Doctor of Philosophy in 1965. His dissertation topic for the Ph.D. at Yale and Pittsburgh was "The Age-Grading Cultures of East Africa: An Historical Inquiry." Hal's major doctorate adviser, at Yale and at Pittsburgh, was George Peter Murdock. During the latter part of his graduate studies Hal served as a Graduate Assistant in the Department of Anthropology, University of Pittsburgh and as Ogden Mills Fellow, American Museum of Natural History, New York.

In 1965 Hal began his long career at Boston University, first as Assistant Professor of Anthropology, then as Associate Professor of Anthropology and Research Associate in the African Studies Center, Boston University (1971–1988), and continuing to this day as Research Fellow in the African Studies Center and Emeritus Professor of Anthropology, Boston University (since January 1989). At various times Hal taught the following courses: Anthropology (Introductory, Cultural, and Social), Primitive Religion, Ethnology of India, Ethnology of the Middle East, History of Anthropology, Theory and Method (Anthropology), Peoples and Cultures of Africa, Ethnology of Northeast Africa, Languages of Africa, Historical Linguistics,[2] Description (Field Methods) in Anthropological Linguistics.

Early in his career at Boston University Hal published a paper that outlined an important taxonomic discovery, his proposal that what had up to then been known as the "Western Cushitic" language family was not a part of Cushitic at all, "but rather constitutes a sixth primary branch of Afro-Asiatic, for which he suggested the name Omotic."[3] Solving taxonomic problems with African languages, and worldwide, has continued to be a major theme of Hal's work.

In August of 1986 Hal had an experience that came to significantly shape his activities for the following two decades. While attending the Ninth International Conference of Ethiopian Studies in Moscow he "accidentally" met the young members of the "Moscow Circle" of historical linguists.[4] Hal was deeply impressed by the "long range

---

2. With internal focus on Phylum Linguistics (also called Paleolinguistics or Prehistoric Linguistics).

3. "The Classification of West Cushitic within Hamito-Semitic." *Eastern African History*. *Boston University Studies in African History*, ed. by D. McCall, N. Bennett, and J. Butler, III: 3–27. 1969. The quote is from Merritt Ruhlen, *Guide to the World's Languages. Vol. 1: Classification.* Stanford University Press, 1987, p. 89.

4. The Moscow Circle at that time consisted of A.Y. Aikhenvald, A. Belova, V.A. Dybo, E. Khelimsky (Helimski), A.Y. Militarev, S.L. Nikolayev, I. Peiros, V. Porkhomovsky, S.A. Starostin, O. Stolbova, V. Terent'ev, T.L. Vetoshkina, N. Zhvania. As mentors they looked to I.M. Diakonov (Diakonoff), A.B. Dolgopolsky (Haifa), and the late V.M. Illich-Svitych. Another member,

linguistic probing … of scholars in Moscow who were trying to extend genetic taxonomy of human languages beyond the levels achieved in the 1950s and 1960s." Since Hal was the only American in the linguistic section of the Conference he was selected by the Moscow Circle to be their "representative" to western scholars.[5]

Beginning in the fall of 1986 Hal began discharging this duty by circulating letters to a large number (ca. 75) of linguists and anthropologists outside of Russia. The second and third letters were labeled Circulars, and by the fourth issue (November 1987) the newsletter had acquired a more formal appearance, the name *Mother Tongue*,[6] and the *Anči* symbol (the mother figure, with a ceramic jar on her head) that has graced every issue of *Mother Tongue* (Newsletter or Journal) since.

In 1989 what had been the "Long Range Comparison Club" was legally incorporated as the Association for the Study of Language In Prehistory (ASLIP), a non-profit corporation. Hal has served as President of ASLIP (1988–1996), Secretary-Treasurer (1996–98), Vice-President and Acting Treasurer (2004 – present), and Member of the Board of Directors (1998 – present). ASLIP's mission is "to encourage international, interdisciplinary information sharing, discussion, and debate among biogeneticists, paleoanthropologists, archaeologists, and historical linguists on questions relating to the emerging synthesis on language origins and ancestral human spoken languages."[7]

In 1989 and 1990 Hal took another trip to Ethiopia for linguistic field research, primarily "to fill out the parameters of the Omotic group." One of the languages discovered on this trip, Ongota (a.k.a. Birale), turned out to be a taxonomic puzzle. Some experts have considered it Nilo-Saharan (with numerous loans from Afro-Asiatic), some (including Hal) have placed it in Afro-Asiatic at some level, and others have regarded it as a mixed or pidgin language. As this volume is being prepared we hear that Hal's new book on the language has been published.[8] In Hal's words, "it features Ongota as a major sub-phylum of Afro-Asiatic and its presence as decisive in arguing for an Ethiopian homeland for that phylum."

Hal is the father of four children: Leslie, Sara, Jennifer, and Alexander. Since 1982 Hal and his wife Nancy (née Meister) have lived in the legendary fishing and quarrying

---

V. Shevoroshkin, was already in the U.S. (Ann Arbor) by this time, and Mark Kaiser was an early American associate of the Moscow Circle.

5. In reconstructing this history much is owed to a letter from Hal Fleming to Edward C. Carter (American Philosophical Society) in March 1987, as well as the circular letters mentioned below.

6. The title *Mother Tongue* was invented by V. Shevoroshkin.

7. Further particulars can be found on the ASLIP website: http://www.people.fas.harvard.edu/~witzel/aslip.html.

8. *Ongota: A Decisive Language in African Prehistory.* Wiesbaden. Otto Harrassowitz Verlag. 2006.

towns of Rockport and Gloucester, Massachusetts. Intensely interested in politics, Hal has served as a member of the Gloucester Democratic City Committee and Ward Chairman of the Democratic Party for the 4th Ward (2002–05).

As an adherent of the Four Field School of American Anthropology, Hal is conversant in physical anthropology, linguistics, archaeology, and cultural anthropology, as well as many other spheres of intellectual endeavor. He has told me that sometimes scholars in each of the four fields have not been aware of his participation in the others, thinking of him exclusively as one of themselves. This typifies another major theme of Hal's academic life: the ability to move comfortably among and through all the different disciplines touching on human prehistory, and the ability to *get scholars to talk to each other,* whether it be across the chasm between East and West, or across the sometimes impenetrable and artificial walls between scientific disciplines. We are all the richer for having been able to know and work with this remarkable man. Well done, Hal, and continued *buona fortuna!*

The origin of this book's title

"[Aharon] Dolgopolsky and his students and Carleton T. Hodge inspired me and others to join in the hot pursuit of deep human prehistory. Ducking the stones thrown by the pettifoggers, we find the whole exercise quite exciting and wonderfully fruitful. Is that not what science is supposed to be?"

Reference

Harold C. Fleming. 1999. Review of Aharon Dolgopolsky, *The Nostratic Macrofamily and Linguistic Paleontology* (1998) and I. Hegedüs, P.A. Michalove & A. Manaster Ramer (eds.), *Indo-European, Nostratic, and Beyond: Festschrift for Vitalij V. Shevoroshkin* (1997). *Anthropological Linguistics* 41:3.417-426.

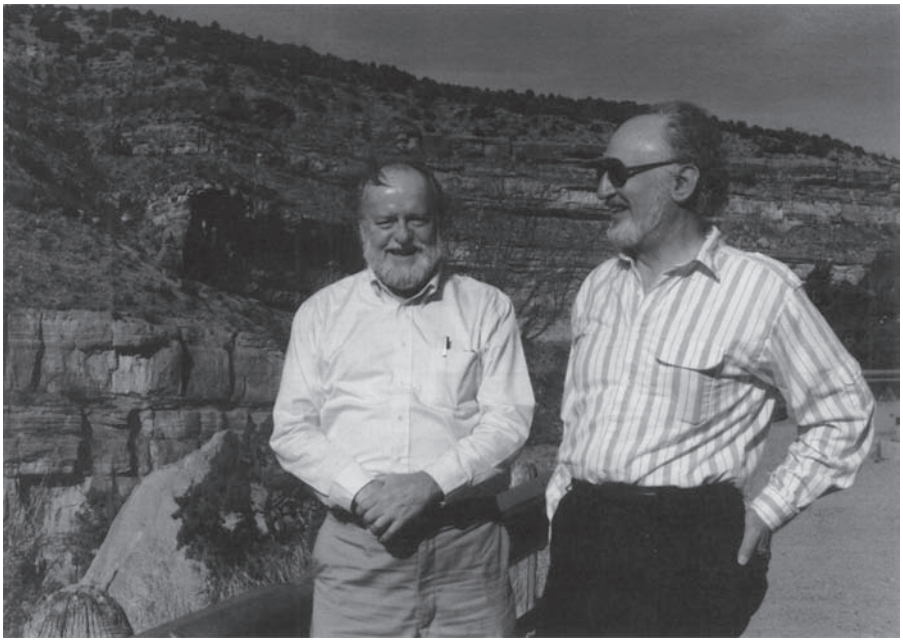
Acknowledgments

The Editor is grateful to all the contributors to this volume. Special thanks (for photographs, technical support, and editorial advice) are owed to Brita M. Bengtson, Ellen Bentzen, Allan R.Bomhard, Hal Fleming, Donald Levine, Herbert S. Lewis, Michael Witzel, and my wife, April L. Rankin.

Photographs

1. Addis Ababa, 1959. American anthropologists (left to right) Herb Lewis, Donald Levine, Hal Fleming, Marcia Lewis, and William A. Shack (1923–2000).
(Photo courtesy of Donald Levine.)

2. Uganda, May or June 1960. Herb Lewis and Hal Fleming at the Equator. (Photo courtesy of Herbert S. Lewis.)



3. New Mexico, 1992. Hal Fleming with archeologist Richard S. "Scotty" MacNeish (1918–2001). (Photo courtesy of Harold C. Fleming.)

Works of Harold Crane Fleming

Field research

1. **1957–60** Ethiopia, especially provinces of Wallega, Kafa, Sidamo, and Gemu-Gofa. Initially coffee plantations were studied *vis-à-vis* culture change in peasantry (major adviser, Sidney Mintz). Remainder of research consisted of linguistic and ethnographic surveys. In the summer of 1960, on the way home, another survey trip was made to Kenya, Tanganyika, Uganda, Congo, and Rwanda.
2. **1971–72** Ethiopia. Primarily ethnographic field research on the Dime of Gemu-Gofa, supplemented importantly by linguistic work on several selected languages of Gemu-Gofa and Kafa. The latter was necessitated by a war which raged around Dime for the whole year.
3. **1989–90** Ethiopia. Linguistic field research. Goals to fill out the parameters of the Omotic group of languages. Restrictions on travel due to war-time conditions in Ethiopia generally. Varying amounts of data, from very much to small amounts of supplementary data on selected aspects, were gathered on Diddesa Mao, Adikas and Jeba dialects of Dizi, Shinasha of Dangur, Hamar, Oyda, Ganjule, Gat'ame or Haruro, Shako, Chara, Gimira, Janjero, and Ongota (Birale). Time was also spent ascertaining that several languages were no longer spoken, to wit, Mesmes, Galila of Lake Wonchi, Garo or Boxa.

Library research, major and unpublished

1966–68 Prepared a general ethnology course on greater (British) India, participated in the informal "Harvard Lunch Group on Indology," read and discussed India with a number of colleagues, reviewed the literature on Himalayan language classification, made new classifications of Indian languages. Prepared grant applications for field research in Uttar Pradesh, Nepal and Maharashtra. All successful up to the point of getting research cooperation from Indian colleagues.[1]

---

1. Washed out by the "Berreman affair," i.e., alarm among Indian academics that anthropologists from the USA might be CIA agents and the subsequent insertion of this into Indian national electoral politics. So I refocused research energies on Africa and the Middle East. Still much of this South Asian research later re-surfaced in ASLIP publications [HCF].

Publications

Books

2006. *Ongota: A Decisive Language in African Prehistory.* Wiesbaden. Otto Harrassowitz Verlag.

Articles and review articles

1964. "Baiso and Rendille: Somali Outliers." *Rassegna di Studi Etiopici* XX. 35–96.

1968. "Ethiopic Language History: Testing Linguistic Hypotheses in an Archeological and Documentary Context." *Ethnohistory* 15. 353–388.

1969a. "The Classification of West Cushitic within Hamito-Semitic." *Eastern African History. Boston University Studies in African History*, ed. by Daniel McCall, Norman Bennett, & Jeffrey Butler, III. 3–27.

1969b. "Asa and Aramanik: Cushitic Hunters in Masai-land." *Ethnology* VII. –36.

1973. "Sub-Classification in Hamito-Semitic." *Lexicostatistics in Genetic Linguistics. Proceedings of the Yale Conference. Yale University, April 3–4, 1971,* ed. by Isidore Dyen, 85–88. Mouton: The Hague.

1974. "Omotic as an Afroasiatic Family." *Studies in African Linguistics, Supplement 5,* October. 81–94.

1975. "Recent Research in Omotic-Speaking Areas." *Proceedings of the First United States Conference on Ethiopian Studies, 1973,* ed. by Harold G.Marcus, 261–278. Michigan State University, East Lansing.

1976a. "Cushitic and Omotic." *Language in Ethiopia,* ed. by M. Lionel Bender, et al., 34–53. [Originally scheduled for 1972!]

1976b. "Sociology, Ethnology, and History in Ethiopia." Review article of *Greater Ethiopia: The Evolution of a Multiethnic Society* by Donald N. Levine. *International Journal of African Historical Studies* IX. 248–278.

1976c. "Omotic Overview." *The Non-Semitic Languages of Ethiopia,* ed. by M.L. Bender & John Hinnant, 299–323. East Lansing, Mich.: Michigan State University.

1976d. "Kefa (Gonga) Languages." *The Non-Semitic Languages of Ethiopia,* ed. by M.L. Bender & John Hinnant, 351–376. East Lansing, Mich.: Michigan State University.

1978. "Ethiopians and East Africans." Review Article *of Ethiopians and East Africans: The Problem of Contacts* by Christopher Ehret. East Africa House, Nairobi. *International Journal of African Historical Studies* XI: 2. 267–282.

1980. "Microtaxonomy: Language and Blood Groups in the Horn of Africa." *Proceedings of the Fifth International Conference of Ethiopian Studies Session B. April 13–16, 1978,* ed. by Robert Hess, 25–49.

1983a. "Chadic External Relations." *Studies in Chadic and Afroasiatic Linguistics.* Papers from the International Colloquium on the Chadic Language Family and the Symposium on Chadic within Afroasiatic at the University of Hamburg, September 14–18, 1981, ed. by Ekkehard Wolff & Hilke Meyer – Bahlburg, 17–32. Hamburg: Helmut Buske Verlag.

1983b. "Kuliak External Relations: Step One." *Nilotic Studies. Proceedings of the International Symposium on Languages and History of the Nilotic Peoples, Cologne, January 4–6, 1982*, ed. by Rainer Vossen & Marianne Bechhaus-Gerst, 423–478. Berlin: Dietrich Reimer Verlag.

1983c. "Surma Etymologies." *Nilotic Studies. Proceedings of the International Symposium on Languages and History of the Nilotic Peoples, Cologne, January 4–6, 1982,* ed. by Rainer Vossen & Marianne Bechhaus-Gerst, 523–555. Berlin: Dietrich Reimer Verlag.

1984. "The Importance of Mao in Ethiopian History." *Proceedings of the Seventh International Conference of Ethiopian Studies*, Lund (Sweden), April 25–29, 1982, ed. by Sven Rubenson, 31–38.

1986. Circular 1 of *Mother Tongue, Newsletter of the Long Range Comparison Club*, 12 pages. (ASLIP began here as a circular letter.)

1987a. "Proto-Gongan Consonant Phones: Stage One." *Festschrift Reinisch (150th Birthday),* ed. by Hans Mukarovsky, 141–159. Vienna.

1987b. "Hadza and Sandawe Genetic Relations." *Proceedings of the International Symposium on African Hunters and Gatherers.* Published as monograph by *Sprache und Geschichte In Africa* [hereinafter SUGIA], ed. by Franz Rottland, Band 7: 2. 157–189.

1987c. Circular 2 of *Mother Tongue, Newsletter of the Long Range Comparison Club*, 21 pp. Partly editorial, partly original.

1987d. Circular 3 or *Mother Tongue 3, Newsletter of the Long Range Comparison Club,* 35 pp.

1987e. *Mother Tongue 4, Newsletter of the Long Range Comparison Club*, 40 pp. Mostly editorial, partly original.

1988a. "Towards a Definitive Classification of Human Languages", Review Article of *Guide to the World's Languages* by Merritt Ruhlen. *Diachronica* IV. 159–223.

1988b. "South Omotic Consonant Phonemes." *Proceedings of the International Symposium on Cushitic and Omotic Languages,* Bonn, January, 1986. *SUGIA Band 8*. ed. by Marianne Bechhaus- Gerst & Fritz Serisko, 12.

1988c. *Mother Tongue 5, Newsletter of the Long Range Comparison Club,* 54 pp. Mostly editorial, partly original.

1988d. *Mother Tongue* 6, Newsletter of ASLIP.

1989a. *Mother Tongue 7*, Newsletter of ASLIP. References to ASLIP publications as such end here. Up to 1989 the issues were largely writ by myself. After that, the authorships were shared. A few large contributions only are mentioned hereafter.

1989b. "Mao's Ancestor. Consonant Phonemes of Proto-Mao. Stage One." *Proceedings of the Ninth International Conference of Ethiopian Studies, Moscow, 26–29 August 1986. USSR Academy of Sciences, Africa Institute, Section IV. Linguistics*. 35–44.

1990a. "The Mother Tongue Hypothesis." *Newsletter of the Department of Linguistics (June)*. 20 pp. Addis Ababa University, Addis Ababa, Ethiopia.

1990b. "A Grammatical Sketch of Dime (Dim-Af) of the Lower Omo." *Omotic Language Studies*, ed. by Richard Hayward, 494–583. London, S.O.A.S., University of London.

1991a. "A New Taxonomic Hypothesis: Borean or Boralean." *Mother Tongue* 14, Newsletter of ASLIP. 16 pp.

1992a. *Mother Tongue Extra.* Special two page announcement of MacNeish's archeological discoveries at Pendejo Cave; discussion of American prehistory; problems of glotto-chronology.

1992b. "Shabo: Presentation of Data and Preliminary Classification." *Proceedings of the Fourth Nilo-Saharan Conference, Bayreuth, Aug. 30 – Sept. 2, 1989,* ed. by M. Lionel Bender, 389–402.

1994. The Dime of Gemu-Gofa: ethnography of a tragedy. *Études Éthiopiennes. Actes de la Xe Conférence Internationale des Études Éthiopiennes, Paris, 24–28 Aôut 1988,* ed. by C. Lepage, *et al.*, vol. 1, 449–451. Paris: Société Française des Études Éthiopiennes,

1995a. Review article of *History and Geography of Human Genes* by L.L. Cavalli-Sforza, et al. *Mother Tongue* 24.8–29. (Following short reviews by Rebecca Cann and Frank Livingstone, pp. 6–8)
1995b. "The 'Sogenannten' Ethiopian Pygmoids." *Mother Tongue* 24. 30–34.
1995c. "Two Aspects of Massive Comparisons." *Mother Tongue* 24. 65–68.
1995d. "Editorial I: Surveying the Damage" and "Editorial II: Afrasian Gets, Better, Better, and Deeper." *Mother Tongue: The Journal*, Issue I. 229–234.
1996. "Looking to the West and North: Nihali and Kusunda Find Links." *Mother Tongue: The Journal*, Issue II. 67–74.
1997a. "Call it Methodology." *Mother Tongue: The Journal*, Issue III. 51–52.
1997b. "Carleton T. Hodge as a Pioneer Long Ranger." *Carleton T. Hodge. Bibliography 1944–1997. Eurolingua,* ed. by Gyula Décsy, 31–33. Bloomington, Ind.: Eurasian Linguistic Asociation.
2000a. "Glottalization in Eastern Armenian." *Journal of Indo-European Studies* 28: 1/2. 155–196.
2000b. "Carleton T. Hodge (1917–1997): A Tribute." *Anthropological Linguistics* 42: 2. 1–4.
2000–2001. "Joseph H. Greenberg: A Tribute and an Appraisal." *Mother Tongue: The Journal*, Issue VI. 9–28.
2002a. "Shabo: A New African Phylum or a Special Relic of Old Nilo-Saharan?" *Mother Tongue: The Journal*, Issue VII. 1–38.
2002b. "Ongota Lexicon: English-Ongota" *Mother Tongue: The Journal*, Issue VII. 39–64.
2004. "Editorial Foreword," pp. 1–2; "Background Information on Australian Languages," pp. 3–6. *Mother Tongue: The Journal*. Issue IX.
2005. "In Memoriam: F. Livingstone & S. Starostin & JHG," pp. 1–7; "The Wee Fossil Folk of Sundaland," pp. 8–13; "The Great Archeological Debate on Human Origins," pp. 14–39; "Out of Africa: Crossing the Bab el Mandeb as Early as 125,000 BP" + "Fossil Homo sapiens in Ethiopia circa 195,000 BP," pp. 40–50; "Archeology in the Great Rain Forests? A Surprise," pp. 59–73; "Chinese Geneticists report on China's (human) Genome," 74–84; "With a Biblical Hypothesis as Preamble," p. 85; "Comments on Trombetti by a Modern Long Ranger," p. 108; "Are Linguists Inherently Conservative?", p. 151. All in *Mother Tongue: The Journal.* Issue X.
2006. "Introduction to Issue 11 of *Mother Tongue*. Asian Remnant Languages and the Year of the Australoid. A Conference held at Harvard University, October 21–22, 2006," pp. 1–22. *Mother Tongue: The Journal*. Issue XI.

Co-authored articles

1979. Feinhandler, Sherwin J. & Harold C. Fleming. "Pre-Columbian Tobaccos in the Pacific," *Economic Botany* 33: 2. 213–226.
1992. Fleming, Harold C., et al. "Ongota (or) Birale: A Moribund Language of Gemu-Gofa (Ethiopia)." *Journal of Afroasiatic Languages* 3: 3. 181–225. (But actually appeared in 1994.) Co-authors Aklilu Yilma, Ayyalew Mitiku, Richard Hayward, Yukio Miyawaki, Pavel Mikeš, J. Michael Seelig. (The team represents five different countries.)
1993. McCall, Daniel, & Harold C. Fleming. "The Pre-Classical Circum-Mediterranean World: Who Spoke Which Languages." *Mother Tongue* 21. 1–15.
1998. An up-dated and slightly revised version of the above paper has been published in *Language and Archeology,* ed. by Roger Blench & Mathew Sprague. Cambridge University Press.
2003. "Tribute to Joseph H. Greenberg." Daniel McCall, Herbert S. Lewis, & Harold C. Fleming. *Sprache und Geschichte in Afrika.*

Note: All *Mother Tongue* references are to what is technically *Mother Tongue: The Newsletter* of which 30 issues have been published. Only a few references are made to things written in *Mother Tongue: The Journal* (since 1995).

Book reviews, short comments and rejoinders

1961. Review of Ad.E. Jensen, ed., *Altvölker Süd-Äthiopiens*. *American Anthropologist* 63. 615–616. (With Herbert S.Lewis)

1963. Rejoinder to Eike Haberland. *American Anthropologist* 65.

1966a. Review of A.M. Jones. "Africa and Indonesia: The Evidence of the Xylophone and other Musical and Cultural Factors." *Ethnomusicology*.

1966b. Review of Helmut Straube. *Westkuschitische Völker Süd-Äthiopiens*. *American Anthropologist* 68.

1968. Review of P.E.H. Hair. *The Early Study of Nigerian Languages*. *African Historical Studies* I. 285–288.

1970. Review of Wolf Leslau. *Ethiopians Speak: Soddo*. *American Anthropologist* 72.

1971. Review of Samuel Baker. *The Nile Tributaries of Abyssinia*. Johnson Reprint Company, 1967. *Ethnohistory* 18.

1972. *CA Treatment = simultaneous review. M.E. Jeffries. "Maize and the Mande Myth." *Current Anthropology.*

1973. *CA Treatment = simultaneous review. D. McCracken. "Lactase Deficiency." *Current Anthropology.*

1974. Review of Gillian Brown. *Phonological Rules and Dialect Variation: A Study of the Phonology of the Lumasaaba* (*Uganda*). *American Anthropologist* 76.

1976. Review of William Shack & Habte-Mariam Marcos. *Gods and Heroes: Oral Traditions of the Gurage of Ethiopia*. *International Journal of African Historical Studies* VIII. 165–169.

1978. Review of Jean Buxton. Book on Mandari of southern Sudan. *American Anthropologist.*

1980. Review of Ivo Strecker & Jean Lydall. *The Hamar of Ethiopia*. Vols. I, II, III. *Canadian Journal of African Studies.*

1983. Review of Christopher Ehret. *The Historical Reconstruction of Southern Cushitic Phonology and Vocabulary. Journal of African Languages and Linguistics* 5. 93–97.

1988. Review of David Dalby & Michael Mann. *A Thesaurus of African Languages*: *A classified and annotated inventory of the spoken languages of Africa*. *International Journal of African Historical Studies* XX.

1994a. Reply to Marcello Lamberti's attack on the Omotic hypothesis. *Anthropos.*

1994b. Rejoinder to Marcello Lamberti's rebuttal to 1994a. *Anthropos.*

1994c. Review of Eike Haberland. *Hierarchie und Kaste: Zur Geschichte und Politischen Struktur der Dizi in Sudwest-Aethiopien. Studien zur Kulturkunde. Stuttgart*: Franz Steiner Verlag, 1993. *International Journal of African Historical Studies* XXVI. 421–426.

1995. Review of Merritt Ruhlen. *The Origin of Language: Tracing the Evolution of the Mother Tongue.* 1994. New York: John Wiley & Sons, Inc. *Anthropological Science: Journal of the Anthropological Society of Nippon* 103: 4. 141–153.

1996. Review of W. Wilfried Schuhmacher, F. Seto, J. Villegas Seto & Juan R. Francisco. 1992. *Pacific Rim : Austronesian and Papuan Linguistic History.* Heidelberg: Carl Winter Universitäts-verlag. In *Word.*

1999a. Review of Aharon Dolgopolsky. *The Nostratic Macrofamily and Linguistic Palaeontology.* Cambridge: The McDonald Institute for Archaeological Research. (Along with the book below)

1999b. Review of *Indo-European, Nostratic, and Beyond: Festschrift for Vitalij V. Shevoroshkin***.** Edited by Irén Hegedüs, Peter A. Michalove, & Alexis Manaster Ramer. *Journal of Indo-European Studies*, Monograph 22. Washington, D.C.: Institute for the Study of Man, 1997. *Anthropological Linguistics*.

2002a. Review of *Somali*. by John Saeed, 1999. Amsterdam/Philadelphia: John Benjamin's Publishing Company. *Word*.

2002b. Review of *A Rendille Dictionary: including a Grammatical Outline and an English-Rendille Index***.** Steve Pillinger and Letiwa Galboran. 1999. Köln: Rüdiger Köppe Verlag. *Word*.

2003a. Review of *Phonological Theory and the Dialects of Italy*. Lori Repetti, ed. 2000. Amsterdam/Philadelphia: John Benjamins Publishing Company.*Word*.

2003b. Review of *Comparative Dravidian Linguistics*. *Current Perspectives*. Bhadriraju Krishnamurti. 2001. Oxford University Press. *Word*.

PART I

African peoples

Geography, selected Afro-Asiatic families, and Y chromosome lineage variation

An exploration in linguistics and phylogeography

Shomarka Omar Keita
National Human Genome Center at Howard University
Washington, DC/Department of Anthropology Smithsonian Institution

While it is known and must be emphasised that language, biology, and culture do not travel as an obligatory package, this does not mean that congruence never exists and may be historically meaningful. In this paper the published Y chromosome data and branches of the Afro-Asiatic language family for which good genetic data exist were examined in an exploratory fashion to determine if there are patterns suggestive of overlap. Most of the Afro-Asiatic speakers shared the lineage defined by Yap descendant called PN2/215/M35. It was further found that that a key lineage – the M35/78 was shared between the populations in the locale of original Egyptian speakers and modern Cushitic speakers from the Horn. Amazigh (Berber) speakers had a high frequency of M35/81. Semitic speakers in the Near East had a higher frequency of non-Yap lineages defined by the M89 mutation, but some M35/M78 is found at a lower frequency in the Near East (vs supra-Saharan Africa). The data for Omotic and Chadic speakers is of poor quality due to sample size. The paper discusses some possible implications of the observed pattern and the need for better data.

The interest in synthesizing or comparing different lines of biological and cultural data in the exploration of historical topics has become prominent in the work of several disciplines. Perhaps most notable is the work of some geneticists who have explored patterns of genetic variation in relationship to artifacts and language family distributions. Patterns of congruence between the spatial patterning of linguistic and genetic variation have frequently been observed depending on which genetic markers are used in the analyses (Cavalli-Sforza 1988; Sokal et al. 1992). This has led to the suggestion that "parallel linguistic and allele frequency changes were not the exception, but the rule" (Barbujani 1997). This general claim has been challenged (e.g., Nettle & Harris 2003). In any assessment of gene-language correspondence the time of emergence and speed of evolution of the language *and* genetic variants under

study have to be considered, as well as demographic and historical processes. Languages can cross population boundaries, no matter how defined, and this adds complexity to the interpretation of any observed patterns. The relative paucity of data, and the extinction of past populations (via any mechanism) is also to be considered, and together with the issue of the pace of linguistic and genetic change makes the construction of historical narratives a challenge.

This exploratory effort will not engage in any polemic concerning the validity of attempting to construct "laws" about the co-variation of genes and languages, and their geographical spread. Rather it will present Y chromosome data from the populations/regions in which some branches of Afro-Asiatic (Afrasian) were apparently spoken in the mid-Holocene, before the periods in which clearly differentiated Semitic languages were introduced into supra-Saharan Africa by the Phoenicians, and into the Horn by South Arabians, the standard view that reflects an interpretation of the data that indicates that Semitic originated/emerged as a distinct entity in the Near East. (Hudson [1977] and Murtonen [1967], suggest an emergence in the Horn of Africa.) The goal is to examine the geographical patterning of linguistic and genetic variation, and discuss the results in terms of various models of population movement that would be most consistent with the findings, and that can be related to the geographical origins of the earliest Afro-Asiatic speakers and the question of the role of population migration in the history and development of the language family versus just the spread of the language. However, no absolute claims will be made because it is believed that the current data are insufficient to make any definitive statements.

1. Y chromosome characteristics and nomenclature

The genetic data used here are from published and unpublished sources that have analysed the Y chromosome. This chromosome is exclusively passed from father to son, and tends to accumulate mutational variation slowly in comparison to other chromosomes. The genetic variants of interest come from the mid-section comprising 95 percent of the Y's length, and is called the non-recombining portion of the Y chromosome (NRY); this means that in cell division this portion does not exchange genetic material with any other chromosomes. Therefore variation found there is due strictly due to mutations found in the male line of descent, and occur only within males. The tips of the Y chromosome, called pseudo-autosomal regions, do exchange genetic material with the X chromosome that does come from the mother.

The DNA in the NRY segment has a useful characteristic for those interested in genealogy. Mutations there have a low rate of occurrence, and hence are called unique event polymorphisms (UEPs). "Unique" here signifies what is believed to usually be *a one time* occurrence with little or no mutational reversal back to the previous state, such that

"in samples of realistic size, they [UEPs] unambiguously define related groups of chromosomes, termed haplogroups" (Stumpf & Goldstein 2001). The males sharing one or more of these UEPs do in fact share a common male ancestor with each other. These men are genealogically related, thus making possible a kind of history. The father-to-son male line of descent of the NRY segment does not provide any information about maternal ancestry, or the male ancestry that comes via the maternal side; the line of descent reflected is strictly uniparental. It is important to note that these UEPs are regarded as being *neutral* genetic variants, and not subject themselves to natural selection.

These UEPs are called biallelic markers and in most general terms denote lines of descent, that can be refined into smaller branches depending on the number of markers looked for by the laboratory. However, in some cases there are no further known markers "downstream" from one marking a recognized lineage, and in this case the lineage can be said to be "underived" which means that the individual has the key marker defining the line or subline of descent and no others that allow it to be further characterized. In theory, assuming little migration and no history of unusual demographic events, a population/region that has the highest frequency of individuals with such markers would be a more likely place of origin of the marker. However this contention should be taken as a "rule."

The nomenclature of the reconstructed NRY taxonomy/genealogy has varied from researcher to researcher, with some general standardization now being widely accepted (Y Chromosome Consortium 2002). In the appendix is an outline of the Y chromosome genealogy as it is now known with key defining genetic markers and the alphabetical nomenclature now commonly used by most researchers. Various names were given to the lines of descent or clades at one time. The system that has now been widely adopted uses letters to name the genealogies (haplogroups) defined by specific markers. From a taxonomic point of view all of these haplogroups are not really of the same rank, because although recognized separately some of them are actually sublineages of larger genealogies and named haplogroups (see the Appendix).

2. Data and methods

The language families or subfamilies considered here are from classifications most commonly used in linguistic work. Afro-Asiatic is usually divided into the following families: Omotic, Cushitic, Chadic, Ancient Egyptian-Coptic, Berber, and Semitic; sometimes Beja is separately acknowledged, removed from the Cushitic branch, and a new division has been added Ongota (Fleming, personal communication). As noted the focus here is on the geography of linguistic branches and the regional descendant populations in the mid-Holocene. This is because Ancient Egyptian/Coptic is not a spoken language, and it is well attested that Phoenicians, south Arabians, and Arabic

speakers migrated into the African continent. There is no evidence that regional populations have been completely or nearly biologically replaced, although in some communities there may have been a much greater impact of migrants.

Presented here are genetic data from the regions in which Cushitic, Egyptian, Berber, and Semitic were spoken to the best of current knowledge. Populations from homeland regions of these language families are represented in Y chromosome studies, albeit inadequately in an absolute sense. Sampling these locales is very important as they likely indicate the source populations of the predominant genetic markers in the region at this time, and likely provide clues to when populations and culture migrated, and in what directions.

The method used here consists of performing a spatial comparison of the distributions of language groupings and Y chromosome lines of descent. It is important to note that the frequency of markers reported in a sample does not necessarily reflect the number of individuals in the founding population who carried that mutation. It only means that in the sample *examined* a certain frequency of individuals have ancestry that can be traced back to men having the DNA variant under discussion. Various demographic and evolutionary processes could lead to a changed frequency even if the mutation is neutral, and not under the influence of direct natural selection. For example, if the sample under evaluation is from a population that once was made nearly extinct by malaria it may be that the individuals who survived had genes that conferred survival, but also had – purely by chance – a particular Y chromosome lineage, and other neutral markers. This Y lineage (and other neutral genes) would then increase in number due to this accident; this situation describes what is called a *selective sweep*. Another example would be situations where social custom permits certain men to have many mates. They would skew future gene frequencies because of their greater reproductive success, assuming all other things to be equal. Differential settlement patterns by migrants would be expected to have a greater effect in those regions. These examples illustrate why it is important to have as much data as possible, including historical information, from as many samples from different regional populations as fit the study criteria. This coupled with good models makes a reasonable interpretation possible. Various mathematical models can sometimes help in determining the range of factors that may have impacted upon gene frequencies, but these will not be presented here.

3. Results: Y variants' distributions in relationship to language units

Tables 1–4 present summary data from various studies and laboratories. While no claim is made that it is exhaustive, some confidence can be expressed in its being fairly representative of the range of results that exists in the literature. The actual markers are

given instead of haplogroup names, but these latter can be found in the appendix. It is important to note unless otherwise specified that the percentage of lineages connoted by a marker combines both the underived and derived lineages of descent.

Table 1. Cushitic region or speakers

	N	M91	M60	RP54Y	M174	M96*	PN2*	M35	M2	M33+M75	M89
1.	201	0.5	1.0	0	0	0	0	80.6	1.5	0.5	15.9
2.	37	24.3	13.5	0	0	0	0	62.2	0	0	0
3.	78	10.3	1.3	0	0	0	15.4	62.8	0	1.3	0
4.	43	7	7	0	0	0	0	**37.0	42	2	5
5.	22	41	0	0	0	0	18	32	0	0	10
6.	88	13.6	10.2	0	0	0	18.2	31.8	**3.4	17	5.6
7.	48	14.6	2.1	0	0	0	10.4	35.4	3.4	0	35.5
8.	126	11.9	1.6	0	0	0	13.6	52.5	0	0.8	19.9

* Underived markers (no further characterizing markers found)
1. Somali, Sanchez et al. 2005
2. Oromo, Scozzari et al. 1999
3. Oromo, Semino et al. 2002
4. Wairak = Iraqw ? due to use of Bantu prefix, Luis et al. 2004
5. Ethiopian Jews, Cruciani et al. 2002
6. "Ethiopians", Semino et al. 2002
7. Amhara, Semino et al. 2002
8. Amhara + Oromo, Semino et al. 2002
** The lineages in these samples were all underived at this marker

The Cushitic region/speakers (Table 1) show in aggregate a high frequency of PN2/M35 of haplogroup E, with the samples of Somali and Oromo having the greatest percentages. In the Horn of Africa M35 derivatives (or more fully PN2/215-M35) are primarily of the M78 subgroup of M215/M35, but there are smaller frequencies M35/123, and others. The Cushitic domain includes notable percentages of M91 (haplogroup A) and sometimes M60 (haplogroup B), both essentially restricted to Africa, and as an aside found in some Khoisan speaking samples in substantial frequencies. Some samples evince substantial frequencies of an underived PN2, the marker "upstream" (ancestral) to M35 and M2, the latter being the marker that is most common in tropical Africa. The M89 lineages, mainly of haplogroup J, a subgroup of F, are of southern Arabian origin. (M89 is believed to have likely arisen in the Near East, although depending on the age of the ancestral mutation it may have arisen in Africa and was taken to the Near East; subsequent population events could have obscured this origin.) The Amharic groups although speaking a Semitic

language, are included here. This is because it is presumed that they are the descendants primarily of Cushitic speakers, who were biogeographically indigenous to the Horn. In other words it is not thought that an immigrant southern Arabian ethnic entity now called Amhara intermarried locally to the point of nearly complete assimilation, although this is possible. Rather local groups adopted Semitic speech and biologically assimilated the migrants, who were likely traders. Of course it is not actually known; only inferences can be made using a totality of data from archaeology, linguistics, and biology.

Table 2. Egyptian region (locale of the ancient Egyptian branch)

	N	M91	M60	RP54Y	M174	M96*	PN2*	M35	M2	M33+M75	M89
1.	58	3.4	0	0	0	0	0	54.4	6.9	1.7	34.3
2a.	25	0	28	0	0	0	0	60.0	0	0	12
2b.	21	0	43.1	0	0	0	0	38.3	0	0	19.1
2c.	46	0	34.8	0	0	0	0	50.0	0	0	15.2
3.	92	3	2	0	0	1	0	43	3	1	45
4.	147	3	0	0	0	0	0	36	2	1	58

* Underived markers (no further characterizing markers found)

1. Egypt, Hammer (n.d.)

2a. Northern Egypt, Scozzari et al. 1999

2b. Southern Egypt, Scozzari et al. 1999

2c. Combined Northern & Southern, Scozzari et al. 1999

3. Egypt, Hammer (n.d.)

4. Sample labeled Egyptian Arab, Luis et al. 2004

The samples from Egypt are almost as diverse as those of the Horn. Some samples have high frequencies of the ancient M60 marker (haplotype B), but the plurality of studies show M35 lineages as predominating. Traces of M2 are found as they are in the Horn. In Egypt the M35 are also primarily M35/M78 with some M35/M81. Egyptian samples show varying frequencies of M89 derivatives and in one sample these actually are the majority; this is not surprising given its locale and the settlement of people from the Near East during the Islamic period. It is also likely that some migration took place during the Neolithic, but this is not the major source of the Egyptian population. Migration to Egypt from the Near East also took place during the pharaonic period. Northern Egypt was more likely affected in the overall sense, but migration to major cities in upper Egypt also likely took place. The most accurate sampling of Egypt would require an extensive study of the places of settlement and good modeling that takes into account marital customs and population growth. The studies to date are tantalizing because of the diversity that they have captured: each sample is notably different.

Table 3. Berber regions or Berber speakers

	N	M91	M60	RP54Y	M174	M96*	PN2*	M35	M2	M33+M75	M89
1.	64	3	0	0	0	0	0	80	5*	2	11
2.	75	0	1.4	0	0	0	0	76	0	2.6	20
3.	29	–	–	–	–	–	–	89.5	–	–	–
4.	69	–	–	–	–	–	–	91.2	–	–	–
5.	20	–	–	–	–	–	–	80.0	–	–	–
6.	56	0	25	0	0	0	0	69.7	1.5	0	1.8

* Underived markers (no further characterizing markers found), including the M2 in #1

1. Berbers, Cruciani et al. 2002

2. North African Berbers, Alonso et al. 2005

3. Marrakesh Berbers, Cruciani et al. 2004

4. Moyen Atlas Berber, Cruciani et al. 2004

5. Mozabite Berbers, Cruciani et al. 2004

6. Moroccan Arabic speakers, Scozzari et al. 1999

Samples 3–5: studies focused on the E haplogroup and only the M35 lineages are reported here

Northwest African Berber speakers are somewhat less diverse than the Horn and Egypt. M35 lineages predominate; these are known to be primarily further characterized by the M81 marker. Non-Berber speakers in some studies have a higher frequency of M78 that might indicate descent from Arabised Nile Valley population (Egyptians). It could also indicate the substantial migration of a non-representative sample of Near Easterners, if Arabic speakers primarily came from the areas of the Levant that have a noteworthy frequency of M35/M78. A sample from Tunisia (not shown) has a frequency of 55% M89 lineages, 45% J and 10% R haplogroups; this sample is not likely representative of all of Tunisia but illustrates the historical Near Eastern and European (Roman specifically) impact on an urban population. It is important to reiterate that such numbers cannot be said to reflect the numbers of original founders, but only the number of individuals in the sample whose lineages trace back to a particular origin; many processes could affect the current frequencies, and one sample may be misleading.

The samples from the core Semitic-speaking region of the mid-Holocene, the Near East, are characterised by very high frequencies of M89 derived groups. M35 (or 215/M35) derived clades are next in frequency and decreases from west (Lebanon) to east (Syria). There are very low frequencies of M2 (PN2/M2). The diversity based on the presence of substantial frequencies of different lineages is less than that for any other region This finding may be due to a lack of studies although the large composite sample may be an accurate reflection of the overall picture. There are local populations who would likely really skew the frequencies.

Table 4. Semitic speaking regions, Non-Jewish

	N	M91	M60	RPS4Y	M174	M96*	PN2*	M35	M2	M33+M75	M89
1.	139	0	0	0	0	0	0	15	2	0	83
2.	121	3	0	3	0	0	0	7	7	2	71
3.	230	0.4	0	0.4	0	0	0	15.2	0.9	0.9	82.2
4.	24	0	0	0	0	0	0	29.2	4.2	0	66.6
5.	91	0	0	0	0	0	0	9.9	0	0	90.1

* Underived markers (no further downstream markers found)

1. Iraq, Zahery et al. 2004

2. Oman, Luis et al. 2004

3. Composite of Palestinians (n = 73), Syrians (n = 91), Lebanese (n = 24), Druze (n = 21), Saudi Arabians (n = 21)

4. Lebanese separately (n = 24), Hammer (nd)

5. Syrians separately (n = 91), Hammer (nd)

4. Discussion

The geographical patterns of the Y chromosome variation and the mid-Holocene distributions of Afro-Asiatic are interesting. Cushitic, Berber, and descendant Egyptians have in common a high frequency of M35 lineages, specifically M35/M78 and M 81. There are small frequencies of M2. The Semitic speaking region is characterized by M89 lineages. There is a declining west to east frequency of M35 in the Levantine-Syrian region.

The descendant Egyptian and Cushitic speakers also have in some samples notable frequencies of lineages sometimes called "ancient" and essentially confined to Africa. These ancient haplogroups (A and B) only sporadically occur outside of the Horn to Nile Valley corridor, and they are not seen in all samples from these regions. Nevertheless there is a relative regionality for these lines of descent in the samples in this study.

The Berber speaking northwest Africans have a very high frequency of the PN2/M35 derivative-M81, with some samples having other lineages, notably M35/M78. It does seem that, whatever population events took place in the past, M81 rose to high frequencies in Berber groups.

At another level the geographical congruence is striking for the language families and Y lineages. Essentially M35 lineages are distributed from the Horn of Africa, down the Nile Valley, and over to the Maghreb, with a spur into the Levant/Near East. Although not discussed here M35 (as M35/78) also is found in the Aegean and Balkans, but it is important to reiterate that the origin of the M35 is in East Africa. These data are potentially useful in exploring questions about the locale of the earliest population speaking earlier stages of Afro-Asiatic, to the degree that any such undertaking for

any language family is possible. M35 clades predominate based on current data in a core group of Afro-Asiatic speakers: Cushitic, Egyptian, and Berber. There are some data for Omotic that are similar in having high M35 frequencies (> 50%) and some of the "ancient" lineages but the only published sample known to the author has a mere 12 individuals. The Semitic and a few Chadic samples (not shown) are divergent from this finding for three of the locales in which distinct Afro-Asiatic families were spoken. The speakers of Chadic in northern Cameroon have a high frequency of M89 derived lineages, which have been interpreted as a footprint of a very ancient return to Africa with extreme drift suggested as the explanation of why these same populations do not have noteworthy "Eurasian" female uniparental lineages as determined by mitochondrial DNA analyses (see Coia et al. 2005). A wider sampling of Chadic speakers – and all others is needed in order to make better inferences.

What are the possible implications of these observations? Some points of interest are outlined below that could form the basis of further research projects.

1. The overall pattern is consistent with a model of the first speakers of Afro-Asiatic having emerged in or near the Horn of Africa or in the Nile Valley. The presence of Semitic in the Near East might be explained as follows. Early pre-proto Semitic speakers (? a unit of undifferentiated Boreafrasians in Ehret's conception, Ehret 1995) would have migrated into Syro-Palestine before the neolithic, being taken there by M35 bearers (specifically M35/78), and adopted by populations bearing M89 lineages. The presence of M35 lineages in the Near East is the objective evidence of this movement, although it is impossible to know the impact of Islamic period migration. This would be analogous to what happened in Ethiopia or supra-Saharan Africa, where the populations' biologies suggest that new languages were adopted by local people having a different Y chromosome heritage, and not that they were genetically swamped. The migrants would have been hunters and gatherers. The data are consistent with the linguistic evidence that indicates that proto-Afro-Asiatic speakers were not farmers but hunters and gatherers (Ehret 1995, contra the opposite opinion that they were farmers). It is likely that the first inhabitants of the late Pleistocene/early Holocene Sahara included Afro-Asiatic speakers; no food production is attested from the earliest periods there.
2. The evidence is also consistent with the biohistorical Africanity of the base populations of the Horn, Maghreb, and Nile Valley. These genetic data give population profiles that clearly indicate males of African origin, as opposed to being of Asian or European descent. It is important to say that being biogeographically African does not indicate any specific set of skin colors, hair type, or facial features; the populations were constantly subject to the forces of microevolutionary mechanisms. The E haplogroup is clearly African in origin as are A and B. The mitochondrial DNA variants in supra-Saharan and Saharan Africa show high heterogeneity, and suggest extensive Eurasian female in-migration, based on the current interpretations of where they arose. The

issue of how much Paleolithic migration from the Near East there may have been is intriguing, and the mitochondrial DNA variation may need to be reassessed as to what can be considered to be *only* of "Eurasian origin" because if hunters and gatherers roamed between the Saharan and supra-Saharan regions and Eurasia it might be difficult to determine exactly "where" a mutation arose. In any case the actual biocultural emergence of the archaeologically and historically attested Amazigh and Egyptians occurred in Africa; these peoples are not settler colonists in Africa.

3. The genetic data do not support a model of demic diffusion by farmers from the Levant to explain the Neolithic in northern or eastern Africa, or the spread of Afro-Asiatic languages into Africa (Ehret et al. 2004). It is significant that the words in Old Egyptian for sheep, goat, wheat and barley are not Semitic; nor are they Sumerian, a language isolate. Some migration no doubt occurred but a thesis of demic diffusion is unwarranted to explain the emergence of cereal agriculture in the Nile Valley or the presence of Afro-Asiatic in Africa. The latter point is supported by the evidence generated by Afro-Asiatic specialists, and is accepted nearly unanimously by scholars in the field; this may not be well known because of the lack of interest in African historical linguistics by non-specialist researchers.

4. The Amazigh (Berber) spoken today apparently shows relatively shallow time depth, based on how little it seems to have differentiated (Newman, personal communication). It is of interest that the Y variant, M35/M81, predominant among Amazigh speakers seems to have a younger origin date than M35/78 (see Cruciani 2004). It has been hypothesized by extrapolation from a genetic analysis that Berber speakers entered Africa from the Near East based on an analysis of Y chromosome-associated rapidly evolving markers (see Arredi et al. 2004). This is totally inconsistent with linguistic evidence, and there are no populations with M81 in the Near East in appreciable frequencies, that can be inferred to be pre-Islamic. There is an obvious and fundamental problem with using genetics to mark the emergence of a language family, and this cannot be overstated. This needs to be remembered. The emergence of a lineage that is now found to predominate in the speakers of a particular language or linguistic family cannot be said to be coterminous with the emergence of the language family. The tension between these ideas might be reconciled if the ancestors of present day Berber speakers are interpreted as being the descendants of restricted early Holocene populations from the northern Sahara who represented a skewed non-representative and non-random sample of the Saharan peoples who were evolving there after recolonisation with the decrease in aridity in the late Pleistocene; such a population could very well have adopted domesticated sheep and goat from the Near East and bioculturally assimilated some immigrants (see Myles et al. 2005 for a dairying hypothesis). This less arid Sahara could be modeled as the home of a reticulated Saharan metapopulation in which M81 differentiated from an underived M35; or it may have already existed before the reoccupation occurred. M81 is found in low

frequencies in the Horn in areas not known to have been visited in great numbers by historically attested Berber speakers. The morphological features (light skin and eyes, etc) of Berbers likely also developed *in situ* in the African context, at least in part, but also likely result from gene flow from Eurasia that was likely primarily maternally mediated. There is a dilemma – there is no convincing evidence to bring M35/M81 lineages from outside of Africa, yet the studies of the lineages restricted to female inheritance (defined by mitochondrial DNA) found in Berber speakers are in most studies are shared with "Eurasians" and thought to be of Eurasian origin. It is not clear necessarily that all of these lineages have to be of Eurasian origin; dates given by some investigators would place their origin very near the time of migrations from Africa. Is it possible that some variants actually arose in Africa, or arose in populations that were constantly migrating in and out of Africa such that it is impossible to say where a particular mtDNA lineage arose? Could unusual demograhic or social factors have played a role in the differential survival of these lineages, that did not affect the male lineages? To reiterate current evidence makes indigenous development a truism: the genesis and history of the Amazigh peoples as a biocultural entitiy occurred in Saharan and supra-Saharan Africa – culturally their speech community emerged there, and their biological ancestral elements merged there in the formation of breeding populations. The same is true of Nile Valley groups in terms of their locale roots along the Nile Valley and its Saharan flanking regions.

5. It is of interest that the M35 and M2 lineages are united by a mutation – the PN2 transition. This PN2 defined clade originated in East Africa, where various populations have a notable frequency of its underived state. This would suggest that an ancient population in East Africa, or more correctly its males, form the basis of the ancestors of all African upper Paleolithic populations – and their subsequent descendants in the present day. It would be tempting to relate linguistics to this, by postulating an ancestral paleo-African language. However, the current models of long range classifications do not hypothesise the existence of a proto-Pan African family that led to Khoisan, Afro-Asiatic, Nilo-Saharan, and Niger-Congo (and perhaps others now extinct). So what does this finding correlate with in the linguistic record, if anything? Although it is not proper to attempt to connect biological and cultural phenomena, or use one to "correct" the other, it may be that one discipline can be used to suggest questions about another. In this case we can ask what historical and biocultural scenarios would account best for the most basic pattern of genetic and linguistic data.

This geographical distribution of biallelic Y chromosome lineages and Afro-Asiatic branches gives researchers some insight into the speakers of Afro-Asiatic in the Holocene (Ehret et al. 2004). The spatial and biological data offer firm ground to generate a range of hypotheses when other evidence is considered, and simulation studies would be of great interest.

Acknowledgments

I would like to thank Harold Fleming for inviting me to participate in his Festschrift. Kudos to you Harold for a life well spent in the pursuit of knowledge to further a positive understanding of the African past. I lament that we do not have a legion of historical linguists who work on African languages, and predominantly from the continent. Thanks also go to Mike Hammer and his associates for sharing their data with me. Some of these data have been published, but not in the format in which I used them. My supervisors at Oxford, Tony Boyce and John Baines, continue to be a part of my life, and I thank them for their insights.

Appendix

Human Y Chromosome Genealogy in Outline form (mutations on the left in bold; not all mutations that have been discovered/defined are listed for each level)

- **I. M91** – Haplogroup A
- **II. SRY 1083.1/ M94 /M139** – Haplogroups B-R all share these mutations
  - **A. M60** – Haplogroup B
  - **B. M168** – Haplogroups C-R share this mutation
    - **1. RPS4Y** – Haplogroup C
    - **2. YAP** – Haplogroups DE
      - **a. M174** – Haplogroup D
      - **b. M40/M96** – Haplogroup E
        - **i. M33**
        - **ii. M75**
        - **iii. PN2**
          - **a. M2**
          - **b. 215/M35**
    - **3. M89/M213** – Macro-Haplogroup F, subsumes G-R
      - **a. M201** – Haplogroup G
      - **b. M52/M69** – Haplogroup H
      - **c. P19/M170** – Haplogroup I
      - **d. 12f2a** – Haplogroup J
      - **e. M9** – Sub-macro-haplogroup K, subsumes L-R
        - **i. M20/M22** – Haplogroup L
        - **ii. M52/ M69** – Haplogroup M
        - **iii. LLY22a** – Haplogroup N

iv. **M175/M214** – Haplogroup O
v. **P27 M4** – Haplogroup P
   a. **P3** – Haplogroup Q
   b. **M207** – Haplogroup R

References

Al-Zahery, N., Semino, O., Benuzzi, G., Magri, C., Passarino, G., Torroni, A. & Santachiara-Benerecetti, A.S. 2003. "Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations." *Molecular Phylogenetics and Evolution* 28: 3. 458–72.

Arredi, B., Poloni, E.S., Paracchini, S., Zerjal, T., Fathallah, D.M., Makrelouf, M., Pascali, V.L., Novelletto, A. & Tyler-Smith, C. 2004. "A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa." *American Journal of Human Genetics* 75: 2. 338–45.

Barbujani, G. & Pilastro, A. 1993. "Genetic evidence on origin and dispersal of human populations speaking languages of the Nostratic macrofamily." *Proceedings of the National Academy of Sciences (USA)* 90. 4670–4673.

Coia, V., Destro-Bisol, G., Vergnelli, F., Battaggia, C., Boschi, K., Cruciani, F., Spedini, G., Comas, D. & Calafell, F. 2005. "Brief communication: mtDNA variation in North Cameroon: lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa." *American Journal of Physical Anthropology* 128: 3. 678–81.

Cruciani, F., LaFratta, R., Santomalazza, P., Sellitto, D., Pascone, R., Moral, P., Watson, E., Guida, V., Colomb, B., Zaharova, B., Lavinha, J., Vona, G., Aman, R., Coli, F., Akar, N., Richards, M., Torroni, Novelletto, A. & Scozzari, R. 2004. "Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa." *American Journal of Human Genetics* 74: 5. 1014–22.

Cruciani, F., Santolamazzo, P., Shen, P., Macaulay, V., Moral, P., Olchers, A., Modiano, Holmes, S., Destro-Bisol, G., Coia, V., Wallace, C., Oefner, J., Torroni, A., Cavalli-Sforza, L., Scozzai, R. & Underhill, P. 2002. "A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes." *American Journal of Human Genetics* 70: 5. 1197–214.

Diamond, J. & Bellwood, P. 2003. "Farmers and their languages: the first expansions." *Science* 300. 597–603.

Ehret, C. 1995. *Reconstructing Proto-Asiatic (Proto-Afrasan): Vowels, Tone, Consonants, and Vocabulary.* Berkeley, Calif: University of California Press.

Ehret, C., Keita, S.O. & Newman, P. 2004. "The origins of Afroasiatic." *Science* 306. 1680.

Hudson, G. 1977. "Language classification and the Semitic prehistory of Ethiopia." *Folia Orientalia* 18. 120–166.

Luis, J.R., Rowold, J., Caeiro, B., Cinnioğlu, C., Roseman, C., Underhill, P., Cavalli-Sforza, L. & Herrera, R. 2004. "The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations." *American Journal of Human Genetics* 74: 3. 532–44. [See Erratum in *Am J Hum Genet*. 74: 4. 788.]

Murtonen, A. 1967. *Early Semitic*. Leiden: E.J. Brill.

Myles, S., Bouzekri, N., Haverfield, E., Cherkaoui, M., Dugoujon, M. & Ward, R. 2005. "Genetic evidence in support of a shared Eurasian-North African dairying origin." *Human Genetics* 117: 1. 34–42.

Nettle, D. & Harriss, L. 2003. "Genetic and Linguistic Affinities between Human Populations in Eurasia and West Africa." *Human Biology* 75: 3. 331–44.

Quintana-Murci, L. & Fellons, M. 2001. "The human Y chromosome: the biological role of a 'functional wasteland'." *Journal of Biomedicine and Biotechnology* 1: 1. 18–24.

Sanchez, J.J., Hallenberg, C., Børsting, C. & Morling, N. 2005. "High frequencies of Y chromosome lineàges characterized by E3b1, DYS19-11, DYS392-12 in Somali males." *European Journal of Human Genetics* 13.856–866.

Scozzari, R, Cruciani, F., Santomalazza, P., Malaspina, P., Torroni, A., Sellitto, D., Arredi, B., Destro-Bisol, G., DeStefano, G., Rickards, O., Martinez-Labarga, C., Modiano, D., Brondi, G., Moral, P., Olckers, A., Wallace, C. & Novelletto, A. 1999. "Combined use of bialldic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations." *American Journal of Human Genetics* 65. 829–846.

Semino, O., Santachiara-Benerecetti, A.S., Falaschi, F., Cavalli-Sforza, L. & Underhill, P. 2002. "Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny." *American Journal of Human Genetics* 70. 265–268.

Stumpf, M.P. & Goldstein, D.B. 2001. "Genealogical and evolutionary inference with the human Y chromosome." *Science* 291: 5509. 1738–1742.

Y Chromosome Consortium. 2002. "A nomenclature system for the tree of human Y-chromosomall binary haplogroups." *Genome Research* 12. 339–342.

A dental anthropological hypothesis relating to the ethnogenesis, origin, and antiquity of the Afro-Asiatic language family

Peopling of the Eurafrican-South Asian triangle IV

Christy G. Turner II
Arizona State University

Diachronic comparisons of dental morphology in North and East African groups shows a marked difference through time in North Africa after the end of the Pleistocene. Further comparisons with archaeologically-derived Near Eastern dental samples suggest that the temporal changes in the North African teeth were due to population replacement or admixture from the north. On this basis, it is further suggested that Afro-Asiatic was introduced into Africa along with the migrating immigrant farmers and herdsmen from the Near East 10,000 to 7,000 years ago.

Keywords: dental morphology; migration; population replacement; Afro-Asiatic; North Africa; Near East; ethnogenesis

Ethnogenesis is a term commonly used in Russian anthropology. In Soviet times it was infused with the biological notion of Huxleyian "evolutionary progress" (Davis 1949: 65), but focused on the Communist belief that cultures will progress economically to their ultimate culmination in communism. Today, the term ethnogenesis has no ideological or predeterministic implication, and simply refers to the changes and conditions seen over time that lead to the formation of ethnic groups, whether they be defined by archaeological, ethnographic, biological, or linguistic criteria. The porosity or genetic leakage of human ethnic groups varies greatly depending on the strength and duration of various isolating mechanisms including geographic barriers, distance, language, culture, traditions, and racial attitudes to mention only a few. Still, "anthropologists, like other scientists, would like to create a single, parsimonious theory of origins that would account, in this case, for world distributions of language, culture,

and physical type" (Moore 1994: 11). However, at the moment, historical particularism holds sway.

As in biological evolution, isolation (a condition), adaptation (a process), and chance (another process) are important factors in ethnogenesis, as the term is loosely used today. For example, island populations may be distinctive due to the make-up of their founders (chance), remain unchanged over time (isolation), or develop traits that function well on a given island (adaptation). On the other hand, migration and gene flow can upset this *in situ* equilibrium, even replacing or mixing markedly with the original inhabitants, which this paper suggests happened in post-Pleistocene north Africa. Hence, I herein present a hypothesis about the ethnogenesis of the Afro-Asiatic language family (Greenberg 1966) based on dental morphological variation as identified over time in North Africa and the Near East. Simply put, the hypothesis proposes that Afro-Asiatic arose in the Levant and was carried by southward moving migrant farmers and herdsmen into North Africa mainly between the time that the Pleistocene ended, about 10,000 years ago, and the middle Neolithic, roughly 7,000 years ago. Southward migration and gene flow continued well into the middle of the second millennium B.C. as can be inferred from Table 1.

The proposed timing for introduction is consistent with the fact that there are relatively few languages in the family despite there being as of 1976, more than 175 million Afro-Asiatic speakers, at least half of this number speaking Arabic (Ruhlen 1976: 60). The hypothesis of a non-African origin of Afro-Asiatic finds support in the fact that the Semitic branch or sub-family is the most important in terms of number of speakers. Semitic languages include Arabic, Hebrew, and Amharic. Additional support is found in Afro-Asiatic's general lack of tonality that characterizes the majority of African languages, among other differences (Ruhlen 1976: 61).

Scholars today who are interested in the classification as well as formation of human groups generally envision evolution as primarily divergent, branching and hierarchical, and less often as converging, mixing, or as Neel, Layrisse, and Salzano (1977) express these two general formation concepts – fission and fusion. While fusion is patently happening today as witnessed by the recent and on-going large-scale movements of people from Africa to Europe, and from Latin America to the United States, its occurrence in the archaeological past is much more difficult to demonstrate due to problems in distinguishing diffusion from migration (Rouse 1986). Even in various island contexts where in theory, migration should be the major cause of saltational biocultural change, developments and change can occur through trading visits where no actual migration takes place. Moreover, similar environments give rise to similar adaptive problems, for which the types of solution may be small at any given moment in time. Given that there is sufficient biological difference between a migrant group and the receiving population, then the best direct evidence for archaeological migration is, of course, the remains of the migrants themselves. The classic case for

Table 1. A sample of African and near eastern dental trait frequencies. (Individual count, sexes pooled, ASU-DAS procedures; after Lipschultz and Turner, n.d.)

Trait	WA	Pleis	Mero	X-G	Chris	S.Lev	E-G	Natuf
Shovel	% 52.6	59.1	42.4	33.3	0.0	9.1	0.0	10.2
UI1	n 10/19	11/22	14/33	2/6	0/7	1/11	0/15	6/59
Mesial Ridge	% 32.4	22.6	32.6	10.0	28.6	12.5	15.0	13.8
UC	n 12/37	4/18	15/46	1/10	2/7	1/8	3/20	8/58
Cusp 5	% 32.4	28.6	11.1	22.2	16.7	4.3	4.8	3.2
UM1	n 11/24	4/14	7/63	4/18	2/12	1/23	4/84	6/189
Enamel ext.	% 37.5	60.0	14.0	11.5	8.3	0.0	15.8	8.8
UM1	n 24/64	21/35	12/86	3/26	1/12	0/21	13/82	6/156
Cusp 6	% 35.3	30.0	6.9	11.8	12.5	0.0	12.7	9.3
LM1	n 12/34	9/30	5/72	2/17	1/8	0/31	7/55	16/172
> 4 cusps	% 83.3	93.9	33.3	14.3	1.0	29.7	17.4	18.2
LM2	n 35/42	31/33	25/75	3/21	1/10	11/37	8/46	32/176

U denotes upper jaw, L, lower jaw; I1, central incisor; C, canine; M1, first molar; M2, second molar. WA denotes West Africa; Pleis, late Pleistocene Nubia; Mero, Meroitic Nubia; X-G, X-Group Nubia; Chris, Christian Nubia, S.Lev, Southern Levant Neolithic; E-G, En Gedi; Natuf, Natufian.

Dating and source: West Africa, 19th and early 20th century, collections of the American Museum of Natural History, New York. Christian Nubia, AD 550–1350, collections of Arizona State University (ASU), Tempe. X-Group Nubia, AD 350–550, collections of ASU. Meroitic Nubia, 100 BC-AD350, collections of ASU. En Gedi, Israel, 200 BC-AD 640, collections of Tel Aviv University (TAU). Southern Levant Neolithic, 10,000–6,000 BP, collections of TAU. Natufian, 12,200–10,046 BP, collections of TAU, and Peabody Museum, Harvard University, Cambridge. Late Pleistocene Nubia, 18,000–12,000 BP, collections of Southern Methodist University, Dallas, Texas.

prehistoric migration is, of course, the peopling of the New World from northeastern Siberia, an inference that has been time and time again verified by independent researchers (Turner 2002, 2005).

Studies of archaeologically-derived late Pleistocene and Holocene human teeth from the Nile Valley of north Africa when compared with the dental morphology of West Africans and other people south of the Sahara, along with comparative Europoid dental samples from England, Spain, Jordan, and Israel (Irish & Turner 1990; Turner & Markowitz 1990; Moyer & Turner 1990; Irish 1993, 2006; Lipschultz 1996; Ullinger, Sharidan, Hawkey, Turner, & Cook 2005; Bailey, Turner and Souich, n.d. Lipschultz and Turner n.d.) strongly suggest that the genetic and cultural isolation from Eurasia believed by G.J. Armelagos and associates for North Africa did not actually exist. This isolationist view has been promulgated for many years by these authors, eventually morphing into a paradigmatic reaction against the use of migrationism to explain cultural change in North Africa (Greene. Ewing. & Armelagos

1967; Van Gerven, Armelagos, & Rohr 1976; Adams, Van Gerven & Levy 1978; Calcagno 1986; Armelagos, Calcagno, Coppa, & Vargin 2002). This view was also part of the new evolutionary biology introduced into anthropology from genetics in the 1940s in which local evolution and adaptation replaced migration as the major concept for explaining biocultural changes that occurred within a given region. In the case of the Nile Valley, Armelagos and associates have gone so far as to label anyone as being a racist who proposes a Eurasian influence on the cultural evolution that occurred in the Nile Valley. (Turner & Markowitz 1990). Even less rational, Armelagos, Calcago, Coppa, and Vargin (2002: 38) challenge our use of the widely used Mean Measure of Divergence statistic and the well established genetic independence of most dental traits. Both the applicability of the MMD statistic and the independent inheritance of dental morphological traits have been so widely accepted and repeatedly demonstrated (Hillson 1986; Scott & Turner 1997) that these authors are obviously trying in desperation to deflect criticism of their flawed assumptions about isolation and local evolution. Clearly, they have thrown the baby out with the bath water and seem to believe that political correctness trumps empirical evidence. Globally, the use of MMD and dental traits has demonstrated strong correspondences concerning biological affinity as determined by multiple genetic polymorphisms (Cavalli-Sforza, Menozzi & Piazza 1994).

A series of papers that examined dental morphology in the northern Nile Valley and the adjacent Levant in late Pleistocene and Holocene times found that a great deal of biological change had occurred, so much so, in fact, that adaptation (i.e., natural selection) was clearly out of the question (Turner & Markowitz 1990; Irish & Turner 1990; Lipschultz & Turner n.d., Irish 2006). Elsewhere in the world large amounts of rapid diachronic dental morphological change have been found only in populations that had been replaced or had received large numbers of migrants over short periods of time (Turner 1976). Regional dental morphology is remarkably stabile over relatively long periods of time, for example the frequencies of many dental traits of North and South American Indians are effectively identical despite separation for at least 10,000 years (Scott & Turner 1997).

The above diachronic comparative dental studies of North Africa showed that there was a significant change in dental trait frequencies, i.e., genetic discontinuity, between the Pleistocene and Holocene populations, whereas continuity has been the situation for most of the later Holocene (Irish 2006). Late Pleistocene Mesolithic Nubians were dentally much like West Africans and other Africans south of the Sahara. This is illustrated in Tables 1 and 2. The Holocene Nubians (Christian, X-Group, and Meroitic) are much like southwestern Eurasians (En Gedi, Israel; Southern Levant; and Natufian, Israel). The latest in these North African-Near East dental studies involved archaeologically-derived materials from Israel, specifically Natufians (epi-Paleolithic) and more recent materials (En Gedi). Analyses showed that the Natufian dentition

was much like more recent Southern Levant people as well as Holocene Nubians. This work (Lipschultz 1996; Lipschultz & Turner n.d.) concluded that the only way to explain these similarities and differences was to propose significant gene flow and/or actual migration from the Southern Levant into the Nile Valley at the end of the Pleistocene and continuing into the Holocene. The people involved may have included Natufians along with their Levantine trading partners, the Mushabians. This inference for migration, based on dental morphology utilizes univariate and multivariate statistical comparisons, is strong, and is as well founded as the broadly accepted inference based on dental morphology that the ancestors of all Native Americans originated in northeastern Asia. Table 1 lists the frequencies of a few morphological traits to show the marked differences between Pleistocene and Holocene north Africa, and the Holocene similarities of north Africa and the Middle East. Table 2 provides the multivariate statistical comparisons using C.A.B. Smith's Mean Measure of Divergence (for discussion of this statistical method, see Harris & Sjøvold 2004). For readers unfamiliar with MMD, Smaller values indicate greater multivariate inter-group similarity, while larger values indicate greater dissimilarity. Hence, in Table 2, dental traits studied in the Southern Levant are multivariately four times more similar with Natufian teeth than are Southern Levant teeth like those of West Africa or Late Pleistocene Nubia.

Table 2. Ranked pairs of Mean Measures of Divergence (MMD)[1]

West Africa		**Christian Nubia**		**X-Group Nubia**		**Meroitic Nubia**	
L. Pleis.	*.040*	X-Group	*.000*	Christian	*.000*	X-Group	*.000*
X-Group	.236	Meroitic	*.003*	Meroitic	*.000*	Christian	*.003*
Meroitic	.237	En Gedi	*.065*	En Gedi	.110	En Gedi	.138
Christian	.327	Natufian	.221	Natufian	.228	Natufian	.233
En Gedi	.332	S. Levant	.277	W. Africa	.236	W. Africa	.237
Natufian	.437	W. Africa	.327	S. Levant	.301	L. Pleis.	.269
S. Levant	.603	L. Pleis.	.396	L. Pleis.	.302	S. Levant	.299
En Gedi, Israel		**S. Levant**		**Natufian, Israel**		**Late Pleistocene Nubia**	
Natufian	.063	Natufian	.137	En Gedi	.063	W. Africa	*.040*
Christian	*.065*	En Gedi	.157	S. Levant	.137	Meroitic	.269
X-Group	.110	Christian	.277	Christian	.221	X-Group	.302
Meroitic	.138	Meroitic	.299	X-Group	.228	Christian	.396
S. Levant	.157	X-Group	.301	Meroitic	.233	En Gedi	.460
W. Africa	.332	W. Africa	.603	W. Africa	.437	Natufian	.508
L. Pleis.	.460	L. Pleis.	.635	L. Pleis.	.508	S. Levant	.635

[1]From Lipschultz and Turner, n.d. Italicized values are *not* statistically significant, i.e., groups are very similar or one has a small sample size.

This paper makes an additional inference. Since there is archaeological and physical anthropological reason to believe that the Natufians were related to modern Semitic-speaking peoples of the Levant, I suggest that some part, if not all of, the Afro-Asiatic family originated north of Africa proper. Since Natufian culture dates to ca. 10,000 to 12,000 years ago it is suggested that the age of the Afro-Asiatic language family might also be about that old. This estimate of Afro-Asiatic antiquity going back into the late Pleistocene is not an absurdity. Using, again, an American Indian comparison, Greenberg's Amerind language family includes most of the native peoples of the Americas, excluding Aleut-Eskimo and Na-Dene-speakers. The best archaeological estimate for the arrival in the New World of ancestral Amerind-speakers places their entry at about 13,000 years ago, that is, Amerind can be inferred to have its temporal origin as long ago as is suggested for Afro-Asiatic. Of course, other language families and phyla could be even older, Australian for example, and several others are clearly younger.

In conclusion, dental morphology has proven to be a helpful indicator of ethnogenesis and inter-group affinity in world population history (Scott & Turner 1997). It is used herein, as it has been used elsewhere in the study of North African population history, to indicate that there was genetic discontinuity between late Pleistocene and Holocene North Africa. A strong case for population replacement or marked admixture has been developed. Using this population replacement finding, it is further suggested that the external migrants, who came from the north, were ancestral Afro-Asiatic speakers.

References

Adams, W.Y., Van Gerven, D.P. & Levy, R.S. 1978. "The retreat from migrationism." *Annual Review of Anthropology* 7. 483–532.

Armelagos, G.J., Calcagno, J., Coppa, A. & Vargin, R. 2002. "Changes in Nubian craniofacial morphology and dentition: evaluating the case for population discontinuity." *American Journal of Physical Anthropology Supplement* 34: 38.

Bailey, S.E., Turner II, C.G. & Ph du Souich. n.d. "Dental evidence for population affinities of the Iberian Peninsula (100 BC–1300 AD) and Balearic Islands: Peopling of the Eurafrican-South Asian triangle III." Under revision.

Calcagno, James M. 1986. "Dental reduction in post-Pleistocene Nubia." *American Journal of Physical Anthropology* 70. 349–363.

Cavalli-Sforza, L. Luca, Paolo Menozzi, & Alberto Piazza. 1994. *The History and Geography of Human Genes*. Princeton, N.J.: Princeton University Press.

Davis, D. Dwight. 1949. "Comparative anatomy and the evolution of vertebrates. *Genetics, Paleontology, and Evolution,* Ed. by G.L. Jepson, E. Mayr & G.G. Simpson, 64–89. Princeton, N.J.: Princeton University Press.

Greenberg, Joseph H. 1966. *The Languages of Africa*. Bloomington: Indiana University Press.

Greene, D.L., Ewing, G.H. & Armelagos, G.J. 1967. "Dentition of a Mesolithic population from Wadi Halfa, Sudan." *American Journal of Physical Anthropology* 27. 41–56.

Harris, Edward F. & Sjøvold, T. 2004. "Calculation of Smith's Mean Measure of Divergence for intergroup comparisons using nonmetric data. *Dental Anthropology* 17. 81–93.

Hillson, Simon. 1986. *Teeth.* Cambridge, Engl.: Cambridge University Press.

Irish, Joel D. 1993. *Biological Affinities of Late Pleistocene Through Modern African Aboriginal Populations*. Ph.D. dissertation, Anthropology, Arizona State University, Tempe.

Irish, Joel D. 2006. "Who were the ancient Egyptians? Dental affinities among Neolithic through Postdynastic peoples." *American Journal of Physical Anthropology* 129. 529–543.

Irish, Joel D. & Christy G. Turner II. 1990. "West African dental affinity of late Pleistocene Nubians. Peopling of the Eurafrican-South Asian triangle II. *Homo* 41: 1. 42–53.

Lipschultz, Joshua G. 1996. *Who Were the Natufians? A Dental Assessment of the Population Affinities.* M.A. thesis, Anthropology, Arizona State University, Tempe.

Lipschultz, Joshua G. & Christy G. Turner II. n.d. "The Southern Levant corridor to Nubia. Peopling of the Eurafrican-South Asian triangle III." Under revision.

Moore, John H. 1994. "Ethnogenetic theory." *Research and Exploration* 10: 1. 10–37.

Moyer, Kathy. & Christy G. Turner II. n.d. "Dental health and variation in Jordan." Paper presented at the 1990 annual meeting of the American Schools of Oriental Research, November 17–20, New Orleans, Louisiana.

Neel, J.V.M. Layrisse & Salzano, F.M. 1977. "Man in the tropics: the Yanomama Indians. *Population Structure and Human Variation. Volume 2. International Biological Programme*, Ed. by G.A. Harrison, 109–142. Cambridge, Engl.: Cambridge University Press.

Rouse, Irving. 1986. *Migrations in Prehistory. Inferring Population Movement from Cultural Remains*. New Haven: Yale University Press.

Ruhlen, Merritt. 1976. *A Guide to the Languages of the World*. Stanford, Calif.: Language Universals Project, Stanford University.

Scott, G. Richard & Christy G. Turner II. 1997. *The Anthropology of Modern Human Teeth. Dental Morphology and Its Variation in Recent Human Populations*. Cambridge, Engl.: Cambridge University Press.

Turner, Christy G. II. 1976. "Dental evidence on the origins of the Ainu and Japanese. *Science* 193. 911–913.

Turner, Christy G. II. 2002. "Teeth, needles, dogs, and Siberia: bioarchaeological evidence for the colonization of the New World. *The First Americans. The Pleistocene Colonization of the New World*, Ed. by Nina G. Joblonski, 123–158. San Francisco, Calif.: Wattis Symposium Series in Anthropology, Memoirs of the California Academy of Sciences, No. 27.

Turner, Christy G. II. 2005. "A synoptic history of physical anthropological studies on the peopling of Alaska and the Americas." *Alaska Journal of Anthropology* 3: 1. 157–179.

Turner, Christy G. II, & Michael A. Markowitz. 1990. "Dental Discontinuity Between Late Pleistocene and Recent Nubians. Peopling of the Eurafrican-South Asian Triangle I." *Homo* 41: 1. 32–41.

Ullinger, Jaime M., Susan Guise Sheridan, Diane E. Hawkey, & Christy G. Turner II. 2005. "Bioarchaeological analysis of cultural transition in the southern Levant using dental nonmetric traits." *American Journal of Physical Anthropology* 128. 466–476.

Van Gerven, D.P., Armelagos, G.J. & Rohr, A. 1977. "Continuity and change in cranial morphology of three Nubian archaeological populations." *Man* 12. 270–277.

African weeks

Daniel F. McCall
Boston University

In honor of Harold Fleming – founder of ASLIP

Hal Fleming and I were colleagues and friends in the Anthropology Department and the African Studies Center at Boston University for years and remained close after retirement. We had even co-taught a course one year trying to establish a diachronic anthropology. The four fields of American Anthropology are the basis for whatever diachronic dimension is attainable in the study of non-literate peoples. Hal is better steeped in Linguistics and Genetics than I, so I will not try to match him there, but this story of a culture-complex that runs through many millennia is one I know he enjoys. He had heard it before, but it has never been previously published.

The planetary week as a culture-complex is traced from its origins in Babylonia, through the cultures of the Mediterranean Basin, and eventually to its development in the Akan cultures of the Guinea Coast of West Africa.

Nineteenth century anthropologists deduced historical connections from geographical distributions of culture items. After a number of these supposed connections had been effectively contradicted, this method of deduction became suspect. A.L. Kroeber offered a caution: a single culture item is inadequate for the purpose of probing for historical linkages, but a culture-complex (i.e., an integrated set of a number of culture items) may be more promising. He laid out the case for a few culture-complexes. One of them is the "planetary week" (Kroeber 1948; Colson 1926).

A "planetary week" is a set of seven days allegedly under the influence of the seven visible celestial bodies in the sky that, in contrast to "stationary" stars, move their positions in a regular and predictable manner. All of these are "planets" – in the original Greek word, which means "wanderer".

A week is the only period of time in a calendar that is not derived from a sidereal periodicity: years are derived from the rotation of the earth around the sun; a month is suggested by the phases of the moon, and a day is the result of the earth turning on its

axis. The week, in contrast, is a socially selected period to aid in planning activities of less than a month's length. Weeks in different cultures have been of varying numbers of days.

The planetary week is a culture-complex because it comprises several distinct ideas:

1. Each day of the week is attached to a particular supernatural power, a god or goddess.
2. Each of these deities is the epiphany of a planet, giving the week seven days because there are seven visible planets: Sun, Moon, Mars, Mercury, Jupiter, Venus, and Saturn.
3. Each of the planetary deities is believed to control events on a particular day of the week.
4. A person born on a particular day has characteristics given to him or her by the deity who "owns" the day of the person's birth.

Kroeber presented an account of the origin and diffusion of the planetary week. It was developed in Babylonia by priests who identified the planets with seven of their important deities. Astronomy and astrology, both of which were in their infancy, were not yet distinct. Between the eighth and sixth century BCE concepts emerged in which each of these deities presided seriatim over a day and each of the deities had influence over events during his or her day.

During Greek over-rule of Babylonia after Alexander's conquest, the new rulers adopted the Babylonian astrology while substituting Greek gods for Babylonia gods and – as Kroeber details – caused a change that shifted the order of the days in the seven day week. Also the Greeks used the associated astrology for ordinary persons, anyone or everyone whereas the Babylonians had reserved astrological predictions exclusively for the king and the interests of the kingdom. This enlargement greatly increased the likelihood of diffusion, which is exactly what soon happened. Thus in its first movement away from Mesopotamia, the planetary week had been modified in two ways by the Hellenistic Greeks.

It is this Greek form of the planetary week that spread to the Mediterranean basin. It became prevalent throughout the Hellenistic world. It was not only a specific type of week that was diffused but also a system of astrology integrated with the seven days; professional astrologers, known as Chaldeans, spread out westward from Mesopotamia. There were other methods of augury, examining the entrails of a sacrificial animal, interpreting the flight of birds, etc., but astrology quickly became popular.

Later the expanding influence of Rome helped spread the peculiar form of week further west and north in Europe and also in western North Africa. Each diffusion saw a substitution of local gods for the gods of the people from whom they borrowed the planetary week. In the shift from Babylonian to Greek to Roman to German and

also to Punic and Berber, the reason for the equation of the deities is usually clear, but occasionally one may be difficult for modern scholars to understand; for example, Germans replaced Roman Mercury, a minor god, with Teutonic Woden, king of the gods! And they let Thor have Jupiter's day. Each diffusion into a new cultural home for the planetary week meant some variation of the quality attributed to particular days; these variations generally derived from the local gods with whom a day was identified and the characteristics of that deity.

Astrology was the motive force that propelled the planetary week into new regions. As a week, seven days was not necessarily better than other existing weeks. Greece already had a ten day week, and Rome had an eight day week; these remained the civil (or official) weeks, and the civil week ran simultaneously with the seven day astrological week after the latter had been introduced. Curiosity, or apprehension, of what was in store for one in the future drew seekers of foreknowledge to astrologers who ran a profitable business.

The planetary week was used only for astrology until the fourth century CE when Constantine, the first Christian emperor of Rome, in 321 CE, made the seven day week the official week. Constantine was not promoting astrology; on the contrary – the old Roman religion was more competitive with Christianity than was astrology, and the Christian emperor wished to disconcert the Senate, many of whose members preferred the old religion and other old ways, including the 8-day week. Reducing the civil week by one day upset habits of the pagan people, while the Church, which Constantine established, attempted to discourage astrology (with minimal success). Constantine also made Sunday the first day of the week instead of Saturday, which had been the first day in both the Babylonian and Greek versions of the planetary week. Some historians think he chose Sunday in order to palliate the worshipers of Mithra, a deity popular among soldiers in the Roman legions. Mithra was a competitor with Christ at the time, and the Christian emperor needed support of the army.

Thus, in the Roman Empire two more changes were instituted: changing the day that began the week and attempting to separate the planetary week from astrology.

In Western Europe remnants of the Roman planetary names of the week remain. The Welsh who were in England, under Roman rule, still have names for days of the week that recognizably reflect the Latin names. When peoples around, but outside, the Roman Empire borrowed the planetary week, they nationalized their new week by giving their local gods ownership of the days. In English we have a Germanic form brought into England by the Angles and Saxons. Sun-day, Moon-day (shortened to Monday), for the two largest lights in the sky, but most of the other names of the days are disguised by Germanic gods substituting for Roman deities, except Saturday, which is Saturn's day. French and other Romance language day names preserve more of the Roman planetary names for days of the week, e.g., *mardi* (Mars' day), *jeudi* (Jove/Jupiter day), etc.

The day of one's birth was believed to pre-determine one's fate, and astrologers claimed to be able to calculate the options an individual had in any crisis. Astrologers caused turmoil in every society where they practiced and many efforts were made by Roman Emperors to suppress them. But yearnings for knowledge of the future made the attraction of astrology hard to eradicate. Religions, as wells as governments, tried, with varying results, to suppress the astrological aspects of the planetary week. Judaism numbers the days of the week, as does the Orthodox Christian Church; the Vatican likewise tried to promote numbered days but succeeded only in Portugal, then an appendage of the Vatican. Islam definitely rejects the polytheistic basis of the deities that were presumed to "own" the days of the week.

When Kroeber wrote in the late 1940's, there were some documents that historians could accept as evidence that the planetary week, in its various modified forms, had been borrowed over a chain of connections from Babylonia via Greece and Rome to all of Europe and to southern Asia (where its astrological aspects remained more prominent than currently in European cultures). It was, in fact, documentary history that enabled Kroeber to substantiate the reality of the diffusion.

There is one other example of a planetary week, which Kroeber did not know about, and for which no historical documents exist. If we can plausibly deduce that it also is ultimately derived from the same source, this should constitute a completion of Kroeber's effort to show that a culture-complex, in this instance a planetary week, can be traced in its paths of diffusion. This previously unremarked planetary week, which is in West Africa, is still recognized by the local Akan peoples (Asante Akwapim, Fante, etc.)

How did this culture complex of the planetary week come to be in the rain forest along the central Guinea Coast of West Africa? Without written reports of observers at the time, the answer has to be hypothetical. We may be justified in deducing that it was transported across the Sahara desert from northern Africa where Rome once ruled, and where, before the introduction of Islam into that region, the planetary week flourished.

Trade of some kind was carried on – to a limited extent – across the Sahara in Roman times, and the Kingdom of Ghana came into existence before Rome's presence in North Africa was finally eliminated. Traders bring more than material goods; ideas are also often transmitted. The Kingdom of Ghana in the Sahel, on the southern edge of the Sahara, is the logical recipient of trade, and cultural ideas, coming across the desert, but the planetary week is not found in the region of ancient Ghana in modern times. If Ghana borrowed the planetary week from traders from the north in Roman times, its disappearance can be attributed to the later arrival of Islam in this region, since Islam abhors the concept of multiple deities in any context and would not let any such "own" the days of the week.

There were other possible trading partners in the Sahel – Gao or Tekrur – but Ghana is considered the earliest in the region, and as we will see, is the most likely

to have been the temporary host of the planetary week because there are claims of connection from Ghana to the Akan, the present day observers of a planetary week, to their southeast.

The argument that the Akan peoples have some historical connection to ancient Ghana remains disputed, but a linkage – of some sort – could explain the observance of the planetary week among these people in the rain forest along the central Guinea Coast, rather distant from the locale of ancient Ghana. In fact, J.B. Danquah claimed the Akan week, this specific peculiarity, as part of his evidence that the Akan had a relationship with Ghana, and he is responsible for the adoption of that name for the colony of the Gold Coast when it became independent from the British Empire. Danquah's suggested origin of the Akan week is slightly different from the route I outline here (he eliminates Rome), but we agree in realizing its historical importance.

Danquah, now deceased, is remembered principally as a politician, but he is the only one I am aware of who recognized the affinity of the Akan week with the planetary week (Danquah, nd.).[1] His writings on ancient Ghana remain controversial among academic scholars. Disdain for Danquah's claim of an Akan connection with Old Ghana may rest largely on general ignorance, even among well educated persons, of the constituent elements of planetary week.

Even J.G. Christaller, who compiled the great dictionary of the Akan language (Christaller 1937), and provided an appendix (B III) on the day-names of males and females born on each of the seven days in the Akan week, did not recognize that the Akan week – through he described it – was a planetary week! Probably he was unaware that the week he used was also a planetary week because the astrological aspects were so eroded in recent centuries.

The rise of science marked the demise of astrology. As late as the 14th century astrologers were regularly advisors to kings in Europe. Even Elisabeth Tudor (d. 1603) sometimes listened to an astronomer, named Dee, but it is harder to tell if she followed advice; she tended to think independently of any of her advisors. The increase of scientific thinking finally repudiated – among educated persons at least – any credibility for the fable of astrology in Europe and America. Copernicus's destruction of the earth-centric concept of the cosmos, and Galileo – and his successors with telescopes – reduced the planets from their godly status *in astrology* to mere celestial bodies without any discernible divinity *in astronomy*. Most people today do not recognize that what we have is a planetary week because its astrological heritage has

1. Joseph Boakye Danquah, B.A., LL.B., Ph.D. I read a preliminary paper, "The Akan Planetary Week," to a meeting of the American Society for Ethnohistory in the 1960's, which was never published. At the time I had not yet come upon this collection of Danquah's works.

been obscured, and many of the planetary names for days have been distanced from the names for the planets.

My discovery of the Akan planetary week came about from being tutored in the Akan language. Michael Osai, an Asante of New Juaben, head master of a school in Koforidua in the then British colony of the Gold Coast, came shortly after dawn in 1951 to tutor me every morning before opening his school.

"When were you born?" he asked me.

"March 3rd 1918." That was the appropriate answer, but not what he wanted. Osai had to elaborate a bit before he got the information he needed: the day of the week! The day of the month, the month itself and even the year was irrelevant to him.

"Ah! You were born on Sunday! That is Kwasida in our calendar, so your day-name is Kwesi."

If I had been born a day earlier, I would have been Kwame, and believed to have commanding strength, or one more day earlier, I would have been Kofi, a likable guy, but if on Wednesday, I would have been Kwaku, a tricky, unreliable person. I was glad to be Sunday-born with no heavy burden to bear because of my natal day. But in one way it was disappointing to be a Kwesi, because all whites (except Syrians or other Muslims) were called Kwesi because they went to church on Sunday, so are assumed to have a special connection to Kwesida. "Kwesi Obroni" (Sunday-born white man) was the usual designation for any white man. Being merged into the whole white population robbed me of some sense of individuality I would have preferred to have. From then on, for the nearly two years I remained in Koforidua, I was "Kwesi Daniel."

The Akan have several ways of acquiring names. Order of birth: first born is "Adu"; third born (considered luck) is "Mensa." Ninth-born, "Nkruma," (3×3) is regarded as powerful. One can also be given an honorific name reflecting some accomplishment. But the day-name is the basic name that is generally used.

A person's name is more than a mere identification to distinguish him or her from other persons; what name you are known by carries a suggestion of what kind of a person you are. Kofi Mensa is as common among the Akan as John Smith is among Americans. This comes about because if one is born on Friday, it is likely that will be stressed as what the "likable guy" would prefer to be known by, and if he is also his mother's third child, and therefore "lucky" that may be advertised as well. In contrast, a Wednesday-born man will prefer to use some other name than "Kwaku" with its unflattering connotations.

It is possible to manipulate these meaningful names. The political leader known as Kwame Nkruma(h)[2] was his mother's only child; thus not an Nkruma. His father

2. The name is found both with and without the final *h:* Nkrumah or Nkruma. [Ed.]

had been an authentic Nkruma with eight older siblings, and when his son was enrolled in a school, the European teacher recorded his father's day-name as the boy's surname. He later came to the United States to attend Lincoln University as Kofi Nkruma, but when he returned to his own country he had taken the name "Kwame" Nkrumah as more advantageous to a political career. This I learned from him in a conversation in 1952.

Curiously, an English folk rhyme gives characteristics to children born on different days of the week. These are, it would seem, remnants of a comparable set of concepts, but poorly remembered or modified to fit the rhyming scheme:

> Monday's child is fair of face
> Tuesday's child is full of grace
> Wednesday's child is full of woe
> Thursday's child has far to go
> Friday's child is loving and giving
> Saturday's child works hard for a living
> But the child who is born on the Sabbath day
> Is bonny and blithe and good and gay.
>
> *Anon.*

The concept, in this English verse, that children born on a certain day of the week will vary in their personal characteristics from those born on other days is certainly similar to the Akan idea of one's personality being determined by one's birth day. We may note that Friday's child is "loving and giving" as Venus's child should be, and as the Akan agree is "charming and lovable." This similarity is what we would expect if both were derived from the Roman planetary week.

The English Wednesday's child "is full of woe." This too is comparable to the Akan conception of Wednesday's child, Kwaku, who is one who plays tricks on others, and as the Akan trickster stories recount usually is caught and discomforted by his own tricks. The owner of Wednesday in the Roman system was Mercury, and he has a tricky personality: patron of markets, and also of thieves, but also the messenger of the gods – he could be a Kwaku. But Germanic Woden (=Roman Jupiter/Jove) would not normally be associated with woe! Though Woden could deal out woe if he wished, but would he give woe to his day's child? It seems the "mercurial" attributes of the Roman "Mercury's day" have been carried over into the Germanic adoption of the planetary week, despite this peculiar identification of a German deity dissimilar to the Roman one. Astrologers no doubt were responsible for the persistence of a Roman-day characteristic to a German day the deity of which is inconsistent with the attributed quality.

The English Saturday's child "works hard" because Roman Saturn was a god of farming, and farmers work hard as anyone who has worked on a farm knows. The

Roman influence is present, one can see, in that Saturn is the only Roman deity identified in the name of a day; all the others have Germanic substitutions.

Monday doesn't say much, but to be "fair of face" is perhaps appropriate for a full moon. Some of the day's offspring in the verse are elusive of any heritage of the old gods, and Sunday, being in Christianity "the Lord's day" gets two lines, instead of one as the others have (enabling eight lines to complete the four couplets) and what it says obviously has to be a non-planetary treatment of that day.

Tuesday's child has a fate that may come from having to rhyme with "face" inasmuch as "grace" is not what one would think of as the primary characteristic of the minor German god, Tiw (=Roman Mars). And the same may apply to Thursday's child's destiny, which is not easy to justify as coming from the planetary week deities. Thor, being "thunder", but standing in for a Roman god, Jove, because Woden somehow in the transition to Germany (then including the Angles and Saxons) replaced Mercury, thereby confusing Wednesday's character); it has to be a powerful force, but is left here undefined.

And we should consider that the terms "jovial" (convivial), "mercurial" (quick-witted, thievish, changeable), "saturnine," (melancholic), "martial" (war-like) as applied to persons is another relict of the belief that planets have influence over human beings. ("Venereal" (lustful) has been distorted by its designation for a disease.)[3]

European names for the days of the week, in several languages, are clearly related to the Roman names; and the number of days in a week (an arbitrary number) is the same, and there is no doubt among scholars who have examined the question that the relation of modern European to ancient Roman in this regard is solidly established. In regard to the Akan situation we can make out a relationship for the names of some days but not all, but here the number of days in a week may be of some significance. The Akan are surrounded by peoples who have four day weeks, eight, or six day weeks. Seven is not an expected number in this region.

---

3. The *Boston Globe*, December 4, 2006, carried this story: "Michael Ward, a British student of literature working on a dissertation on [C.S.] Lewis … realized [that Lewis] aimed to retrieve the medieval worldview, in which there were seven "planets," or heavenly bodies, each thought to be associated with certain human characteristics. Those born under the sign of Jupiter, also known as Jove, were "jovial" (convivial). Mars was a "martial" (warlike) planet, Mercury was "mercurial" (both good and evil), Venus "venereal" (lustful), and Saturn "saturnine" (melancholic). The sun and the moon, finally, were believed to represent actions and emotions, respectively, or lucidity and dreams." This was to explain why Lewis wrote seven chronicles in his series "as opposed to, say, six or eight?" Ward's research was quoted in reference to the film made of one of Lewis's novels. It elucidates the worldview once prevalent in Europe. It is the conceptual framework that must have inspired the English ditty about Monday's child, etc.

The Akan concepts of some of the days of a week are easily equated to Roman days and even to English days. A person born on Friday (Kofe if male, Efua if female) is "charming and lovable" because that is the characteristic of the supernatural being who owns Friday. A person born on Wednesday is "tricky" because that is the day that the planet Mercury is dominant. Not all the other Akan days are as transparently duplicates of North African "owners of the days," but three out of seven is better than chance. And all seven days display the trait of personality being derived from the birth day.

Christaller would not call the seven supernatural beings who owned the days "gods" because he did not find them as deities in the religion of the Akan peoples, and he guessed that the system of days and day-names had been borrowed. He did not suggest a source.

Friday in the Roman week belonged to Venus (Frigg was the Germanic goddess of fertility and love). The Akan concept of the owner of Friday is strikingly in accordance with Rome's Saturn, who owns Saturday, was in Roman North Africa merged with a Carthaginian god, and considered by the local peoples as more powerful than Jupiter. Note that Saturday in Europe is still usually attributable in several languages to Saturn, as it was in Roman North Africa. But in Europe, as in the English jingle, there is no influence of the Carthaginian concept of great power, but south of Carthage where the planetary week is found among the Akan, the Carthaginian influence in that aspect is palpable.

The primacy of Saturday in the Akan formulation of the week, as reflected in the power attributed to the character of Kwame, is a plausible "continuity" from a putative Roman North African source – recalling the equating of Saturn with Baal Hamoun, the head of the Carthaginian pantheon. The quality attributed by the Akan to Saturday indicates the area of Carthaginian influence, and not Egypt, as the probable donor of the Akan week.

Friday and Wednesday have the qualities for Akan that were found in the Babylonian, Greek and Roman varieties of the planetary week. Not all the other Akan days are as transparently duplicates of North African "owners of the days", but an anthropological axiom is that "adopt means adapt," so modification is to be expected here in this African location as is found in other locations where modification is documented, e.g., Germany (demotion of Woden, who should be the counterpart of Jove) and Roman North Africa (the elevation of Saturn).

In its travels from the Mediterranean Basin to the Akan in West Africa something was lost: the identification of the visible planets with the spirits, or "genii," as Christaller referred to them. In previous diffusions the characteristics of the planetary deities were modified, as noted above, but losing the planetary connection altogether might seem to negate the Akan week as being a planetary week. However, we started out considering a culture-complex, and the culture-complex of the planetary week

has four culture-items; the connection to the planets is listed (above) as number two. The retention of seven for the number of days in a week is, one might say, a ghostly reflection of their former place in the system. The other three culture-items in the culture-complex are present.

We should note that all the peoples in the regions surrounding the Akan peoples have a different type of week. Theirs is a common premodern type of week among farming peoples – the market week. The early Romans and Aztecs had market weeks. A market week is based on the number of days in the periodicity of an area's market. In an area where a market is needed every day, among the neighbors of the Akan, the buying and selling might take place in four different locales within a local area, each market offering easier access to a quarter, more or less, of the area's population. Most market weeks in western Africa seem to be four-day weeks, but an eight-day market week is not uncommon. The eight-day week may be simply a doubling of a four-day week. The name of the day is the name of the market that is open on that day. A six-day market week also existed but seems to have been less common.

Doing fieldwork north of the Akan in the 1970's, I never knew where the market was on any given day. It was not a problem for Edmond Zinaku, my interpreter and assistant. He merely had to know what day of the four-day week it was and he could tell me where we needed to go to buy some provisions.

Into a vast region of market weeks, the planetary week intruded but remains the exclusive possession of a single group, the Akan. As Christaller recognized, the peculiarities of the Akan week suggest that it was borrowed from some outside location.

Long distance trade in Roman times did not directly reach from North Africa all the way to the rain forest of the Guinea Coast. An intermediary connection is therefore needed. The Sudanese kingdom of Ghana could have been the linkage. We cannot say that the people of Ghana, about a millennium ago, had a planetary week; our information on them is limited to some comments by a few writers in Arabic (before the kingdom was overthrown by Islamic jihadists), and a modicum of archeological excavation, in neither of which is any indication of a planetary week. Given the nature of extant evidence this lack is not surprising. "A lack of evidence is not evidence of a lack."

The planetary week arrived among the Akan, we may deduce, by traveling along the routes of the gold trade. The earliest reference to Ghana – "the land of gold" – was by the Arab geographer al-Fazari in the late eighth century (Law 1978: 203). The Arab invasion of Western North Africa, which thereafter became known as the Maghreb, was in the mid-seventh century, approximately a century prior to al-Fazari's account. During much of that century the Arab invaders were engaged in overcoming the resistance of the remnants of Byzantine opposition and the Berber, Punic, and retired and settled Roman legionnaire (who were Latin-speaking) inhabitants. So it seems the new masters of North Africa turned their attention to

trans-Saharan exchange of salt and textiles for gold about as soon as pacification of the Maghreb permitted.

The absence of a Latin reference to Ghana causes some scholars to wonder if there was any contact between Romans and any people south of the Sahara. It is probable that no individual from Italy ever reached that far south, but that does not exclude traders from the Roman provinces of "Africa" or "Mauritania" carrying on commercial or other exchanges with Black Africa. Traders in the Carthaginian tradition would be secretive of their source of riches, and the Roman overlords may have been ignorant of this over-the-desert trade. The character of the Akan "owner" of Saturday, consistent with the North African perception of Saturn's day rather than the Roman "owner" of that day, suggests that Romans were not the conveyer of the planetary week, who must have been identical with the trans-Saharan traders. This Carthaginian element in the Akan planetary week argues for Punic traders and not Roman traders. Absence of a mention of "Ghana" in Roman writings could be due to the fact that the word "Ghana" as the name of the country was a mistaken Arabic creation (Levtzion 1973: 3). The word *Ghana* was one of the titles of the king; another was *Tunka*. The name of the country in the oral traditions, as Levtzion points out, is Wagadu. If the Romans knew of the kingdom they would have known it by some other name, perhaps as mistaken as the Arabic appellation which has persisted via translations of writings until it is our common modern designation for the kingdom.

Gold is not a common commodity. It is found in only a small number of localities. The area occupied by the Akan people is one, and unlike the others in western Africa south of the Sahara, it is in the rain forest (difficult to penetrate), feared by outsiders as the "jungle". The reason traders from the grassland north of the rain forest wanted to penetrate this forest was to obtain trade goods. There was kola, a nut desired for its qualities (e.g., to staunch thirst), and salt from salt-ponds on the coast, but especially there was the gold, which was of greater value than the other forest products. There is no doubt that gold was the most alluring (and profitable) attraction (Sutherland 1959: 8).[4] The people of Old Ghana did not have gold in their own territory but obtained it from people who lived in areas where it was mined or collected. The gold producing areas outside the forest were the earlier suppliers of gold for Ghana but eventually the Akan area was also developed as a producer of gold. Gold suggests a sufficient reason for the linkage of the Akan to Old Ghana. Historians and anthropologists will continue to

4. "One of the rarest of metals, it had gained a peculiar and indeed unique place in human experience. Taste and style may change … but there are few in any age who are not immediately affected by the colour and luster of gold." The profit from the Akan gold came only partially to the Akan producers; Ghana kept its share from the trade and the trans-Saharan traders probably acquired the largest portion.

debate the connection, but they should no longer ignore the planetary week among the Akan. How did it get to the forested coast of Guinea if Ghana was not the donor?

References

Christaller, J.G. 1937. *Dictionary of the Asante and Fante Languages.* 2nd Edition. Basel: Basel Evangelical Missionary Society.

Colson, F.H. 1926. *The Week.* Cambridge, UK: Cambridge University Press.

Danquah, Joseph Boakye. nd. *Historic Speeches and Writings on Ghana.* Compiled by N.K.A. Akyempong, Accra, Ghana.

Kroeber, A.L. 1948. *Anthropology*. New York: Harcourt, Brace and Co.

Law, R.C.C. 1978. "North Africa in the Hellenistic and Roman Periods, 323 BC to AD 305." *The Cambridge History of Africa,* Ed. by J.D. Fage & Roland Oliver, Vol. 2, 148–290. Cambridge, UK: Cambridge University Press.

Levtzion, Nehemia. 1973. *Ancient Ghana and Mali.* London: Methuen.

Sutherland, C.H.V. 1959. *Gold.* London: Thames & Hudson.

PART II

African languages – Synchronic studies

Gender distinction and affirmative copula clauses in Zargulla*

Azeb Amha
Leiden University

In Omotic, grammatical gender that is expressed through agreement between associated words and equally affecting animate and inanimate nouns is not common. Rather, the system of gender-marking in Omotic mostly involves direct-marking on the noun and it is semantically determined, i.e., the system classifies (large) animate nouns according to their biological gender and leaves inanimate nouns unmarked or marks all inanimate nouns categorically as either masculine or feminine. In Zargulla feminine and masculine gender is morphologically distinguished in the following areas: in verbal subject-agreement marking, definiteness marking in nouns and in third person singular pronouns. Moreover, Zargulla has two copula markers: *-tte* and *-tta*, the distribution of which partly corresponds to masculine and feminine gender respectively. The focus of the present contribution is on these two morphemes. We show that the gender-marking role of the two morphemes is diminishing and that they are grammaticalizing into two different discourse-functional morphemes.

1. Introduction

According to the classification of Fleming (1976), Zargulla is a member of the east Ometo branch of the Omotic language family. Other languages from this branch include Zayse, Koorete and Haro. In the present contribution, I discuss data from Zargulla that bear on one of the grammatical features Fleming mentioned as highlighting the difference between Cushitic and Omotic languages. This concerns morphemes marking (grammatical) gender. In his overview of Omotic languages, Fleming (1976a: 36) writes:

> The status of grammatical gender is different in Omotic languages as compared with Cushitic: they either lack grammatical gender altogether or have it to a much lesser extent, or use different indicators for it than is the case with most Cushitic languages. For example, many Cushitic languages classify nouns, pronouns, and demonstratives by means of /k/ for masculine and /t/ for feminine. This k/t

*I would like to thank Thilo Schadeberg for comments on an earlier version of this paper.

> distinction is apparently totally absent from the Omotic languages, although traces may survive in the o/e distinction in Kafa nouns. The *o* resulting from the influence of a lost *k*, *e* from a lost *t*.

In this paper we will show that the final vowels *e* and *a* in the copula markers *-tte* and *-tta* in Zargulla indicate masculine and feminine gender respectively. Of the two endings, *-e* formally corresponds to the masculine gender marker which Fleming mentioned above but instead marks feminine gender in Zargulla.[1]

The gender distinction in copula clauses is not consistently used in Zargulla. Speakers sometimes use the two affixes interchangeably. There are different uses of nonverbal predicates that are marked by *-tte* and *-tta*. Firstly, the alternation between the two copula markers indicates gender distinction; in this case free alternation of the two morphemes is not possible. Secondly, in dialogues involving questions and answers, an affirmative copula clause with *-tta* expresses that the respondent is "annoyed" with the question or "impatient" with the interlocutor whereas in the same communicative context copula clause with *-tte* provides a simple informative response to the question. In addition to this, the morpheme *-tte* has an extended use of marking contrastive focus in Zargulla. Because of these extra meanings we will not indicate a morpheme boundary between the geminate consonant *-tt* and the vowels *-e* and *-a* in the present contribution. We will also use the same transliteration (i.e., COP) throughout, thereby indicating the assumption here that the various uses developed from the same gender-sensitive copula morphemes. The goal of the paper is to show that even where traces of common cognate morphemes are available between Cushitic and Omotic languages, there are significant differences (innovations?) among the two languages that need to be accounted for.

The paper is organized as follows. In Section 2, we will describe copula constructions in Zargulla, separating the parts which do indicate gender distinction from those where no such distinction is made. In Section 3 we will show that *-tte* and *-tta* are also used to express functions which seem to have nothing to do with gender distinction. We also briefly discuss the use of the morpheme *-tte* (but not *-tta*) as a fully grammaticalised contrastive focus marker (Section 4).

2. Copula constructions in Zargulla

When a noun, adjective, pronoun or demonstrative is used predicatively, it is often affixed with the copula markers *-tte* or *-tta*. The two suffixes may be used interchangeably as in examples (1–2)

---

1. As Hayward (1984) showed, the *k*/*t* distinction did not disappear without traces either; according to him these are grammaticalised as copula markers in different languages (cf. also Hirut Woldemariam 2003). One also finds a *k* element in greeting forms of Zargulla with masculine and plural addressee: *wiashidee* "how is she?" vs. *waakidee* "how is he?"

(1) a. *haí tá maatsú-tta*
this 1SG.GEN thing-COP
'This is mine'

b. *haí tá maatsú-tte*
this 1SG.GEN thing-COP
'This is mine'

(2) a. *zargul-eď é gádé-y žillá-tte*
Zargulla-PL land-NOM green-COP
'The land of the Zargulla (people) is fertile'

b. *paranj-eď é ?úmmá-y boótsu-tta*
westerner-PL hair-NOM white-COP
'The hair of westerns is white'

In conversation, sometimes the copula affixes are simply dropped. Thus, as a reaction to the utterance in (3a), the speaker gives the response in (3b), where none of the copula suffixes is attached to the nominal *s'amáta maáts* "fresh milk". In (3a) the vowel *u* before the copula marker *-tta*, is an epenthetic vowel. Similarly, in (4) the copula marker is not attached to the word *šaato* as expected.

(3) a. *maáts hís-oo-s-i s'amáta maáts-u-tta*
milk:ACC say:CAUS-IMPS:REL-NMZ fresh milk-u-COP
'What you consider (good) milk is (only) fresh milk'

b. *hoó s'amáta maáts*
yes fresh milk
'Yes, it is fresh milk'

(4) *birá bíšš-att-i šaato ?angúss maák'k'-eé-s-í*
earlier wife-F:DEM-F:GEN child:ACC eldest become-IPF.REL-NMZ-NOM
'It is the first wife's child that is considered the eldest'
[i.e., this child is responsible to manage the family and inheritance,
even if he is younger than the son(s) of the second or third wife]

In some cases where the copula markers are used, the choice between *-tte* and *-tta* makes a semantic difference. The choice between the two depends on the gender of the subject of the copula clause. When the subject is masculine the copula suffix with the vocalic ending *-e* is attached to the non-verbal predicate; when the subject is feminine the copula with *-a* ending is used for the same purpose. However, this gender-distinction is not always consistently used since sometimes speakers use the two forms alternatively. Before illustrating gender marking in copula clauses, a brief note on gender marking in Zargulla is in order. In this language two (semantic) gender-distinctions are made. The distinction is based on the sex of the referent: a noun that refers to a female being is differently marked from those that refer to male beings. Inanimate nouns and some animate nouns for which such female-male distinction is

not obvious, are generally treated as feminine. Gender is marked on the noun it self through the selection of distinct definiteness markers, e.g., *?imats* "a guest", *?imats**ází*** "the guest (M:NOM)", *?imas'**att**oí* "the guest (F:NOM)". It is marked in third person singular verb forms through *-s-* and *-iš-* as in, *maák'ŏtt**es**ínne* "he returned", *maák'ŏtt**íš**ínne* "she returned" and in third person singular pronouns and demonstratives which have distinct masculine and feminine forms as well as in copula clauses, as we will show next.

In identificational clauses in which the nucleus of the predicate is a pronoun, and the subject is covert (eliptic), the gender distinction is consistently made. Examples of these are given in (5). Note that, in such predicative pronouns, the consonant of the copula morpheme is not geminated (i.e., *-t-* in example 5). Moreover, the last segment of the pronoun is dropped when the copula is added to it (compare the predicative pronouns in (5) with their corresponding subject and object pronouns given in brackets).

(5) *?ís-ta* "It is her!" (cf. *?ísí* "she (NOM)", *?ísó* 'her *?í* "her (possessive)")
?és-te "It is him!" (cf. *?ésí* "he", *?ésa* "him", *?é* "his")

The expression *hoó ?ísta* "yes, that is it!" (lit. "That is *her*") is frequently used as exclamation when an addressee volunteers information (e.g., a name of a person, place or some object) which the speaker is trying to remember.

When the subject of the copula clause is a nominalised demonstrative, gender distinction on the predicative element is judged important (6):

(6) a. *ha ?ísó-y s'eena-áde-tta ?aála-so-íš hakíme-so yeéde*
this 3FS-NOM health-father.ACC-COP why-LOC-F doctor-LOC come-PAST
'This one (F) is healthy. Why did she come to the clinic?'

b. *ha ?ís-í s'eenaáde-tte ?aála-so-s hakíme-so yeéde*
this 3MS-NOM health-father.ACC-COP why-LOC-M doctor-LOC come-PAST
'This one (M) is healthy. Why did he come to the clinic?'

In the examples in (6) *-tte* and *-tta* cannot be freely exchanged without changing the gender of the referent. The nucleus in these examples is a compound adjective and like all non-verbal predicates in Zargulla, it occurs in the Accusative case. On the other hand, subject-complements always occur in the Nominative case.

Speakers often volunteer simple adjectival predicate clauses in such a way that the copula markers *-tte* and *-tta* correspond to masculine and feminine genders respectively, as in (7). However, when explicitely asked whether it is possible to exchange the copula markers in these cases, their response was positive.

(7) a. *?ésí túlle-tte*
3MS:NOM deaf-COP
'He is deaf'

b. *?ísí* *túlle-tta*
3FS:NOM deaf-COP
'She is deaf'

The suffixes *-tte* and *-tta* are attached to a predicate nucleus when this is in a present or non-past tense. In the past tense, the verb *yéšše* is used. An example:

(8) *?eeress* *yéšš-inna* *baye* *tá* *ham-é-ss-í* *alámu-so*
small BE.PAST-TEMP NEG.Q 1SG go-PF-REL-NOM Alemu-place
'Wasn't it when I was small that I went to Alemu's place?'

The verb *yéšše* in (8) above is the past tense form of the existential as well as copula verb *yene* "it is/there is":

(9) a. *zargúla* *gídda-iš* *yéne* *yílo-y*
Zargulla inside-F exist Yilo-NOM
'Is Yilo (place name), within Zargulla territory?'

b. *hoó* *zargúla* *gíddá-tte* *yéne*
yes Zargulla inside-COP exist
'Yes, it is within Zargulla'

For only an existential expression the verb *yése yesa* "there is" is used as in (10) which can be analysed as a locative copula in the sense proposed by Hengeveld (1992).

(10) *keté* *keetsé* *gid-é-s-í* *malo-í* *yésa*
earlier house:ACC happen-PF:REL-NOM stone-NOM EXIST
'Earlier there was a stone as big as a house'

The distribution of the copula markers *-tte* and *-tta* corresponds the overall gender-marking pattern observed within noun phrases and in subject agreement-marking in verbs (see discussion above). That is, when the subject of a verbal sentence is inanimate, it is often treated as a feminine noun, e.g., through using third person feminine agreement inflection on the verb. Similarly, when the subject of a predicative nominal is inanimate, the commonly used copula morpheme is *-tta*, which is related to the feminine gender (11).

(11) a. *mins'ŏ-y* *dic'ŏ-tta*
tree-NOM big-COP
'The tree is tall/high'

b. *?ísí* *dees'ŏ-tta*
3FS:NOM heavy-COP
'It is heavy' [referring to a plastic jerrycan.]

This is consistent with the use of the third person singular feminine subject-agreement marker in main verbs for co-reference with inanimate subject nouns.

In "weather expressions" involving a non-verbal predicate, the copula marker *-tta* is used.

(12) a. *gádé-y* *?awá-tta*
earth-NOM sun-COP
'It is sunny'

b. *gáde-í* *?írá-tta*
earth-NOM rain-COP
'It is rainy'

Other contexts where the copula marker *-tta* is consistently attached to the predicative nominal is when the subject is the infinitive:

(13) a. *muús-í* *ló?ó-tta*
eat.INF-NOM good-COP
'eating is good' (Infinitive: *muús*)

b. *?úšš-í* *ló?ó-tta*
drink-NOM good-COP
'drinking is good' (Infinitive*: ?úšš*)

c. *zargúla* *yew-é-y* *metut-s-eé-s-ta*
Zargulla come-INF-NOM be_in_trouble-CAUS-IPF: REL-NMZ-COP
'coming to Zargulla (area) is difficult' (Infinitive: *yewe*)
(*lit.* 'coming to Zargulla is one which causes one to be in trouble')

With demonstratives indicating inanimate subjects too, we find the feminine copula marker *-tta*. Unlike pronouns, the consonant of the copula marker attached to demonstratives is geminated as in noun phrases:

(14) a. *hií-tta* *s'unduk'a* *híyy-útt-e-s-í*
that-COP *s'unduk'a* say-PAS-PF:REL-NMZ-NOM
'It is *that* which is called *s'unduk'a* (*s'unduk'a*' = a kind of food)

b. *sénní-tta* *ló??ó-y*
that-COP good-NOM
'*That* is better' (said in reference to a suggestion on where to place a guest for the night)

The use of the feminine form *hoó ?ísta* "yes, that is it!" (lit. "That is her") in reaction to an addressee's voluntary information which we mentioned earlier is also parallel to the selection of the feminine subject agreement form in verbs or the use of the feminine copula form when a nominal predicate is an inanimate subject.
Plural predicative-nominals also take *-tte* or *-tta* ending as demonstrated in (15).

(15) *mins'i* *zagar-eď e-í* *dic'-eď é-tta*
tree:GEN branch-PL-NOM tall-PL-COP
'The branches of the tree are big'

As we demonstrated in this section, the copula markers *-tte* and *-tta* are related to masculine and feminine gender respectively. However, it seems that gender-marking through these morphemes is becoming of secondary importance, since there are many examples where the two can be used interchangeably. We assume that cases like those in examples (5–6), where the two forms contrast, reflect retention of a formerly widely used gender distinction. Moreover, there are other semantic contrasts expressed in non-verbal predicates that are marked with *-tte* and *-tta*, as discussed in the next section.

3. Other meanings of *-tta* and *-tte*

In some contexts, the copula markers *-tte* and *-tta* express specialized meanings which are not related to gender. For example, in questions and answers, a predicative construction with *-tta* expresses that the respondent is annoyed by the question. In the same context, a predicative construction with the copula *-tte* is interpreted as a direct informative response to the question. Compare the interpretation of (16b) and (16c) which are given as response to the question in (16a):

(16) a. *ha sínnó ments-é-ss-í ?oóde-wa*
this cup-ACC break-PF:REL-NMZ-NOM who-COP:Q
'Who broke this cup?'

b. *tán-te* (*dem-áčče*) *ments-é*
1SG-COP see-SUBR:NEG break-PF:REL
'It is me who (accidentally) broke it' (Lit. "It is me who broke it, (without seeing it)")

c. *tán-ta ments-é*
1SG-COP break-PF:REL
'It is me who broke it'

Example (16c) implies that the speaker is annoyed by the question or that he broke it purposely. It can be understood for example, as "It is mine and I broke it, why do you bother". In example (16b) the predicative form cannot be replaced by *tánta* "it is me" because this would be incompatible with the meaning of *demácce* "without seeing/accidentally". If the predicative form in (16c) is replaced by *tante* "it is me", the speaker is simply stating the accident as a matter of fact and most likely he is not expecting any reproach from the speaker who uttered (16a) or he would offer an excuse or explain how the situation took place.

A related use of the copula *-tta* is illustrated in (17), which expresses a warning that the speaker wants that no body else interferes in the plan and makes the trip to Arbaminch. He might make this utterance even when he is going to somewhere else at the moment of speech.

(17) *arbamínč'e* *hang-eé-s-í* *tán-ta*
Arbaminč go-REL:IPF-NMZ-NOM 1SG-COP
'I will go to Arbaminch' (lit. 'The one who will go to Arbaminch is me')

On the other hand, the same sentence with the copula marker *-tte* does not have such an interpretation. It can be used as a factual statement, for example when the speaker wants to introduce him/herself to someone who would drive him to the destination mentioned:

(18) *arbamínč'e* *hang-eé-s-í* *tán-te*
Arbaminč' go-REL:IPF-NMZ-NOM 1SG-COP
'I will go to Arbaminch' (lit. 'The one who will go to Arbaminch is me')

As we have seen earlier the copula marker *-tta* normally co-occurs with feminine gender subject complements. It is not plausible to relate the expression of "annoyance" and "threatening" to the function of marking gender. A more plausible analysis is that as gender distinction in copula constructions is losing its importance, this opened a possibility for a reinterpretation of *-tta*.

4. Focus and the copula marker *-tte*

The morpheme *-tte* has also acquired an extended grammatical function of marking focus. It may be attached to a verb or any other contrastively focused constituent within the sentence. In focused verbal clauses, on the other hand, the focus marker is immediately followed by one of a set of seven suffixes which co-vary according to the person, gender and number of the subject noun. In the examples in (19), the highlighted constituent (in italic) is focused.

(19) a. na?á-z-í kátsa *bays-í-tte-s* maa?ó šang-í
child-M-NOM grain:ACC sell-CNV-COP-3MS cloth:ACC buy-CNV
t'um-us-í yeénne
be.dark-CAUS-CNV come:PAST
'The boy *sold* grain, bought cloth and came late'

b. na?á-z-í kátsa bays-í *maa?ó-tte-s* šang-í
child-M-NOM grain:ACC sell-CNV cloth:ACC-COP-3MS buy-CNV
t'um-us-í *yeénne*
be.dark-CAUS-CNV come:PAST
'The boy sold grain, bought *cloth* and came late'

c. *na?á-z-í* *kátsa* *bays-í* *maa?ó* *šang-í-tte-s*
child-M-NOM grain:ACC sell-CNV cloth:ACC buy-CNV-COP-3MS

t'um-us-í *yeénne*
be.dark-CAUS-CNV come:PAST

'The boy sold grain, bought *cloth* and came late'

d. *naʔá-z-í* *kátsa* *bays-í* *maaʔó* *šang-í*
child-M-NOM grain:ACC sell-CNV cloth:ACC buy-CNV

t'um-us-í-tte-s *yeénne*
be.dark-CAUS-CNV-COP-3MS come:PAST

'The boy sold grain, bought cloth and came *late*'

e. *naʔá-z-í* *kátsa* *bays-í* *maaʔó* *šang-í* *t'um-us-í*
child-M-NOM grain:ACC sell-CNV cloth:ACC buy-CNV be.dark-CAUS-CNV

yeett-á-tte-s-ínne
come-TV-COP-3MS-PAST

'The boy sold grain, bought cloth and *came late*'

The masculine copula morpheme *-tte* is also used in cleft constructions, which is the strategy to express subject focus.

(20) a. *kaat-í* *bíšáʔa-tte* *waatsé* *kís'-a* *ham-é*
king-NOM daughter-COP water:ACC draw:INF go-PF:REL
'It is a king's daughter that went to draw water'

b. *taa* *s'aď́-tte* *míy-é*
1SG:OBJ flea-COP eat-PF:REL
'Fleas bite me' (lit. 'me, it is fleas that bite')

c. *zilangó-tte* *táná* *meičč'-e-í* *yaá-tte-ínne*
ant-COP 1SG:OBJ bite-PF:REL-NOM that_say-COP-PAST
'(The monkey) said 'It is ants that bite me"

The only difference between cleft constructions and focus and copula constructions is that in cleft constructions, there is always a relative clause following the constituent that is highlighted in the cleft construction. Also, note that in such constructions, the subject agreement marker does not occur after the focus/copula marker.

As we tried to show through a quotation from Fleming (1976) in the introductory section of this paper, in Omotic languages, grammatical gender is not widely used. Frequently only two gender distinctions, masculine and feminine are made. The distinction between these two is important for larger animate nouns. All other nouns may be generalized to one of the two classes of animate nouns or they may alternatively take feminine or masculine gender depending on size or other extra linguistic criteria. Nevertheless, some Omotic languages, including Zargulla have various means to indicate feminine and masculine gender in singular nouns. In Zargulla, the distinction

is morphologically marked on nouns, third person singular pronouns, on the verb and, as we demonstrated in the present contribution, partly on copula construction. However, in the latter case, the distinction is not consistently used.

Abbreviations

ACC	accusative	M	masculine
CAUS	causative	OBJ	object
CNV	converb	PAST	past tense
COP	copula	PF	perfective aspect
DAT	dative	PL	plural
DEF	definite	Q	question
F	feminine	REL	relative clause
GEN	genirive	S	singular
IMPS	impersonal	SUBJ	subject
IPF	imperfective aspect	TEMP	temporal (when white)
LOC	locative	TV	terminal vowel

References

Fleming, Harold. 1976. "Cushitic and Omotic." *Language in Ethiopia*. Ed. by M. Lionel Bender, et al., 34–53. London: Oxford University Press.

Hayward, R.J. 1989. "The notion of 'default gender': a key to interpreting he evolution of certain verb paradigms in East Ometo, and its implication for Omotic." *Afrika und Übersee* 72: 17–32.

Hengeveld, Kees. 1992. *Non-verbal Predication: Theory, Typology, Diachrony*. Berlin, New York: Mouton de Gruyter.

Hirut Woldemariam. 2003. *The Harro Language*. Ph.D. dissertation, Cologne University and Addis Ababa University.

Pustet, Regina. 2003. *Copulas: Universals in the Categorization of the Lexicon.* Oxford: Oxford University Press.

Stassen, Leon. 1996. *Intransitive Predication*. Oxford: Clarendon Press.

Wetzer, Harrie. 1996. *The Typology of Adjectival Predication.* Berlin, New York: Mouton de Gruyter.

Riddling in Gidole

Paul Black
Charles Darwin University

This is a brief account of riddling in the Gidole and Konso languages of southwestern Ethiopia, including two examples of Gidole riddling exchanges and ten additional Gidole riddles.

I owe a huge debt of gratitude to Hal Fleming for his support over a few years in my few years as a Cushiticist in the later 1960s and early 1970s. I first made contact with him sometime before 1970, when I went to Ethiopia to undertake doctoral research on the Eastern Cushitic languages there. During 1971–72 he also came to Ethiopia and I was able to spend time with him and his family, benefiting greatly from both his academic and emotional support. He became an associate supervisor for my doctoral thesis, and as I struggled to write it up in early 1974, he and his family kindly took in my wife and infant daughter for several months so I could focus on my work. I'm afraid I repaid my great debt to him poorly, by abandoning Cushitic languages as a research fellowship took me to Australia to work on the indigenous languages there.

My doctoral thesis (Black 1974) was a comparative study of the Lowland East Cushitic languages, and for that purpose I was focussing primarily on vocabulary and grammar. As part of this work I gathered data on varieties to the southwest of Lake Chiamo that constitute either a chain of highly differentiated dialects or else a set of very closely related languages (see Black 1992). In the northern part of this area I worked mostly on one of the more western varieties that were spoken around the hill where the town of Gidole was located and thus were collectively known as Gidole. I recorded the indigenous name of this western variety as *kap D'iraasháte*, the language of *D'iráash*, while eastern varieties were called *D'iraytát*. In the rolling country along the western side of the valley south of Gidole I studied varieties of what is generally known as Konso, although the indigenous name of those I worked most on was *áfaa Karátti*', the language of *Káratti*.

For the most part I focussed on the grammar (Black 1973a) and vocabulary (Black 1973b; Black & Otto 1973) of these languages. While I also recorded a number of texts, I did not attempt to report on them before, aside from a reference to an account of Konso origins in Black (1976). In Hal's honour I here provide an account of some Gidole riddles I recorded in early 1971.

The Gidole (and Konso) root for "riddle" *hiipp-*. This root alone can be used as a verb meaning "to tell a riddle", while the corresponding noun for "riddle" is Gidole *hiippót* (Konso *híippaa*). These riddles seem to be conventionalised, with answers that are difficult to guess; as in the case of Western knock-knock jokes or the so-called elephant jokes, it seems that you either know the answer or you don't. They seem broadly similar to the Oromo ("Galla") riddles published long ago by Cerulli (1922); I have not yet seen Mous' (2000) more recent paper on riddling in Iraqw.

I did not observe the actual practice of riddling, but I was told that riddles might be exchanged by people working on some common task, such as (to use a Konso example) women taking seeds out of cotton bolls or spinning thread. For Gidole my informant fabricated two examples of the sort of interaction that might be involved, and after I give these I will list ten other riddles that I recorded.

I regret that my transcription is imprecise and I have not been able to check it fully. Gidole has a system of five vowels *i, e, a, o, u* that can occur both short and long, with long vowels written as double. While I ended up transcribing some words with final consonants, I now suspect that morphophonemically, at least, these are better taken to have final short vowels, while the final short vowels I recorded are morphophonemically long. I've marked contrastive accent as acute, but I never worked out its complications. Gidole has stop consonants *p, t, c*, and *k* that vary from fully voiced (as between vowels) to voiceless (when doubled). Like Amharic it has glottalised stops *p', t', c'*, and *k'*, and like Konso it also has an implosive *d'*. In other environments the apostrophe (') is used to represent a glottal stop. Voiceless fricatives are transcribed as *f, s, sh,* and *h*, nasals as *m, n, ny*, and *ng* (although I agree with Hans Jürgen Sasse (personal communication) that *ng* it merely an allophone of *n* found before velars and the glottal stop), and there is a lateral *l*, rhotic *r*, and the semivowels *w* and *y*.

In the first example of Gidole interaction speaker B is able to guess the answer to A's riddle:

A:	*Hiippó.*	[Said by someone ready to tell a riddle.]
B:	*Hipáyt.*	[Response to the above.]
A:	*Himpírru mínas mánaw hiip.*	A fine house up on Himpirru [a place?].
B:	*K'órayyám?*	[Is it] wood?
	Maném?	A house?
	K'udd'éetem?	A thorn?
	Kaasa kamanáte?	Horn of a cow?
	'Innúkk'oyyi.	[expression of puzzlement?]
	'Innole kalanem?	Is it living with us?
	'Innole hekiyanem?	Is it with us?
A:	*Heyyé. Hekiyani.*	Yes. It's is.

B:	*K'ŏrayyám?*	Is it wood?
A:	*Heyyé k'ŏrayya.*	Yes, it's wood.
	Sherk' sip mina kiyyu (kiď á).	(Say) the hook at the end of a spindle.

Frankly, I have no idea why A conceded the answer to B here; the hook of a spindle seems a far cry from B's guess of *k'ŏrayya*, a plural of *k'oyr* "tree" that I also recorded as meaning "firewood" and which I have thus translated as "wood" here. Note also how the word *kiď á* "say" at the end of the answer is apparently optional: for this riddle it was given only when the basic riddle was repeated later, without the interaction, and you will also see it in the answer of some but not all other riddles. A grammatical feature rather obvious in the above is the final question marker *-m*.

The second Gidole example shows what happens when B fails to guess the answer: B has to give A a "country", which in this case is Konso.

A:	*Hiippó.*	[Said by someone ready to tell a riddle.]
B:	*Hipáyt.*	[Response to the above.]
A:	*'É 'attátt 'on'áalel táwé.*	I am white and I'm going away to 'Aaleel.
B:	*Háyď am?*	Is it fat?
	Fuuttám?	Is it cotton?
	Fák'anam?	Is it wood?
	Púsurk'ám?	Is it chips?
	Maaná?	What is it?
A:	*Teek 'ampayyaam mok?*	Are they poor old people?
	'Innu' ď aattinkiď aw.	Say what you'll give us.
B:	*Maan a'íshaaki?*	What do you answer?
	'Ádď 'afoottintáne.	You can't.
A:	*Maan 'ahintane?*	What do you want?
	Maan 'an'inaď aanin?	What will we give you?
B:	*Káratt.*	Konso.
A:	*'Am Karatt 'inaď annahinam.*	We don't want to give you Konso.

Here are ten examples of Gidole riddles, beginning with the above and its expected answer.

1.	*'É 'attátt 'on'áalel táwé.*	I am white and I'm going away to 'Aaleel.
	'Ánc'uha hawk'etat kiď á.	Say spitting out remains of sugarcane stalk.
2.	*Márakatta mac'ámmang 'aan.*	I went on the slope with a shoe.
	Írrimash.	Anthill.

3.	*Namse heel tooyu d'ama d'ayi.*	Give food to the man who's looking at you.
	Fild'a.	A small hole or crack [e.g., in a pot].
4.	*Usaakan het 'inanshaw pat róopane*	When I put butter on my daughter's back,
	'am k'ar 'elíyyam.	it doesn't stay.
	Láappa káat k'ar het hak'ak'aan	When we put water on *kaat* leaves,
	'am k'ar 'elíyyam.	it doesn't stay.
5.	*'Immawn 'ollat 'allap torp;*	I throw my son into battle;
	'iyat 'ollo 'akkanataamale mitaat.	he returns without danger.
	K'olm kid'á.	Say a stirring stick.
6.	*Hetting 'ant 'aningkiyyo 'anto' he'at,*	When I'm going I am white,
	helpa 'ant mitaatingkiyyo 'ant heroom.	When I return I am red.
	Dáma.	Food.
7.	*Helpa 'ant 'aningkiyyo 'ant hetuf,*	When I'm going I am hungry,
	helpa 'ant mitaatingkiyyo ant hek'uuh.	When I return I am full.
	'Áakal kidá.	Say sack.
8.	*Hintílí k'oomód'.*	Bend up and down.
	C'úhuuna.	A worm.
	Shokke páac'a kid'á.	Another [answer] is a sickle.
9.	*'Ilshiw 'awmit Móoha tooy.*	My eye always looks to God [or Venus].
	Síitta teltetet kid'á.	Say tail of a goat.
10.	*Páana Mashóole márk'am mark'a.*	The road to Mashile is meandering.
	Sínd'a horma kid'á.	Say the urine of an ox.

References

Black, Paul. 1973a. "Draft sketch of Konso phonology, morphology, and syntax." Unpublished manuscript.

Black, Paul. 1973b. "Gidole dictionary." Unpublished manuscript.

Black, Paul & Shakko Otto. 1973. "Konso dictionary." Unpublished manuscript.

Black, Paul, 1974. *Lowland East Cushitic: Subgrouping and reconstruction*. Unpublished Ph.D. thesis, Yale University.

Black, Paul. 1976. "Some linguistic evidence on the origins of the Konsoid peoples." *Proceedings of the First U.S. Conference on Ethiopian Studies*, Ed. by H. Marcus & J. Hinnant, 291–302. East Lansing, Mich.: African Studies Center, Michigan State University.

Black, Paul. 1992. "Konsoid: An example of extreme dialectal differentiation." *Jimmon Shakai Kagaku Kenkyuu / Journal of Humanities and Social Sciences* (Waseda University, Tokyo). 32. 99–104.

Cerulli, Enrico. 1922. "The folk-literature of the Galla of Southern Abyssinia." *Harvard African Studies*, vol. 3, Ed. by E.A. Hooton & Natica I. Bates, 9–228. Cambridge, Mass.: African Department, Peabody Museum, Harvard University; electronic version obtained from http://www.samizdat.com/cerulli.html and http://www.samizdat.com/cerulli2.html on 27 June 2006.

Mous, Maarten. 2000. "Ka babay kababiyo: Riddling in Iraqw." *Afrikanistische Arbeitspapiere* 62 (June). 39–69.

PART III

African languages – Classification and prehistory

A lexicostatistical comparison of Omotic languages

Václav Blažek
Masaryk University (Brno) &
Western Bohemian University (Plzen)

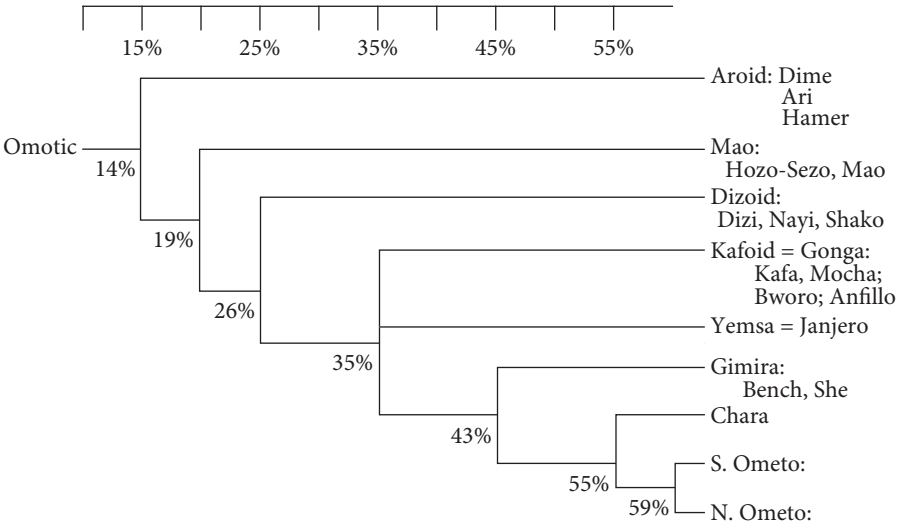
The taxonomy of the Omotic language family is thoroughly analyzed. The hypotheses of "West Cushitic" *vs.* a "Cushomotic" super-branch, *vs*. separating Omotic as the sixth branch of Afroasiatic (Fleming) are compared. Earlier classifications of the Omotic family are presented, followed by the author's version based on S.A. Starostin's variant of glottochronology. According to the latter the North/South Omotic disintegration can be dated to the beginning of the 5th mill. BC. Borrowings can be distinguished from inherited lexicon, though often one can only speculate on the direction of borrowing. Nilo-Saharan parallels were also taken into account. The numbers of cognates common to Omotic and other Cushitic branches on the one hand are contrasted with those common to Omotic and the other Afroasiatic families on the other. It seems most natural to see in Omotic an independent member of the Afroasiatic macro-phylum.

1. Internal classification
2. Phonetic correspondences
3. Basic lexicon
4. Etymological comments

1. Internal classification

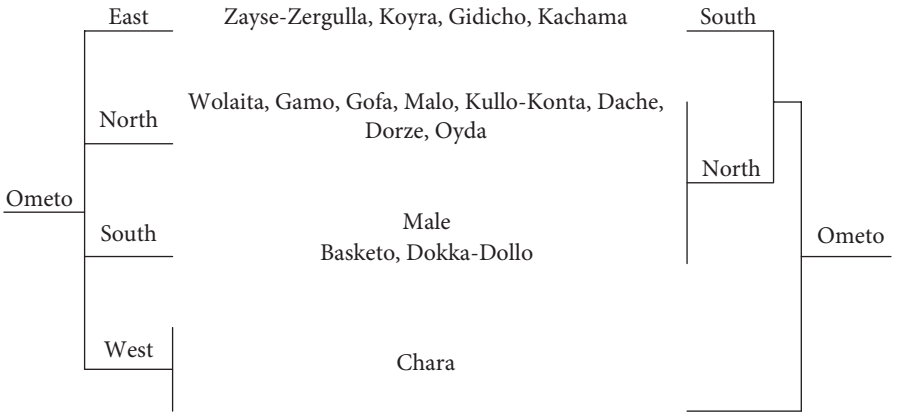
The Omotic languages represent the most controversial part of the Afroasiatic language (macro-) family. For a long time these languages were classified as "West Cushitic", and to this day some scholars (Lamberti & Sottile) maintain this position, while others posit a "Cushomotic" super-branch (later Zaborski and Bender). On the other hand, there are some arguments (especially lexical) for separating Omotic as the sixth branch of Afroasiatic (Fleming 1969; Bender 1971, 1975). Bender has devoted a series of studies to the problems of genetic classification of Omotic (1971; 1988. 132–33; 1990. 587; 2000. 2) (see Table 1).

Table 1. Internal classification of Omotic (Bender 1988.133)



Recently two models of the genetic classification of the Ometo dialect continuum have been proposed: by Fleming (1976) and by Bender (1988): here left and right respectively. Their results can be compared as follows (see Hayward, *OLS* xi–xii):

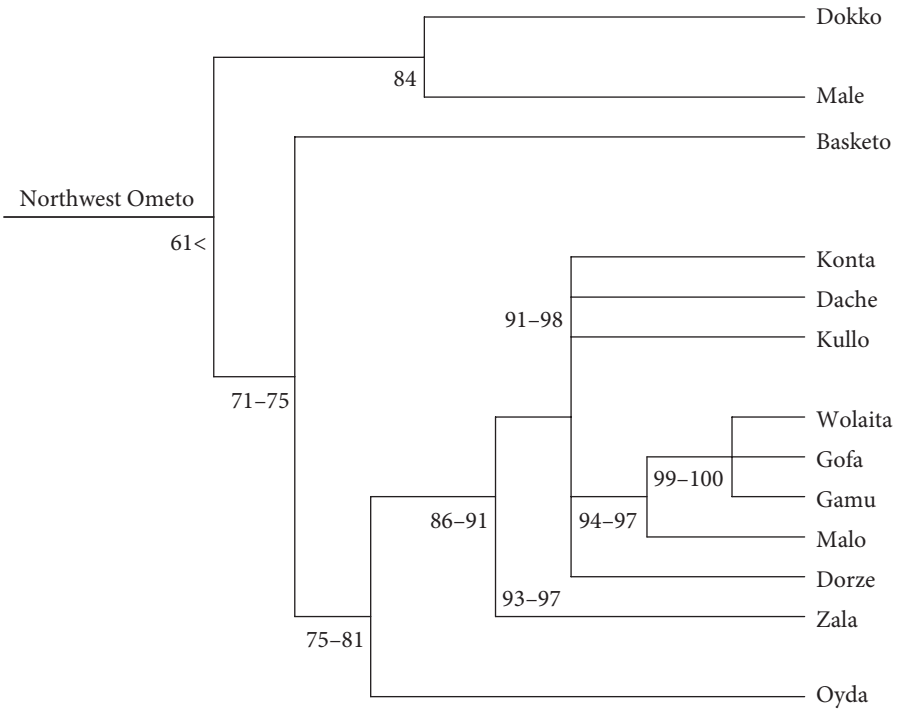
Table 2. Classifications of Ometo dialect continuum (Fleming 1976; Bender 1988)



In his most recent monograph Bender (2003) offers more exact results. He divides the Ometo cluster into three parts, Northwest, Southeast and Chara. The internal structure of the Northwest group of Ometo can be constructed on the basis of

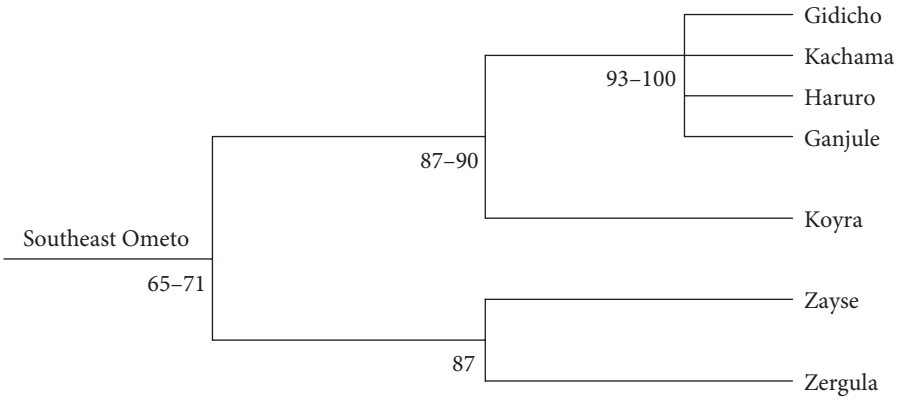
Bender's figures as follows (Table 3; the figures represent percentages of common cognates):

Table 3. Internal classification of Northwest Ometo (Bender 2003. 51)



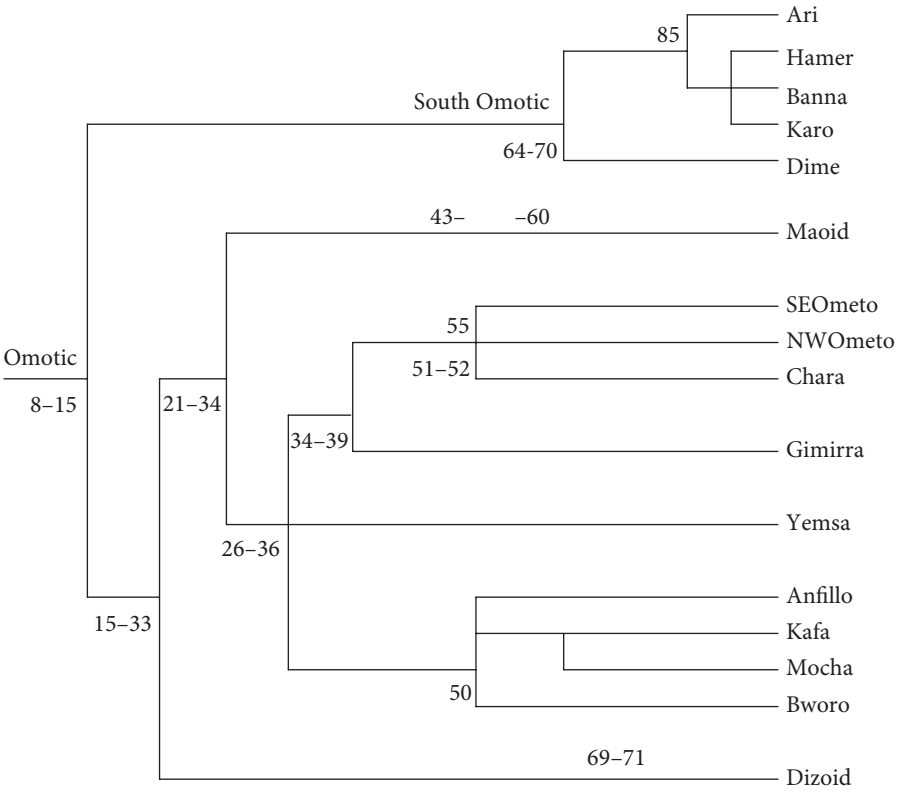
The internal classification of Southeast Ometo (Table 4):

Table 4. Internal classification of Southeast Ometo (Bender 2003. 79–81)



The Omotic language family is classified by Bender (2003) as follows (Table 5):

Table 5. Internal classification of Omotic (Bender 2003. 156–57, 203–04, 286)



2. Phonetic correspondences

There have been several attempts to establish regular phonetic correspondences among the Omotic languages. Since 1990 four important studies have been devoted to this problem: Lamberti & Sottile (1997), Bender (1988), Hayward (1988), Ehret (1995). Lamberti & Sottile (1997) compared 17 Omotic languages in the following table (Table 6). The abbreviations are: Wo Wolayta, Za Zala, Da Dawro = Kullo, Ma Malo, Gf Gofa, Ga Gamu, Dc Dache, Zy Zayse, Ky Koyra, Ka Kachama, Ca Chara, Ye Yemsa, Bw Bworo, Kf Kafa, Mc Mocha, Be Bench, Ar Ari.

Table 6. Phonetic correspondences of Omotic languages (Lamberti & Sottile 1997. 253–260)

Wo	Za	Da	Ma	Gf	Ga	Dc	Zy	Ky	Ka	Ch	Ye	Bw	Kf	Mc	Be	Ar
b-	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b
-b-	w	b	b/w	b	b/w	b/w	b/w	w	b	b	b	b	b	b	–	b
-p'-	ʔ	b	p'	p'/ʔ	p'						p'	p'	p	p'/f	p'	p'
f-	f	f	p	p/f	p	f	f/p	f	p	f	f	f	h/f	f	p	f
-f-	f	f	p	f/p	f/p/b	f/p/b	f/p/b	f/b	p	f	f	f/w	f	f	p	f/b
m-	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
-m-	m	m	m	m	m	m	m	m	m	m	m	m/w	m	m	m	m
w-	w	w	w	w	w	w	w	w	w	w	w/0	0	0	0	w	w
-w-	w	w	w	w	w	w/y	w	w/y	w	w	w	w/b	b	b	b	0/b'
d-	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
-d-	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
t-	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t/d
-t-	t/ts	t/ts	t/ts	t/ts/c	t/ts	t/ts	t/ts	t/ts	t/ts'	t/ts	t	t/š/c/r	t/š/c/d	t/š/c/d	t/ts	t/ts/d
t'-	ts	t'/ts'	s'	ts'	ts'/d'	ts'/d'	ts'/d'	ts'/d'	t'/ts'/d'	c'/ts	t	ts'	c'/t'	c'/t'	ts'	c'/d/ts'
-t'-	ts'	ts'/t(')	s'	ts'	ts'	ts'	ts'	ts	ts'	–	t/r'/ʔ	ts'	c'/t'	c'/t'	ts(')	ts/d
z-	z	z/s	z	z	z	z	z	z	z	z	z	j	y	y	z	z
-z-	z	z/s	z	z	z	z	z/s	z/s	–	z	s/z/y	z	j	j	z	z
s-	s	s	s	s	s	s	š/s	s	s	s	s	š	š	š	s	š/s
-s-	s	s	s	s	s	s	s	s	s	s	s	š	š/c	š/s	s/š	s
š-	š	š/s	š	š	š	š	š	š	š	š	š	š	š	š	š	š

(*Continued*)

Table 6. (Continued)

Wo	Za	Da	Ma	Gf	Ga	Dc	Zy	Ky	Ka	Ch	Ye	Bw	Kf	Mc	Be	Ar
-š-	š	š	š	š	š	š/c	š/c	c/š	š	c/š/s	š	š/y/z	š/y/c	š/y/s	c/ch	c/š
n-	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n
-n-	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n
l-	l	l	l	l	l	l	l	–	l	n	l	n	n	n	–	l
-l-	l	l	l	l	l	l	l	l	l	l/r	r/l	r/l	r/l	r/l	r/l	l
-r-	r	r	r	r	r	r	r	r	r	r	r	r	r	r	r	r
g-	g	g	g	g	g	g	g	g	g	g	g	g	g	g	g	g
-g-	g	g	g	g	g	g	g	g/k	g/k/k’	g	g	g/k’	g/k’	g/k’	g/k’	g/k
k-	k	k	k	k	k	k	k	k	k	k	k	k	k	k	k	k
-k-	k/k’	k	k	k	k	k	k/g/h	k/h	k/h	k	k	k/h/p	k/p	k/p	k/p	k/x
k’-	k’	k’	k’	k’	k’	k’	k’/ʔ	k’	k’	k’	k	k’/g	k’/g	k’/g	–	k(,)/g
-k’-	–	k’	k’	k’	k’	k’	k’/ʔ	k’	k’	k	ʔ/k/g	k’/0	k(,)/g	k’/g	k’/s	k’/ʔ/x
c’-	–	c(ʔ)	c’	c’	c’	c’	c’	c(,)	c’	–	t/c/k	ts’/c’/k’	c’/k’	c’/k’	ch’/c’	c’
-c’-	c’	c’	c’	c’	c’	c’	c’/ʔ	c’	c’/t’	c’	c/t/ʔ	c’/t’/ts’/š	c’/k’	c’/t’/k’/s	s	c’/k’
-c-	c/c’	c	c’/t/š	c	c/t	c/t	c/š/t	š/c(,)	s/t/c(’)	c/ts	c/j/š	ts/d/š/ts’	c/c’/d	c/c’/d/t’	c/ts’/d	š/c
y-	y	y	y	y	y	y	y	y	y/0	y	y/0	0	0	0	y	0
-y-	y	y/0	y/0/ʔ/z’	y/0/ʔ	y/0/ʔ	y/0/ʔ	y/0/ʔ/š/s	y/0/ʔ/š	y/0/ʔ	y/0/ʔ/š	0/s/š(y/w)	0/š(y/w)	0/c(y/w)	0/s(y/w)	0/y/s	y/0
ʔ-	ʔ	ʔ	ʔ	ʔ	ʔ	ʔ	ʔ	ʔ	ʔ	ʔ	ʔ	ʔ	ʔ	ʔ	ʔ	ʔ
-ʔ-	ʔ	ʔ/h	ʔ	ʔ	ʔ/0	ʔ	ʔ/0	ʔ	ʔ/0	ʔ/0	0	ʔ/k’	0/g/ʔ	ʔ/k’	ʔ/k’	0/x/k(’)
h-	h/ʔ	h/ʔ	h/ʔ	h	h	h/ʔ	h	h/ʔ	h	ʔ	ʔ	ʔ/h/k	h	h/k	h	ʔ/k/h
-h-	h	h	h	h/k	h/k/g	h/k/0	h/k	k/g	h/k	h/b	0/b/w	g/h/f	g/h/p	h/g	k	x

Bender (1988) reconstructed proto-Omotic consonantism on the basis of the following correspondences (Table 7: additional abbreviations: Oy Oyda, Ba Basketo, Me Male, Gn Ganjule, Gi Gimira, An Anfillo, Dz Dizi = Majo, Sk Sheko, HS Hozo-Sezo, Mo Mao, Ha Hamer, Di Dime):

Table 7. Phonetic correspondences of Omotic consonants (Bender 1988. 136)

	Labials				**Dental/Alveolars**							**Palatals**					**Post-Pals**	
	pʾ	ɓ	p	f	tʾ	tsʾ	ɗ	ts	dz	s	z	cʾ	c	j	ž	ɲ	ŋ	ʔ
Wo	pʾ		p	f	tʾ, ɗ		ɗ			s	z	cʾ	c	j				ʔ
Da		ɓʔ		f	tʾ, ɗ	tsʾ	ɗ	ts		s	z		c	j				ʔ
Ma	(pʾ)	(ɓ)	p		ɗ	tsʾ	ɗ	ts	(dz)	s	z	cʾ	-c-					ʔ
Oy			p		ɗ	tsʾ	ɗ	/ts	(dz)	s	z	cʾ	/c					ʔ-
Ba				f	ɗ	tsʾ	ɗ	ts		s	z	cʾ	-c-		(ž)		/ŋ	ʔ
Me		ɓ	p	f	ɗ	tsʾ	ɗ	/ts		s	z	cʾ	c	j	ž			ʔ
Zy			p		ɗ, ts		ɗ	ts		s	z	cʾ	c	ž	ž			ʔ
Ky	-pʾ-	(ɓ)	p		ɗ	tsʾ	ɗ	ts	(dz)	s	z	cʾ	c	j	ž			ʔ-
Gn		ɓ	p	f-	ɗ	tsʾ	ɗ	/ts	(dz)	s	z	cʾ	/c	j				-ʔ-
Ch			p	f	tʾ	tsʾ	ɗ	-ts-		s	z	cʾ	-c-	j				-ʔ
Gi			p	f	tʾ	tsʾ		/ts		s	z	cʾ	/c	(j)	ž		/ŋ	ʔ
Ye			p	f	tʾ					s/š	z		c	-j-		-ɲ-	-ŋ-	-ʔ-
Kf	pʾ		p	f	tʾ	cʾ				s		cʾ	c	j				ʔ
An	pʾ?	**bʔ**	p	f	tʾ	tsʾ		ts				cʾ	c	j				ʔ
Sh	-pʾ-			f	tʾ	tsʾ	ɗ	ts			z	cʾ		j	ž	ɲ		ʔ
Dz				f	tʾ	tsʾ				s	z	cʾ	c	j	ž		-ŋ	
Sk				f	(tʾ)	tsʾ				s	z	cʾ	-c-		/ž			
HS			p		tʾ	tsʾ	0	(t)s		(t)s	z							
Mo	pʾ		p		tʾ	tsʾ	0			s	z	cʾ?	(c)				ŋ?	
Ar		ɓ	p	f	(tʾ)	(tsʾ)	ɗ	/ts		s	z	cʾ	c		ž		/ŋ	ʔ ʔ
Ha		ɓ	p		tʾ		ɗ	ts		s	z		c	j		ɲ	ŋ	
Di				f			ɗ	ts		s	z	cʾ	/c	-j-ʔ	/ž		/ŋ	

Analyzing the Omotic sibilants, Hayward (1988) assumed the following system (Table 8).

Table 8. Phonetic correspondences of Omotic sibilants (Hayward 1988)

Initial position	***s**	***z**	***ts'**	***š/ş**	***ž/ẓ**	***č'/ç'**
Wolaytta	s	z	t'	š	y	č'
Zayse	s	z	ts'	š	ž	č'
Bench	s	z	ts'	š/ş	ž/ẓ	č'/ç'
Yemsa	s	z	t	š		č/t
Kafa	s	y	č'	š		č'
Mocha	š	y	č'	š	š	č'
Ari	s	z	ts'	š	ž	č'

Medial position	***s/*ss**	***z**	***ts/*tts**	***ts'/*tts'**	***š/*šš**	***ž**	***č/*čč**	***č'**	***ç'**	***ç**
Wolaytta	s	z	tt	t'/tt'	/šš		š/čč	č'	č'	
Zayse	s/ss	z	ts/tts	ts'/tts'	š/šš		č/čč	č'	č'	
Bench	s		ts	ts'			č		ç	ç
Mocha	y/ss	y	š/čč	č'/čč'	y/ss	y	š/čč	č'	č'	š
Kafa	y/ss	y	š/čč	č'/čč'	y/čč	y	š/čč	č'	č'	
Ari				ts'	š		š/čč			

Finally Ehret (1995) has presented his own set of correspondences (Table 9):

Table 9. Phonetic correspondences of Omotic consonants (Ehret 1995. 10–12)

P-Om	**Wo**	**Me**	**Zy**	**Ye**	**Kf**	**Mc**	**Be**	**Dz**	**Di**	**Ar**
*b/ʔ	b	b	b	b	b	b	b	b	b	b
/V_V	w	w	w	w	b, w/_ʔ	w	b	b		b
*p/ʔ	f	p	p	f	p	p	p		f	f
/V_	f	f	p	f	f	p	p	p	f	f
*p'	p'	p'	'p	b		b, p/V_	p'	b		p'~ɓ/V_
*d/ʔ_	d	d	d	d	d	d	d		d	d
/V_	d		d	d	d	d	d		d	d
*t	t	t	t	t	t	t	t	t	t	t
*z/ʔ_	z	z	z	z	y	y	z		z	z
/V_		d		d	y	y	z		d, z	d, z
/Vy_	z	ts	ts	z	y	jj	z	d		
*zz					jj	jj				
*s	s	s	s	s	š	š, s/_	s	s	s	s
*ts'	t'	ts'	s'		č'	č	ts'	ž/V_	ts'	ts'
*g	ɗ	ɗ	ɗ	ɗ	ɗ	ɗ	ɗ	ɗ	ɗ	ɗ
*k	k	k	k	k	k	k	k	k	k	k
*x_1/ʔ_	k	k	k	k	k	k	k	k		

(*continued*)

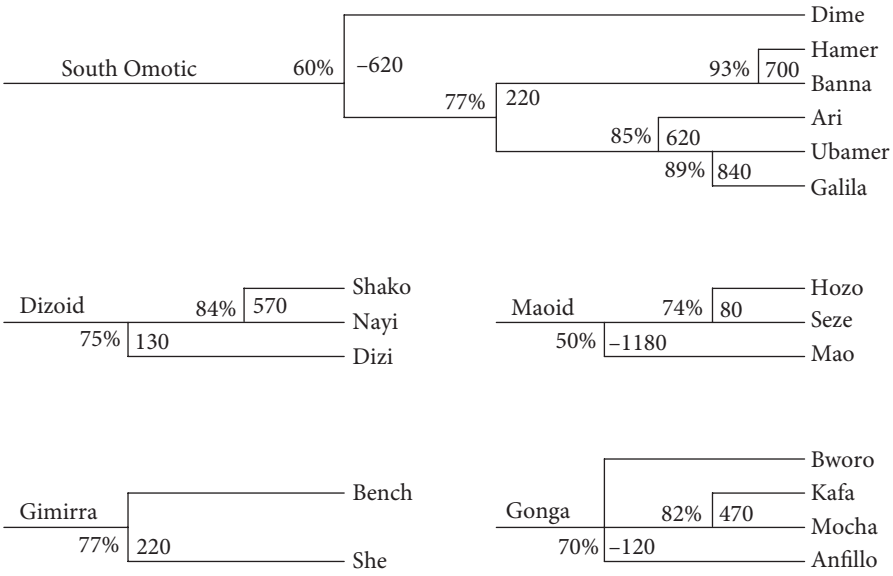
P-Om	Wo	Me	Zy	Ye	Kf	Mc	Be	Dz	Di	Ar
/V_		k				k				ɗ
*x_2/ʔ_	h	h	h	k	k	k	k	k		ɗ
/V_V	0, k/_ʔ	ɗ	0, ɗ/V_ʔ n/_	0	0	0	0, k/	0		ɗ
*kʾ/ʔ_	kʾ	kʾ	kʾ	kʾ	kʾ	q [kʾ]	kʾ	kʾ	ɗʾ	q
ž/ʔ_			ž		š	š	ǯ, ž/_e	č~ts		ž
/V_			ž	s	č	s/_ʔ	ž		ž	ž
*č/V_	č	č	č	?č	č	č	ts	č	ts	ts
*ts/V_	s	s	ts	s	š	š, s/_ʔ	ts	č/i_	š/i_	
*tsts/V_	t	ts	ts							
*$š_3$/ʔ_	s	š	š	š	š	š	c	č(ʔ)		
/V_V			š				č	č		
/V_ʔ	š	š	š	š	č	s	c		č	č
š ($š_1$)	š	š	š	š	š	š	š	š	š	š
*$š_2$/V_	š	š	č	š	š	š, čč	č	č	čč	z
*sʾ/ʔ_		čʾ				čʾ	tsʾ	tsʾ		
/V_		s	s, č		čʾ	čʾ	tsʾ	tsʾ	ts	ts
*čʾ	čʾ	čʾ	čʾ	čʾ	čʾ	čʾ	čʾ, ç̌ʾ/_u	čʾ	čʾ	čʾ
*m	m	m	m	m	m	m	m	m	m	m
*n	n	n	n	n	n	n	n	n	n	n
*ɲ	n	n	n	n	n	n	ɲ			
*ʔ/V_	ʔ	ʔ	(ʔ)	(ʔ)	ʔ	ʔ	0	0	ʔ	ʔ
*h_1/ʔ_	h	h	h	h, 0		h, 0	h, 0	h		0
*h_2/ʔ_	h	w	w	w	w	w	0	h		
*l/ʔ_VC	l	l	l	n	d	d	d	l		l
/ʔ_Vb					n	n				
/ʔ_VN	n	l	n	n			n			
/V_	l	l	l	l	l	l	l	l	l	l
*dʾ/ʔ_	tʾ	dʾ	dʾ	t	tʾ	tʾ	tʾ	tʾ		dʾ
/V_	dʾ	dʾ	dʾ	r	tʾ	tʾ	tʾ			dʾ
*r/V_	r	r	r	r, ll	r	r	r	r	r	r
*w/ʔ_	w	w	w	w	w	w	w	w	w	w
*y	y	y	y	y	y	y	y	y	y	y

Note: The symbols ç̌, ş̌, ǯ, represent the retroflex sounds designated as ç, ş, z̧ respectively by Hayward (see above).

3. Results of the present study

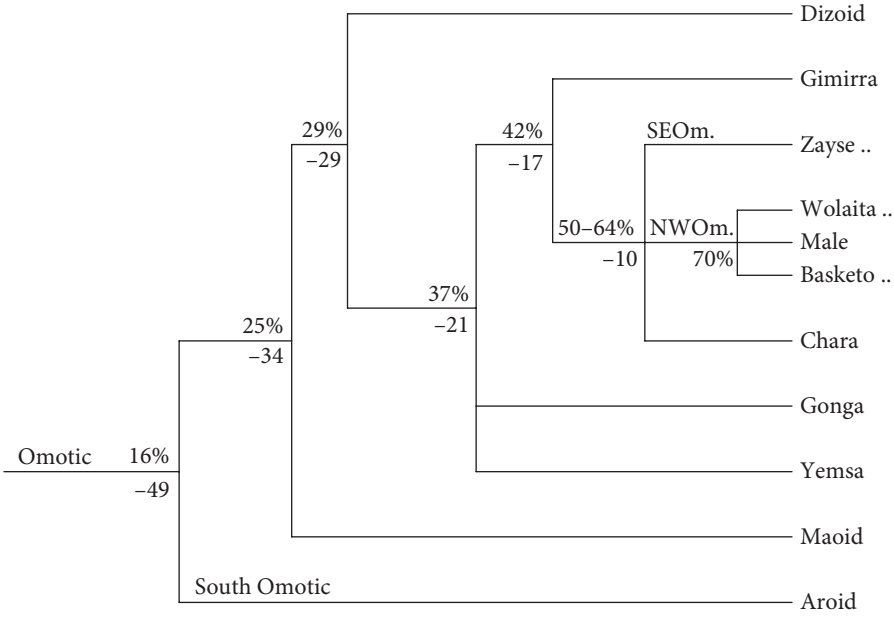
Applying the recalibrated variant of glottochronology formulated by Sergei Starostin, for the partial classifications of the individual Omotic groups the following figures resulted (Table 10).

Table 10. Partial classifications of Omotic subgroups (A)



In the following diagram the absolute chronology is in centuries (Table 11).

Table 11. Partial classifications of Omotic subgroups (B)



4. Basic lexicon of Omotic (Table 12)

Table 12. Standard 100-word-list with some additional items

lg. / gloss	1. all	2. ashes	3. bark	4. belly	5. big	6. bird	7. bite	8. black
Basketo	woytsi	bùdà	k'ura	mɪlla; lippe	bʿɛts	kafa	d'ag	karts
Dokka	wəytsi	buda	k'urra	mɪlle	bɛče	kafi, kapi	d'ak-	kartsi
Male	bija	bidinsi	k'u:ri	gobi	dʿɛ:pi; pu:pi	kaɸi	daij-	karts'i
Wolaita	ubbay	bidin-ta A sukʿunta	fok-uwa; Ce guobba	ul-uwa	gita A wogga	kaf-uwa A kap'-uwa	saʾ-	kareetta
Kullo	ʾubba	bidinsa	p'ok'-uwa	gaawo	dar-uwa	kap'-uwa	saʾ-	keretsa
Cancha	wuuri	bidinssa	p'ok'o	ʾul-uwa	dara	kap'o	sass'-	karets
Malo	ʾubái	wudó	ɸúk'ɔ	ʾúlʌʌtʰɔ	dámmo	káɸo	sats'-	kárts
Gofa	ʾubbay	bud'o	fok'o	ʾulo	gitaa	kap'o	sass'-	karetsa
Zala				daratya	gita	kafo		karetsa
Gamu	L ubbay	bid'insa	p'ok'ets	gawo	bogga	kafo	sas'-	karets
Dache	bʾue	muk'o	fok'o	k'antse	giyta	kafo	sats'-	karets
Dorze	wuurí	bud'o	p'ok'o	gaawo	dara	k'ep'e	sass'-	karets
Oyda	ubba	B bu:dɔ	k'uro	gənč'e	dəmə; gitá	ləndə	d'aʾo	kartis
Zayse	ʾuddá	muk'ɔ́	urrɔ́	gawó	adó	kaɸɔ́	méyc'-	kárts'
Zergulla	gudɛsa	muko	kokopoko	gau	hata	kapo	mɛ:č'i	kats
Ganjule	uddá	muk'ɔ́	p'ɔkɔ́	gawɔ́	arʾó	kaɸɔ́	sa:ts'-	karts'i
Gidicho	ʾuddi	muk'ɔ́	fók'o	gawó	arʾó	kaɸɔ́	sa:s-; áykki!	kársi
Kachama	ʊddí	muk'ɔ́	S poko	ga:ʾɔ́	arʾɔ́	kaɸɔ́	sa:ts'-	ka:rts'i
Koyra	uddi	muk'ɔ́	urɔ́	gawɔ́	arʾɔ́; ʾórce	kaɸɔ́	sa:ts; ɔ́ewa	ka:rts'

(*Continued*)

Table 12. (Continued)

lg. / gloss	1. all	2. ashes	3. bark	4. belly	5. big	6. bird	7. bite	8. black
Chara	úräri; fitta	bénda	gónga; gúrʾa	ʾíːna	méːsa	káfa	sás'-	kärtá
Bench	kang, kep	ts'yakn, ut'n	k'omp'ar	šil	ez; B yaːz	šoyt	sats'	ts'id
She	kaŋk	ben	šan	šɪl; CR šeːn	ik'a	šoit	sats'a	ts'id
Yemsa	zuttera; kup'ira	bedna	goŋa	kata	akama; innya	kap	saaʾra	kara
Bworo	unnooni	tulla	k'ook'ra	maac'a	eena	kafa	šaats'-	aak'a
Anfillo	manno	bɛrɪtso	opp'o	maːts'o	oːgo	kaːfo	saːts'	aːʾo
Kafa	bulla; uba	tullo	goro	maač'o	oge; qabe	kafo	šač'o	aʾo
Mocha	búllo	tullo	óqqo	ma·č'o	ò·go	kaᵽo	šà·čč'i-	a·ʾo
Dizi	šowitel	s'iaᵏn	orgn	čoun-u	yak-ɪs	kɛbi	wʌs'-	s'an-ɪs
Shako	keːtà	s'iak`n̥	örk'n̥	B bo	yats	kə̀fà	wós'	s'ans
Nayi	B bula	s'eak`n̥	B orkn	B yiːnu	ʾeac; B yats-iz	kaf; B kab	wos'	s'ahn
Mao	mūkésɛ̀	p`ʊzɛ̀	k'òk'éšɛ̀	tíːlé; F šin	kémté	káɸɛ́	ʾà ttáːzā	tìːšindɛ̀
Seze$_{(1/2)}$	mʊ´kkè	pɔ́sí	kɔŋkɔ́ší	ʾoːbbi/k'ani	ʾɔ́ggátɛ̀	káwìʾ	saːssaːsa	k'ikk'iːni
Hozo	dimɛ	peːndi	šʊweli	ʊbbi	ɛtzɛ	dorŋa	šaːtɪ	tʊːta
Dime	kull; F wuːf	bindi	ʾoxom	č'ola; F šiːne	gɪčo; F gad	ʾəɸtʰe; F ipt	hatsɪn	zuːb;F tsʾaːno
Hamer	wul	F bidin	wok'umpa	ʾi	gɛbɪ; L gʾɪb	afti	g(a)'-	F ts'ia
Banna	wull	bidini	loš	ʾi	gɛʾbi	afti	gaʾ-	č'alɛ-mi
Karo	wul	bidɨni	uŋglo	ʾíː; M ʾiːna PL. *abdomen*	gɛʾbi	afti	ɨdi M gaʾ-	ts'ia
Ari	muda	bɪndi	loʾota	norti	dɛs'-mi	afti	gaʾ-/gak-	s'ia
Ubamer	mudda	bɨndi	o·fri	nortí	dɛts'-mi	afti	gaʾ-	č'əlɛ-mi
Galila	tunni	bɪndí	o·fri; lahági	lip'a; nok'orti *intestines*	andɨr; baka; B ɗ'ek'a	afti	gaʾ-	č'alɛ-mi

	9. blood	10. bone	11. breast	12. burn	13. claw / fingernail	14. cloud	15. cold	16. come
Basketo	suuts	mɛkˈɛts	ʾəmmì	micc-	sˈuŋgɪts	B ša:ra	B zak	(a)yɛ
Dokka	su·ts	mikˈəts	əmma	mičč-	sˈuŋuts	ša·ra	zá·ki	y-
Male	sugutsi	mɛgɛtsi	dˈansi	mitsˈ-	tsˈuŋgutsi dapo	ša:ri	šimɛ; tsʿi:s	mokɛ:ni
Wolaita	suutta; Ch maccamié	mekˈetta	tʿaanta Ch tera	tʿuug-	tʿuguun-ta	šaara; Ce guma	dooš-uwa A iirta B meegwa Ce biraa	y-Ce hay
Kullo	sutsa	mekˈetsa	dˈaantsa	sˈug-	sʿuguntsa *f.* zaap-iya *c.*	šar-iya	meg-uwa	y-
Cancha	suts	mekˈetsa	tˈansi	sˈugett-	sʿuguns *f.*	guule	irsʿa	ʾukk-
Malo	súts^{h}	mɛkʿɨ́tsʿ	dˈʎns	tsˈugɪnt-	tsʿɨngúts	šá:ra	ʾɪrtsʿ-mé:go	je'-
Gofa	sutsa	mekˈetta	antsa	sˈugett-	dapˈo *c.*	šara	ʾisˈ-	y-
Zala	tsutsa				tseguntsa			y
Gamu	suuts	mekˈets	dˈans	sˈugett-	sˈuguns *f.* worsˈofa *c.*	guule	irsˈa	y-
Dache	suts	mekˈets	dˈa:nts	tsˈug-	tsˈugunts	damɛna	me:go	y-
Dorze	suts	mekˈetsa	tˈansi	sˈugitt-	sˈuguns *f.*	guule	irθa	y-
Oyda	suts	mɛkˈɛts	B dˈants	mič/š-	sˈogəts	dˈone	čʼiš; toya	yi'-
Zayse	súuts'	mɛkˈɛ́ɛte	dˈáants'	ʾɛ́ɛtsˈ-; mɛg	gumɛ́	šaará	mooš; tóyya	y-oo'a!
Zergulla	su:ts	nɛkɛtɛ	dans	dɛ:tsi	tsˈugumi	duluŋgi	ʾɪrtsa; mo:še	j-
Ganjule	súuts'	mɛkˈɛ́tɛ	dˈáants'	ʾɛ́ɛtsˈ-; tsˈub	tsˈuumɛ	dˈumá	tóyya	y-
Gidicho	súutsˈi	mɛkˈɛte	dˈáantsˈi	ʾɛɛtsˈ-	tsugúntsˈi	dumáncʿa dɔgɔ́ncʿa	tóyya	y-oowa

(*Continued*)

Table 12. (Continued)

	9. blood	10. bone	11. breast	12. burn	13. claw / fingernail	14. cloud	15. cold	16. come
Kachama	súuts'ɛ	mɛkʿétee	d'áants'ɪ	ts'úb-	ts'ugúnts'i	ʾuma	tɔ́yya	y-ɔɔa
Koyra	súuts'	mɛkʿɛ́ɛte	d'áants'	ʾɛ́ɛts'-; ts'ug	ts'ugúnts'e	d'umá	tóyya	y-ɔɔwa
Chara	sú:ta	mertá	t'áma/d'áma	míč'-	š'úgna	túgna	čalká; šáča	wóy-
Bench	sut	mert	t'yam	kets'; bunkʿ	ts'ung	duldul	k'og	yeʾ-; anyo!
She	sut	mɛrt	t'iam	saŋka	ts'uŋg	cʿom	k'og	wu
Yemsa	anna	mega	tama	sooka	difna	šaaru	koočo	yo; wu!
Bworo	ts'atts'a	makʾəttsa	gerha	mits'-	ts'unguttsa	dawna	akk'a	waa-
Anfillo	ts'antso	šaušo	t'a:ntso	mɪ:tsʿi	ts'ungɨtso	k'u:ffo	ak'o	wɔ:be
Kafa	dammo	šawušo	t'ano	mič'o; at'io	gač'e- -gommo	qufo; kafo; šibo; iro	aqqo; qewo, qayo	womo; šaggo; beddo
Mocha	damo	ša·wúšo	t'änó; gäto	mìčč'i- ʾà·tt'a-	čʿù̱ŋgətʿó	gu·ppo	qäwo	wa·(hä)
Dizi	yɛrm-u	us	t'iam-u	ka:m-; Y ats-	hial-u	di-u	čo-u	y-
Shako	yärm	u:su	t'em; s'ezga gùb *chest*	bónk'	šišn̰ keri *f.*	gìbu	k'etʾ n̰s; šow	yà̰g
Nayi	yarbm	ʾu:s	t'eam	bonk'us B taso	šišk keli *f.*	di:	šow	ye:
Mao	hándɛ́	má:ltʿɛ́	ʾárɛ́	hà kɛ́nɛ̀gá	wɛ̀nzɪbɛ̀	šɛ̀:lɛ̀	sʿūš/sʿūpʰ	ʾà kʲá
Seze$_{1(/2)}$	hámbìlɛ̀	bàkʿílí	ʾámmɛ̀	bìssɛ́	tsʿúkùmi	šɛ̀:la	tássì	kʷa
Hozo	hambilɛ	bakʿilɛ	ammi	k'as'i	tʊʊmi	šekʿili	tatˢi	kʷa
Dime	maχse; F dzumt	kʿoss F kʿʊs	ʾɛmɛ	F a(:)ts; F a(:)f	guš; F šukum	čʿi:čʿ	B kʿɛž-ɪn F kʿɪž-	ʾoxs BF a(:)d-
Hamer	zumʾi	le:fi	ami	ku:ʾ-; at-	guš; F košum	liile	k'až	niʾ-
Banna	zump'i	lɛfi	ami	at-; ko:ka	gušo; košumʾ	B li:le	k'aji	niʾ-
Karo	mək'əs	lefi	M ʾami		gušo	lubo; d'ummi	k'ajamu	

Ari	zom'i	lɛfi	ami	as'-	guša; H šuqlí	lila	k'aži	ad-
Ubamer	mək'əs ~ -ɣ-	lɛfí	amí	atts-	šuk'(u)ma	lila	ɣaži/'aži	a·d-
Galila	mək'əs	lɛfí	ami	aj-; rrob'e	guša;B gil'a *f*	uppá	k'aží; zá:z	ad-; B kor-

	17. die	**18. dog**	**19. drink**	**20. dry**	**21. ear**	**22. earth**	**23. eat**	**24. egg**
Basketo	hayk'	kana	ušk	B mɛlla	waytsi	Ce bitta	A moy-	B muk'a
Dokka	ayk'-	kana	ušk-	mele	waytsi	bi·tti; sa·ze	m-	muk'a
Male	haitɛni	kani	uš-; makoja	mɛlɛzi	waizi	sar	mo-	bulla
Wolaita	haykk'-	kana	uyy-	mela	haytta	biitta	m-	k'uk'ull-iya A p'up'uliya
Kullo	hayk'-edda	kana	'us'-	mela	haytsa	sa'a	m-	k'uk'uliya
Cancha	hay'-ires	kana	'uy-	melala	hayts	biitta	m-	p'up'le
Malo	hajk'éza	kána	wuʒ-	méla	hʌ́je	bí:ta	m-	b'up'ʌle
Gofa	hayk-is	kana	'uy-	mela	haytsa	biitta	m-	p'up'le
Zala	ayq	kana	uy		(h)aytsa	bitta; gade	maa-	p'ap'ulle
Gamu	hayk'-ires	kana	'uy-	mela	haytsa	sa'a	m-	p'up'ule
Dache	hayk'-	kana	(u)w-	mɛla	hayts'e	L biitta	m-	b'ub'ule
Dorze	hay'-ires	kana	'uy-	melala	waye	biitta	m-	p'up'ule
Oyda	hayk'-	kana	až-	mɛla	B ha:ye	gədde; sa'a	mu'-	b'ub'ule
Zayse	hay'-	káná	úšš-	mɛ́lla	waayɛ́	S 'ala	m-	L b'ub'ulle
Zergulla	hai'i	kana	'ušša	mɛla	wai	gade	m-	bubɪllɛ
Ganjule	hay'-	káná	úšš-	mɛ́lla	waašɛ́	S sɛxa:	m-	b'ub'úlɛ
Gidicho	hay'-	kaná	úšš-	mɛ́la	waašɛ́		m-	b'ub'uulɛ́
Kachama	hay'-	kaná	úšš-	mɛ́lla	uwaašɛ́	S 'alla	m-	b'ub'úlɛ
Koyra	hay'-	káná	úšš-	mɛ́lla	waayɛ́	S saha	múuwa	b'ub'úlɛ

(*Continued*)

Table 12. (Continued)

	17. die	18. dog	19. drink	20. dry	21. ear	22. earth	23. eat	24. egg
Chara	háy-	kána	úš-ša	kó:la	wó:ya	dúkʻa	ḿ-na	mó:la
Bench	hayb(kʼ)	kyan	uš̩(k)	kol	(h)ay	buyt'; dod	mʾ	mul
She	aikʻa	kian	uš	kol	ai	CR dod	mma	mul
Yemsa	kitu	kana	uša	išma, ičma	odo	da(a)	me	kewa
Bworo	kʻirr-	kana	uš-	šuukʻa	waaza	gaawa	maa-	keettsa
Anfillo	kʼe:ra	kano	ʊšš	kə:tso	wa:jo	G desʼo	mə	ke:nno
Kafa	qito; kašo; aqqo; čʼimo	kunano/e	uyo; tʼamo	šuʼe, šoke	wammo; kendo	showo; mogo	mammo; čʻokko	gabetʼo Ce kʼabitʼo
Mocha	qìti-	kũna·no	ʾussi-	šù·qqi- v.	wa·mmo	šawo	ma̲·(hä)	gabatʻo
Dizi	šub̲-	kian-u	bɛ-	kal-ɪs	a:i	gob; Y tu:r	m-	mialg-u
Shako	šúb	keanu	kʼi̲š	B kɔrɪnts	a:y	tú:rú	`m̥	meaku
Nayi	šubm	ke:anu, kyonu	B wob̲kesk	B kola	B ha:y	tu:r	m-	moala
Mao	ʾà hékʻà	kànɛ́	hà ʾíšá	káŋgílt'ɛ̀	wá:lɛ́	kʼɛ́:tsʼɛ́	hà míjà	kʲɛ́lɛ̀
Seze$_{(1/2)}$	zízzítɛ̀	wiš[š]i	ʾíšmć	kjá:mì	wɛ́ɛ̀	kʼɛsi	má:mɔ́	wá:kjà
Hozo	zizi	kana	išɪ	kikyamtɛ	wɛɛra	S kɛtsi	maa	-kɛŋgi
Dime	BF la:ʾdʼ-	kɛnɛ	ʾučʻu	wučum	kʼa:me	hilə; B yi:l	ʾičɨn	mɔlu
Hamer	di-	kʻɑski	wəcʼ-	woc-	kʼa(:)m-	pe	kʊm-	bʊla
Banna	di-	kʻaski	wʌčʼ-	wʌčča	kʼami	pe	its-; kum-	mukʼa; bʼʊla
Karo		kaski	naʾap; M wucʼ-	M wɔcʼɪ	kʼami	F pe; tore	isidi	mukʼo
Ari	dɛʾ-	aksi/aksɛn	wʌčʼ-	wʌči	kʼami	pʻɛčʼa	its-	muɣa
Ubamer	dɛʾ-	aksi	wəčʼ-	wʊ·ča	ɣ/kʼa·mi	fičʼa	ʾits-	mukʼa ~ -ɣ-
Galila	dɛʾ-	akši; kani	wɛčʼ-	wʊča	kʼa·mi	pʻičʼa	ič-	mu·ka ~ -ɣ-

	25. eye	26. fat	27. feather	28. fire	29. fish	30. fly n.	31. foot	32. full
Basketo	af	B mɔd'a		B tama	mola	wusʿɪla	to^{h}e	
Dokka	af	šəlga		təma	mola	wusʿɪla	toh	kúmsi
Male	ʾa:pi	ma:li; mo:rɛsi	gaɸani	tami	molo	wus'ali	toki	kumotsi
Wolaita	ayf-iya A ayp'-iya	modd'-uwa	baall-iya	tama	mol-iya	udunt'-iya	toh-uwa Bk gáddea	kum-/mitt- A kumetta
Kullo	ayp'-iya	modd'-uwa	kʿepeta	tama	mol-iya	guduns'-iya	ged'-iya	kumetsa
Cancha	ayp'e	modd'o	balle	tama	mole	guduns'e	toho	kumets
Malo	ʾáɸe	mɔ́t'e	šugúlʌ; kʿɛφɛ	tamá	mɔ́lɔ	wutsɨ́tsɔ	t^{h}ɔ́hɔ	kʊltsi
Gofa	ayp'e	modd'o	balla	tama	molo	ʾuduns'a	toho	kumets
Zala	ayfe			tama	muoliya		toho	
Gamu	ayp'e	modd'o	balle	tama	mole	ʾuduns'e	toho	k'umes
Dache	ayfe	mɔ́d'o		təma	mole	wəzɪntse	to^{o}	
Dorze	ayp'e	modd'o	balle	tama	mole	guduns'e	toho	kumets
Oyda	ápe, ayfe	mod'o	kʿɛfe	təmmə	molo	wus'us'o	tohe	ubbə = 1
Zayse	ʾáaɸɛ	handé; mɔd'ɔ v.	S sɛ:lo L paange	buddɔ́	molɛ́	wuts'ínts'i	túki	ʾakukʊŋgis
Zergulla	ʾa:ɸe	modebe	kʿɛfɛ; goena	budo	mole	wuzuntse	tuke	komuts
Ganjule	ʾáaɸɛ	handé	S ba:lo	tamá	molo	wuts'ints'e	túke	S kʊmə
Gidicho	ʾáaɸɛ	handé		tamá	moló	wus'únts'e	túke	
Kachama	ʾáaɸɛ	handé; mɔ́ɔlla v.	S ba:l̥i	tamá	moló	wus'únts'e	túke	pumi
Koyra	ʾáɸɛ	handé	ɸaŋgɛ	taama	molɛ́	wus'ints'e	tokɛ	kumɛ
Chara	á:pa	kúwa	kúsna	táma	molá	úsʿla	tá:ya	s'os'á

(*Continued*)

Table 12. (Continued)

	25. eye	26. fat	27. feather	28. fire	29. fish	30. fly n.	31. foot	32. full
Bench	ap	ko’	kakn *wing*	tam	oru	bur; kaskas	to’	ç'onç'- *fill*
She	af	ko’		tam	woři	wuts'uts'	to’	
Yemsa	aafa; kema	faru	koso	geya	ts'op'a	kura	orgo; tibo; zooma; očo	tuuma H tuun-u
Bworo	aawa	kʿoottsa	banga	tawa	mutts'a	zaanza	tufa	ts'een-
Anfillo	a:fo	še:o		tamo	ha:šo	ya:’unzo	tʿa:mmo	
Kafa	affo, aho	auyo; qočo	wale; čʿero	qaqo; šelo	yešo; ollo	yammo; yobo	batto; deč	č'ene
Mocha	á·po	’awwó	g/qošo	qa·qqo	ha·šo	yäwuǧǧo	gano	č'e·nno
Dizi	ab-u	kobab; Y kowu	bi	al-u; Y tamu	urgit	B wɛngi	ašš-u	s'oz Y s'os'
Shako	á:b	kòw	bì’í	ta:mu	o:rù	’uts; B wʊtsʿi	á:šú; ša:nu	s'ós'á
Nayi	’a:f	kow		toamu	B aša	’ots; B utsi	a:šu	s'os'a
Mao	á:fɛ́	málɛ́	kʷá:p kʷi:ntʿɛ̀	kámɛ̀	kʿókʿɛ̀	tsʿíŋɛ̀	čɛ́:mɛ́	ts'ó:nɛ́nɛ̀
Seze$_{(1/2)}$	a:b, á:wi	kʿɔ́ssi	’ùnsʊ́lɛ̀ / báŋkì	tàmmì	was'ɛ	šiaŋi	dàkká / tʊgg[Ý]	ts'ó:nì / ts'unewi
Hozo	abbi	kʿotˢa	baŋka; S wá:ntoli	taamɛ	was'i	šɛŋgi	kɛšɪ	S sʿa:nnì
Dime	’afe, ’aɸe	kʰuštʰu	kʰɛlɛ; F kam	nu:nu	’ɔrɔxo	kʰobzo	dotʰu	F s'u:s'-; k'ox
Hamer	api, afi	mɔr; F durfi		nu	kara	kʊsuubo	roti	F ts'ots'
Banna	afi	durfi	siti; sɪle	nu’	toila; ka:ra	k'us'ʊbu	roti/rʊ:/rro	s'os'i
Karo	afi	durfi	F suti	no = M nɔɔ			rro	
Ari	afi	durfi; B lada	sil’a	noa	toyla	H gols'a; ka:-b'i; B majak	duti	s'os'i; H míš-
Ubamer	a·fi	durfi	silla	noha	molta	suŋgula	du·ti	s'os'i
Galila	a·fi	durfi	ke(:)fi	no’a/noha; B lokʿa	mola	šuŋgula	du·ti	č'o·č'i

	33. give	34. good	35. green	36. hair	37. hand	38. head	39. hear	40. heart
Basketo	B ʊmə	kɔše		k'omme 38	B kʊš	B k'ɔmma	sisi	buda
Dokka	umm-	košše	č'alla	isins	kušu	k'$^{(w)}$om	sis-	bu·da
Male	ʾɪgɛni	kɔši	ts'ɔl̓e	to:ki gami	kuči	to:ki	waiz-	saza; ʾina
Wolaita	imm-	loʾ-uwa	maata-mala maata *grass*	binnana	kuššiya	hup'-iya Bk kómmo	siy-	wozana
Kullo	imm-	loʾ-uwa	mata-gidiya	huup'-iya 38	kuš-iya	hup'-iya	sis-	wozana
Cancha	imm-	loʾo	mata-malets	iske	kuše	huup'e	s-	wozana
Malo	ʾɪŋg-	lɔ̃ʾɔ	čililó	binána; č'ímba	k^{h}úše	k'ómmo	sizóbʌ	wɔzʌ́nʌ
Gofa	imm-	loʾo	mala	binnana	kuše	huup'e	siʾ-is	wazana
Zala	im	loʾa		midd'insa	kuše	huʾya		mutsuro
Gamu	imm-	loʾo	c'ililo	iske	kuše	huuʾe	s-	wozana
Dache	im-	loʾo		biynɛna	kuše	hue	s-	wəziyna
Dorze	imm-	loʾo	c'ililo	iske	kuše	huup'e	šuy-	wozana
Oyda	ɪŋg-	s'ɛgo	č'ililo	č'imbə	kúšé	B o:mma	saʾ-; B si-	B wɔzɛna
Zayse	ʾiŋg-	lɔ̃ʾɔ	H žillá	its'ínkɛ	kúc'ɛ	ʾúumma	s-iiya	mutsʿirɔ; wozani
Zergulla	dɪŋgi	ləo	ʒila	ʾumma 38	ku:č'e	ʾumma	s-ija	wazane
Ganjule	ʾiŋg-	lɔ̃ʾɔ	S ʒilla	bínaana	kuc'e	ʾóomma	s-iiya	muts'urɔ
Gidicho	ʾiŋg-	lɔ́ɔkɔ		bináana itúnts'i	kúuše, kúc'ɛ	ʾóomma	s-íida	muzuróu
Kachama	ʾiŋg-	lɔ̃ʾɔ	S ʒela	bináana itúnts'i	kúuc'i	ʾúumma CR qino	S s-i:tanini	muzɔrɔ́
Koyra	ʾiŋg-	mɔ́ɔdɛ	S ʒilɛta	uɸutsɛ binana	kuc'ɛ	k'ínne	s-iiwa	múts'ɔró

(*Continued*)

Table 12. (Continued)

	33. give	34. good	35. green	36. hair	37. hand	38. head	39. hear	40. heart
Chara	ʾín-	ónsa	č‘íla	kúsnaʾ	kučá	to:yá	hayt'na	ma:ya 48
Bench	uts	digam; B s'e	Ra č'il	toh; B kʊ:st	kuč	deb; Ce qe:l	sis(k)	muzar;B may
She	uts	ts'ea		kust	kuc	to	sis	fikn
Yemsa	imi	maʾar	čiro	soma	kušu	teeta	weeso; odo	niba
Bworo	im-	doʾa; še:ŋga	moc'o-k'oyi	ts'iira	kiša	toka	šiš-	nibba
Anfillo	ɨmm	bərgɛts	mɔ:ts'ɔ	tɪgi:ro	ke:šo	to:kko	ši:ʸe	yɪbbo
Kafa	immo	gowo	gare	qello; et'o	kišo	gimbo; ogo; Ce qello:	wayo; ello	mullo
Mocha	immi-	dóʾo	qà·ğği- v.	toʾo; č'e·ro t‘igaro	kišó	qällo	qäbbi-	níbbo
Dizi	ta(s)-	jež; šɪšk-ɪs	č'ɪl-ɪs	sar; Y tsicu	kʊ̀č̲-u	geli	siss-	lɪb; B čwonu = 4
Shako	á:c	žens	č'ir`n̥s	sicu	kʊ̀č̲ú	gärí; F mo:tu	si:s	šon
Nayi	a:ts	šišk'ns	c'il-u	tsicu	B kucu(s)	mot	sisne	B šunus
Mao	ʾá tá	nɔ́kítɛ̀	tæ´:kʊmtɛ̀	tɔ́:k wínt'ɛ́ G untili; tis'i	kúsɛ́	tɔ́:kɛ̀	k‘ɛ́wá	ɛ̀:ŋɛ̀
Seze$_{(1/2)}$	né:ntì/ tí:tìjà	màtt'í	tʷá:kàm ʾɛ́:wɪ	tú ónsìl	kúsi	túì	šeše/ k'ʷa:k'ʷa	níbbà /šínt'ù B šiinɪ
Hozo	neeni	matt'i	S jɔ́:swe	-ʊntili	kutˢi	tokɪ	šiši	níbba
Dime	ʾɨmzɨn	ʾɪŋgu	F c'ɪlo;šɪmaji	bande	ʾane	mətʰe	k'amse	budə
Hamer	ɪm-	paya	ʾɛk'	sʊ̀tsi	a:ni	F mata	k'ans-; H sɪ:k	wɔ:ylɪm
Banna	im-	fayya		siti	ani	mata	k'ans-	sada
Karo	inkara			sáts'i	ani	mɛti		
Ari	im-	laɣa-mi	č‘alɛ-mi = 8	sɪs'i	a:ni	mata	ɛsɛr-; H wúr-	saza
Ubamer	ɪm-	laɣ-mi	posa	sɪs'i	a·ni	mɛtá	ɛsɛr-	lɪp'a/lib'a
Galila	um-	urma	č‘anəč‘al	šič'i	a·ni	mətá	ɛsɛr-	bu·da

	41. horn	42. I	43. kill	44. knee	45. know	46. leaf	47. lie	48. liver
Basketo	B ʊšɪnš	tana	B wəda-	buk'	ɛra-	šitta		mayz
Dokka	ušɪ´nč	tana	wəd'-	buk'ɪ	er-	mits waytse *ear of tree*		mayz
Male	gurani	ta:ni	wod'ɛni	go:ši	'ɛrɛni	'ilaši	lahɛni	maizi
Wolaita	kac'iya	ta(ni/na)	war-/wor-	gulbata	er-	haytta; Ce bončwa	zi'n-	tir-iya
Kullo	kac'-iya	tana	wokk'-edda	gulbata	L er-/ir-	haytsa	bokk-edda	tir-iya
Cancha	kac'e	tani	wor-ires	suule		hayts	zinn'-ires	tire
Malo	k^{h}áč'e	táni	wod'é:za	búŋk'e	'ɛréza	jéčɛ	gá:ŋgəteza	t^{h}íre; kilahó
Gofa	kacc'e	tani	wod'-is	gulbate	L er-	haytsa	zenp'ott-is	tire
Zala		ta		iquorsa	er			
Gamu	kac'e	tani	war-ires	gulbata	L er-	banʒo	zempet-ires	tire
Dache	kac'e	tena	wə'-		ɛr-	wa:ye		tire
Dorze	kac'e	tani	wor-ires	suule		waye	zinn'-ires	tire
Oyda	k'ač'e	tana	wod'-	bunk'e	er-	haye		šɪnfilɪ
Zayse	uššúmɛ	tána S tai	wɔd'-	bɔ'inte	'ɛrr-	maatá waayɛ́ *greens ear*	S zabadatɪs	maayyé; L tire
Zergulla	kač'e	tana	wod'i	gulbɛta	'ɛri	wa:je	wei'i	ma:e
Ganjule	bak'ɛ	tána	wɔd'-	bɔ'inte	'ɛrr-	waašɛ́	S gea-	maayyé
Gidicho	uššúume	tána	wɔd'-	bɔ'unte	'ɛrr-	báale		maayyé
Kachama	uššúume	tána	wɔd'-	bɔ'unte	'ɛrr-	waašɛ́	S dɛ:ta-	maayyé
Koyra	uššúumɛ baka	tamba	wɔd'-	bɔ́hɛ	'ɛrr-	waayɛ́	S 'asalɛ	maayyé; Ce kille
Chara	úša	táni	wót'-	k'uma	ár-	wúya	gétik-	ma:yá
Bench	pel	ta(-n)	wot'	k'um	er	koč; kar	Ra šāt	may; B žə:š

(*Continued*)

Table 12. (Continued)

	41. horn	42. I	43. kill	44. knee	45. know	46. leaf	47. lie	48. liver
She	ʊš; F fäl	B ta-na	wota	guraət	eřa	aisi; F kari		mai
Yemsa	kala	ta	woru; keǰa	keʾa	aru	toša	kuna	zogara
Bworo	k'alla	ta	uud-	gubra	daan-	maarha	k'ey-	afara
Anfillo	k'a:ro	to-ʾoni	wu:r'a	gulma:tto	ɛrri	ma:so		afa:ro
Kafa	qaro; tiro	ta(-ne)	wut'o; čʿičo	gutino	arriyo	wollo	koto; nokko	qammo
Mocha	qáro	B ta	wùt'i-	gurato	arì·(hä)	wa·mmo 21		ap̱á·ro
Dizi	ušum	B inu	dɛb-uš-	kol-u	tʿus-	tarš; B a:i	sog-	bo; Y bow
Shako	fera; ušń	nȁta	wú:š	k'umu bòka *thigh*	tʿu:s	kȁrá		B bo aško
Nayi	ʾuša	B na	wuš	B gumtu	tʿus	ai		bow
Mao	ʾi:šìmɛ̀	tíjá	ʾà fíjà	túlk'ùmɛ̀	hà ʾàldá	ʾí:nts wà:lɛ̀	ʾà tɛ́:ŋā	kʿɛ́ndɛ́là:fɛ̀
Seze$_{(1/2)}$	k'álì; šákali	hà:šɛ́	pʲɛ́ŋpʲɛ̀ŋá	gùšúmɛ̀/ gùbbì	ʾálá ʾàlá	wɛ̀ʲ	hɛ́:mɔ̀	bèì
Hozo	k'ara	naŋga S dɛŋgà	pɛŋgi S hɛdzídí	kʿʊʊmÝ	šoi S arɛti	yaats'i S já:zwe	S hɛ:tì	bɛyɛy S mɛ́ʾi
Dime	ʾʊšʊm; F baltu	ʾatə	BF de:s-	boáχ F bux, boh	kʿɛsɪn; F tʿɛs; B dʿɛ:s	k'ame; F ɛrgu	gɨχngɨχntɪn	tʰaχte
Hamer	k'ʊšʊm-	inta	däs-	buk-	dʿes; F deʾs	F k'ulʾa; aca	k'ʌns-	tʿɔrɔbʿɔ
Banna	k'ušu(m)ba	inta	des-	b'uk'o	dʿes-	k'alʾa	rat- = 77	torugo
Karo	k'ušuba	inta	M dɛs-			M k'alb'o		
Ari	k'ošma; H šóqa	ʾíta	des-	bora; H gulbé	ɛs-	k'alʾ-a, -be		turi H tirá
Ubamer	γ/k'ošma	ita	de·s-	buk'á/buγá	ɛs-			turi
Galila	ba˙li	itá	dɛy-	buγa	əs-	k'alb'a; haca; kakši		túri

	49. long	50. louse	51. man	52. many	53. meat	54. moon	55. mountain	56. mouth
Basketo	A p'arents	čʿu·čʿi	= 65	B wɔylɪnts	B a:yš	B a:gən	B ánda	no·na
Dokka	b'arɪntse	čʿu·či	atinaʾe naʾe child	woylɪntse	ayššе	agane	bučʿa	no·na
Male	ʾodossi	ts'əguts	ʾasi = 65	mɛrge	ʾaški	ʾagɛni	dʿuka	daŋka
Wolaita	adussa	c'ucca A c'u33a	attuma	A dar-uwa L c'ora	aš-uwa	agena	der-iya	doona
Kullo	aduussa	cʿu33a	= 65	dar-uwa	aš-uwa	agena	faga	doona
Cancha	giira	cʿu33	= 65	daro	ašo	agina	zuma	doona
Malo	ʾadɪstsé:	čʿúč'	ʾas = 65	dáro	ʾášo	ʾagé:na	deré	dɔ́nʌ
Gofa	adusse	cʿu33a	= 65	daro	ašo	agena	dere	doona
Zala			uura		ašo	agina	darya	duna
Gamu	gita	c'u33	= 65	dar-uwa	ašo	agino	zuma	doona
Dache	adusé	c'uce	= 65	dero	ašo	agɛn[a]	a:fo	duna
Dorze	giira	c'u33	= 65	daro	ašo	agina	zuma	duuna
Oyda	p'argo	B c'u:c	d'irɛts	ga·ma	aššo	aginə	dere	B do:na
Zayse	zambála	c'úuc'e	ats'; Ce ade	laagɔ́	acʿɔ́	agúuna	bɔc'ɛ́	baadɛ́
Zergulla	di:tso	č'u:č'e	ats	mač'o	ʾačo	ʾagunna	zuma	ha:ʾe
Ganjule	gallála	c'úuc'i	atsi	laagɔ́-kɔ	acʿɔ́	agúuna	bɔc'ɛ́	baadɛ́
Gidicho	galláalla	cúuc'i	náʾi	laagɔ́-kɔ	acʿɔ́/aaššɔ́	aagúuna	bɔc'ɛ́; zumá	baadɛ́
Kachama	galláala	c'úuc'i	náʾi	laagɔ́	aaššɔ́	agúuna	bɔc'ɛ́	baadɛ́
Koyra	gallala	c'úuc'e	ats'e	laagɔ́-kɔ	acʿɔ́	aaguuna	baka	ʾáaša
Chara	bʿäd$^{()}$á	čʿučʿá	átsa	bordá	áča	dóza	mú:ta	no:ná
Bench	p'ad	ç̌,ʿuç̌ʿ	ʾyan; B a:ts	p'uč'	aç̌	ʾyarp	t'ul; gag	non
She	p'ad	c'uc'	ats	k'eməŋuš	a:šša	erf	gud	non

(*Continued*)

Table 12. (Continued)

	49. long	50. louse	51. man	52. many	53. meat	54. moon	55. mountain	56. mouth
Yemsa	diča	tu'a	atk'asu	šoyo/šowo	aša	assa	aama; kuma	noono
Bworo	geenza	ts'uts'a	ašša = 65	aayya	meetsa	aašittsa	gura	noona
Anfillo	gɛnzo	ts'ʊts'o	ašo = 65	ɔ:rrɔ	k'ɪŋgo	wəro	da:go	nɔ:no
Kafa	gendjo	č'uč'o	anamo	meeti; woddo	meno; mayo	agene	oge-geppo	nono; koko
Mocha	gänğa(yé)	č'ʊ̀t'o	dò·no; wuró	me̲·tto	qóŋgo	'agane	gäppo	no·no
Dizi	šedn-ɪs	č'uzž-u	ñakus	yak-ɪs	ačk-u	atsɪm	t'umo	ɛd-u
Shako	šə̀d`n̥s	c̲̀'ʊc̲'n	Y neaku	ángá; p'úč'á	ašku	B a^{y}kn	ge:ra	e:d
Nayi	šad	c'uc'u	B iab = 65	dofa	ašku	ats	ge:ra; B mai	e:du
Mao	k^{w}ætsʲítɛ̀	k'íšɛ̀	'é:sɛ̀	g^{j}æa:zæ	'ɔ́škɛ̀	'á:nsɛ́	šó:wɛ̄lɛ̄	pɔ́:nsɛ̀
Seze$_{(1/2)}$	ša:mi	bìzílɛ̀	mɔ: / maj	nàmmé/ gà:néwì	'ɔ́sí	'ɛ́msí	nɔ̀ŋkɛ̀/ g`ʊlì	wa:ndè
Hozo	-tistɪ	konni	S mɔ:	amsa; S ts'əmo	ot^{s}i S ɔtsi	ɛɛmsi	gura	waandi
Dime	gudum	garse; F k'as	'ʌje; F goštu	gɪčo = 5; F c'utt-;B dɪldo	wʊhʊ	'ɪrɸɛ	mɛ:zɛ; F ɛ:d	'afe B 'app-
Hamer	gʊdʊb	k'asa	aŋg	= 5	wa	arf/p-	d'uk'a; gɛ:mɛr	ap-
Banna	arroti-mi; orma	k'asa	angi	s'os'i 32	wa	arpi/arfi	d'uka	afa
Karo	gutup		aŋg		wa(h)a			M 'apo
Ari	ɪrotu-mi	k'asa	ang	bɛdɪ-mi	wa:	arfi	gɛč'a; B gemar; balá	afa
Ubamer	gažmi	ɣasa	aŋina	bɛdi	wahá	arfi	bəlá; gɛča	afa
Galila	k'əlmi, g'al-	k'asa	aŋina, aŋ	bɛdi	wə'a	árfəná	muká; geča	afa

	57. name	58. neck	59. new	60. night	61. nose	62. not	63. old	64. one
Basketo	B sumsa	k(ˌ)ootsa	B ɔrɛs	B k'amma	sints	A base; -kkaye		pɪtt'ɛn; B fɛtɛn
Dokka	suntsa	k'o·ts	oratse	k'əmme	si·nts			pɛttɛni
Male	sunsi	bakʾana	ʾaki	wa:nte	sieɗʿi	pɛtɛtaʾo; ʾɛijʾɛ *no*	ts'imi	ʾipɛtɛ
Wolaita	sunta	k'oor-iya	(h)oraatta	k'amma A 'omarsa	sir-iya	gid-eenna (gid- = *be*)	c'eegg- v.	is-ta / issi
Kullo	sutta	k'ood'-iya	ʾoratsa	k'amma	siid'-iya	gidenna	ʾec'a	itta
Cancha	sunts	k'oore	ʾolatsa	k'amma	sire	akkay	gal'a	issino
Malo	sʊns	k'ɔtʿe; mɔrgé	ʾorótsi	ʾomátsə	síd'e	bá:ja	ts'íma	$p^hɛt^hɔ$
Gofa	suntsa	k'oode	ʾorats	ʾumarse	siide	gid-enna	gal'a	issino
Zala		k'od'iya			sid'e		čʿima	ista
Gamu	sunts	k'oore		ʾomarse	siire	akkadi	gal'a	issino
Dache	sunts	k'od'e	orɛts	omars	siyd'e			isiyno
Dorze	sunts	k'oore	ʾolatsa	k'amma	sire	akkay	galʾa	issino
Oyda	suntsu	ganna	orrats	k'ama	sid'e			fɛto
Zayse	čʿúučʿe	wɔ́ɔllɔ; k'ɛnɛ́	hagɛ́lɔ	k'áma H čʿemmo	kuŋké	S ʾiihe	čʿima	biizɔ́
Zergulla	su:ns	wollo	ʾorats	čʿe:mo	kuŋki	hijɛ	čʿimma	bizo
Ganjule	ts'únts'i	ʾɔ́ɔnt'a; gɔ́nɛ	kɪllə	k'áma	kuŋkɛ	S ʾai	čʿuma	biizɔ́
Gidicho	ts'únts'i	ʾɔ́ɔnt'a; gána	kíilɛ	k'áma	kuŋké			biizɔ́
Kachama	ts'únts'i	ʾɔ́ɔnt'a; gɔnɛ́	kíilɛ	k'áma	kuŋkɛ́	S wa:ŋkawa	čʿima	biizɔ́

(*Continued*)

Table 12. (Continued)

	57. name	58. neck	59. new	60. night	61. nose	62. not	63. old	64. one
Koyra	súuntsi	baala	kíilɛ	k'áma H č'emmo	siid'ɛ	S ʾɔ'ɔ	čıma	biidzɔ́; Ce wolakko *one of two*
Chara	sumá	borká	ká:ya	únta	sínt'u	káyay	gäná	issá
Bench	sum	bar; ul	k'ay	urd; t'uman *at night*	sint'	Ra a̰ʾa̰a̰ *no* Ra -árg *not*	erg; ba	mat'
She	sum	ui	k'ay	dutan	sint'			mat'
Yemsa	suna	suuta	gaddo		siya	ánè	kuʾru	isa
Bworo	šuutsa	k'eetta	andra	t'uwa	šint'a		dora	ikka
Anfillo	ši:go	k'ɛ:tto	andi:ro	di:hro	šɪ:nto			iko
Kafa	šiggo	qetto	andire	t'umo	muddo	tonone	geno; wonne	ikko
Mocha	šəgo	qätte mit'ó m. = *tree*	andiro	t'úmo	šit'ó		gè·na- vb.	ʾikko
Dizi	sɪm-u	kum; morkn	k'al-ɪs	gwot	sin-u		Y ʾyag-us	k'oy
Shako	su:m	ku:m; B fɔ:ri	k'ärˋn̥s	go:tà	B sınt'		íg ˋn̥s	k'òy
Nayi	su:m	B feli	B k'anu	B g^{w}ata	si.n			B isn
Mao	jè:škɛ́	kílíšɛ̀	tùmɛ́	kɛ́mbílàwɛ̀	ší:nt'ɛ́	né:pájà	k'ò:mɛ́	íškɛ̀
Seze$_{(1/2)}$	nì:ší	ké(té)šé	ʾɔ́:lɛ́	k'ɛ́sɛ́ mà:wì /galgali	ší:nté	kɛ́:wɛ̀	k'wíns'í / mò:fɛ́	ɔ́nná / ìšílɛ̀
Hozo	iiši	kiiši	hɛlli	šibbɪ	šini	hʊŋkí	S k'omašè:	onna
Dime	mɨze; F na:b	F ɛɣɯts; ge:te	wolgo	ʾɑχsti; F sutu B gɛlt-	nʊkʊ	ʾe:kai *no*	ʾatsi; F gɛš	wɔk^{h}ʌl = BF wokkil
Hamer	nam-	qɔ:ca	(y)alli	F suti	nuki	F -ma	F gec(c)o	kəlla
Banna	na(a)bi	k'orči;lukum- izɛk'i	(y)ali alli	soyti	nuki nuki	-itini; -bodi	F gec(c)o	kala kala

Ari	nami	bari; B d'aŋg	killa	soiti	nuki	-ai	B galta	wʌllaya
Ubamer	na·mi	ba·ri	kılá	soyti	nuki			wəlləkˈə
Galila	la·mi	ba·ri; murt	kíllá	soytsi; geltí	nuki		gɛš	wələkˈə

	65. person	**66. rain**	**67. red**	**68. road**	**69. root**	**70. round**	**71. sand**	**72. say**
Basketo	B a:se	B ɨr	A zokats	goytsi	sˈabˈa		B mašınt	B ga:-
Dokka	asé	ire	zukˈatse	góytsi	sˈəbˈa	mu·mme	kˈačˈe	nɛk-; iyya
Male		ʾırzi	zokˈɛ	giitsi	sʾabi		ma:šala	gɛʾɛni
Wolaita	asa	ira	zoʾ-uwa	og-iya A ged-iya	tˈapˈ-uwa Ce dulliyaa	yuuš-uwa	šaf-iya A šapˈ-iya	g-
Kullo	asa	ʾira	zoʾ-uwa	ʾog-iya; ged-iya	sˈap-uwa		šapˈ-iya	g-edda
Cancha	asa	ʾira	zoʾo	ʾoge	sˈapˈo		ancˈo	gi-ires
Malo	ʾás	ʾíra	sóʾo, bazóko	ʾogɛ́	sˈɛ́pˈo		šafé	gé:zʌ
Gofa	asa	Ce ira	zoʾo	ʾoge	sˈapˈo		šapˈo	gi-is
Zala	asa	ira	zoʾo	ogiya				g-
Gamu	asa	ʾira	zoʾo	ʾoge	sˈapˈo		ancˈo	gi-ires
Dache	as	iyra	zoʾo	ogé	sˈɛbˈo		sili:mo	hızg-
Dorze	asa	ʾira	zoʾo	ʾoge	sˈapˈo		ancˈo	gay-ires
Oyda	a·si	e·ra; buk- v.	zokˈo	ó·ge	sˈabˈo	gımbe	ša·fe	yag-
Zayse	gɛ́r	ʾírra	zɔʾɔ́	oγɛ́:; S goge	tsˈabˈɔ́		šɛɛcˈɛ́	yi-
Zergulla		ʾira	zo:	hɔke	sˈabo		silemo	j-esi
Ganjule	dɛrɛ	ʾírra	zɔʾɔ́	ogɛ	tsˈabɔ́		šɛɛcˈɛ́	híi-
Gidicho	ʾáatsˈi, aassí	ʾírra	zɔʾɔ́	obɛ́/oγɛ́	tsˈabˈɔ́		šɛɛcʾɛ́	hii-
Kachama	gɛ́ɛrɛde	ʾírra	zɔʾɔ́	óγɛ; S pogɛ	tsˈabˈɔ́; S hanta		šɛɛcʾɛ́	híi-

(*Continued*)

Table 12. (Continued)

	65. person	66. rain	67. red	68. road	69. root	70. round	71. sand	72. say
Koyra	gɛ́rɛ	ʼírra; buk- v.	zɔʼɔ́	ogɛ́	ts'abɔ́		šɛɛcʼʼɛ́	hii-
Chara		í:ra	zúwa	boká	s'áb'a		ámša	díʼen
Bench	ats (atsin *woman*)	wol	zok'	par; bod	ś'ap'm	kar v.	k'amč'	mak; B ayt‘ gah
She	= 51	wol	dis	bod	ito		k'ac'	makɛsku
Yemsa	asu	iro; yeešu	šeʼa	ugna	tasa	zu:rà-kibbí	fuutu	i; u
Bworo	ašša	awšša	bira	werha	ts'ap'a	guura	šiya	et-
Anfillo	= 51	amɪtsso	ts'e:lo	gɔ:tto	ts'a:po		ša:po	k‘asa
Kafa	ašo; yaro	amiyo	č'elle; kawi; damme	woreto; qat'o; gommo	č'ammo; t'eppo	gudo	qač'o	getto
Mocha	ʼašo	ʼamiyo	č‘é·llo gà·mi- v.	worato	č'ammo	gú·do	ší·šo	hätti-
Dizi	yab	iru; vb. diɛb	sub-ɪs	kudu; Y ko:k	k'aŋg; B c'wažu	bullʌbul	harč	gɛg-
Shako	ya:b = 51	irú	sub`m̥s	ko:kn̥	kändu		k‘amč'u	ge:
Nayi	ya:b = 51	i:ru	B turbm	ku:t; B buřun	zɪga(s)		B bolu	B maho
Mao		ʼùmbʊd'ɛ̀	zà:ŋkítɛ̀	pò:mbɛ́; kʼòɸɛ̀	šášɛ̀		šá:ʷà	hà wìjá
Seze$_{(1/2)}$		ʼins$^{(,)}$í	tà:ntí	kʷei	s'áppʻɛ̀		šà:k[ɛ́]wí	wɛ́ŋ kʷá:mɔ́
Hozo	mo	ɪmsi	tanta	kaššɪ	t'appʻi		šakiwɪ	kooi
Dime	B eyyu	dɪ:bɛ; F k'an	F zu:-/B zo:-	gaš	č'eč'e	F mɪlu	ša:jɛ; F šayy	F geem-
Hamer	edi	F dubi	z-	F goiti	c'ac'		F šami	gi-; am-
Banna	edi	dubi; vb. k'an-	zia/za:wu; dɪrr	goyti	č'ač'i		šami	gi-
Karo	edi = M ɛ:di M ʼe:na	dobi	zʌwi			zola	šami	

Ari	ɛd	do:bi; vb. kaɣanda	ze-mi	go:gi	čʿačʿi; B tuudi	dıfın	ša:mi	ga-; B yiik
Ubamer	ɛd	do·bi	ze·-mi	go·gi; gɛrda	čʿačʿi	očıta	ha·ča	ga-
Galila	ʾɛ·d	dó·bi	zé·-mi	gó·gi	čʿa·čʿi	yıulla; mulla	p'ič' k'ač'	ga(y)-
	73. see	**74. seed**	**75. sit**	**76. skin**	**77. sleep**	**78. small**	**79. smoke**	**80. stand**
Basketo	B bɛkʾə-	B zɛrs	B dɔʾa-	B gəmma	B surkʾa-	B gɨlla	B cʾuyáž	B ekʾa-
Dokka	bɛkʾ-	zertse	duʾ-	gəmmá	sur-	gılla	čʾuž	ık'-
Male	zag-	zɛrtsi	dɛʾ-	zɛdʾi; ʾılʾi	gınʾ-; lah-	dʾaka; šai	ts'ubi	ʾɛk-
Wolaita	be'- A t'eel-	zeretta Ce gistiyaa	utt-	galba goga	tʿesk-; Ce zinʾ-	guutta	c'uwa; Ce tuumoo	ekk'-
Kullo	s'ell-	dʾoliyawa	wott-	galba	giis-edda	lap'a	c'uwa	ʾekk'-
Cancha	s'ell-	zirets	ʾutt-		ʾeʒʒ-ires	guuts	c'uwa	ʾekk'-
Malo	beʾ-	zérétʰ	bɛtɛ́:za	gallʌmáʾa	wɛiʾéza	ts'ít'a	čʾúja	ʾits'-
Gofa	beʾ-	zeretsa	ʾutt-	galba	zinʾ-is	gutsa	c'uya	ʾekk'-
Zala	beʾ-	gistiya	ut		zandʾ			
Gamu	beʾ-	zerets	ʾutt-	galba	ʾeʒ-ires	guts	cʿuwa	ʾek'-
Dache	beʾ-	zɛr	ut-	galba	zinʾ-	guts	cʿua	ık'-
Dorze	s'ell-	zirets	ʾutt-	galba	zinʾ-is	guuts	cʿuwa	ʾe'-
Oyda	bɛʾ-	budíts	bɛ́d-	galba	wɔ̲ʾ-	erats	čʾuggə	ɛk'-
Zayse	ts'ɛɛ'l-	diic'á	ʾut-	kɔ́mʾɛ	gɛh-	ɛrés	c'u'a	ʾɛɛʾ-
Zergulla	ts'ɛlli	gude; bič'ɛtta	ʾu:t-i	ʾite	wɛŋ-i	ʾɛréz	čʾua	ʾɛ:ʾi
Ganjule	bɛɛ-	sɛrɛ; S dič'	ʾut-	išila kɔ́lʾɛ	gɛh-	guuts'i	c'uua	ɛɛʾ-
Gidicho	bɛɛ-	séra	ʾut-	kɔ́mʾɛ	gɛh-	gúuts'i	c'úwwɔ	ɛɛʾ-
Kachama	bɛɛ-	séré	ʾut-	kɔ́mʾɛ	gɛh-	gúuts'i	c'úwwɔ	ʾɛɛʾ-

(*Continued*)

Table 12. (Continued)

	73. see	74. seed	75. sit	76. skin	77. sleep	78. small	79. smoke	80. stand
Koyra	bɛ-	sɛrɛ	ʾut-	išila kɔ́Rmʾɛ	gɛh-	ʾɔ́ɔd'e	cuwwa	ɛɛʾ-
Chara	béʾ-	dúk'a	kalm-; utna	gongá	gíʾ-	gútsa	č'úya	aš-ne; neyt'na
Bench	bek'	zar	wog	bet	sur	yars	ç̣'ub	yit'
She	beka	zar	oga	gel	suřka	uškəsuš	šarr	ita
Yemsa	biya; ičma	zala	di/duwu; mučo; kewa	go:go; ko:da; woma	afaʾo; gigitto	garo	čuwa	yero
Bworo	bek'- E ts'il-	šooka	bey-	k'eedda; gook'a	k'ey-; tokər-ey-	guuttsa mukk'a	ts'uwa	eed-
Anfillo	ts'i:n	ya:ro	t'ɛ:pa	gɔŋ:k'o	tok'ɨra	ts'ʊ:nt'o	s'u:ggo	e:ʾra
Kafa	beʾo; č'ino	yare	kotto; bero	ačo; goqo	tokketo	gamnani	č'uffiyo	net'o; timmo
Mocha	bäqqi-	yäro	kòta-; yäbba-		qè·yĭ-	gu·ččó; gi·šó	č'uṗó	nè·t'a-
Dizi	sʌg	butknɪs	al-	kut	sog-	yas-ɪs	č'ub-u	aš-
Shako	säg	B zara	wög	B ba:c'i	só:k; boʾe	kota	č'ubu	a:š
Nayi	B semo	B yaru	kalbm; B aša	B fatu	sok'-	B c'oant'ai	B cubu	B aškio
Mao	ʾá ʾínt'á	šókɛ́	há kù:wá	kútɛ́	ʾà há:là	nì:ts'íšʲè	ts'úwɛ́	hà n`ʊ:ŋkà
Seze$_{(1/2)}$	šɔ́šɔ/káɸàbá	šʷákk kìndì	kʷɔʲ	kús gɔ`ŋkà	héhɛ̀:já	má.nzì	s'ʊ´wì	dɔ́:k'ɛ́
Hozo	šoši	šok-	kobi	goŋk'a	hihɛ	t'ot'zi; S manza	s'ʊbi S ts'óbi	do-ɪ
Dime	B ɪŋg; F yeb	mʌšitʰ	F dah-	bɨče	F na(a)xt-	B toŋg-; bɪddɪ:k; F c'ɛkk'-; lek'	č'ubu	ʾuijujɪn
Hamer	šed-	F beta; maš-	dork'-	bɪc-	rat-	toko; F l*ɯ*kka	c'uba	woy-
Banna	šɛd-	beta	doɣ-	eba; za:ra	rat-	lɪkka	č'uba	w-

Karo		p'eta			wət-	šoli	c'ubo	wedi
Ari	šɛd-	p'eta	dork-	gofta; B ib'; gʌmma; zɛna	rat- B gin'-	toko-mi	č'oba	wʌ'-; d'áam-
Ubamer	šɛj-		doɣ-	ɣ/k'urra; guftá	rat-	tokº-mi	č'uba	
Galila	šɛd-	beta; məšá	doɣ-	zəna; gufta	rat-	ninč'a, liŋša	č'uba	wə'-
	81. star	**82. stone**	**83. sun**	**84. swim**	**85. tail**	**86. that**	**87. this**	**88. thou**
Basketo	B ts'olınta	B šuca	B awa	B wa.dá-	B gɔlʸsi	A yita	A yi	ne·na
Dokka	bɛrse	šučče	awá	piŋ-; gutt-	golse			ne·na
Male	ʒogo; wuts'atsi	šutsi	'abi	wajz; s'ub-	'uŋki	jeja	haja	nɛ:ni
Wolaita	t'oolint-iya	šucca	awa-'afiyu A hawa	wadd'- A wall'-	goyna A goyina	he-gge m. he-nna f.	ha-ge m. ha-nna f.	ne(-ni/-na)
Kullo	t'olint-iya	šuʒʒa	awayketta	wa'-	dup'-iya	hewa	haa	nena
Cancha	s'olinte	šuʒʒ	'awa	wat-	goyina	sekkisi	haytsi	neni
Malo	ts'olínto	šúč	'awá:tə	hʌtsed'-	gʷɔjnʌlʌ	jéji	hái	néhni
Gofa	s'olinto	šuʒʒa	'awa	wad'-	goyina	hinayssi	hayssi	neni
Zala	tsolentsa		awa		guoyina	killi		ne
Gamu	s'olinte	šuʒʒ	'awa	man'-	goyina	hetsi	haytsi	neni
Dache	ts'olinte	šuc	awa	lim	goyna		haysa	ne:na
Dorze	s'olinte	šuʒʒ	'awa	wat-	goyina	'ekkisi	haytsi	neni
Oyda	s'olınto	šú·čɪ	áwə	bək'-	goyna	yɛndea m.	hánno m.	ne·na/i m/f
Zayse	ts'ɔlínte	malɔ́	k'ɔ́ɔs	dakado	fiits'ɔ́	S sɔja	haya	neei
Zergulla	ts'ulıntə	malo	'awa	ts'ubi	goina	ja	haja	nɛna

(*Continued*)

Table 12. (Continued)

	81. star	82. stone	83. sun	84. swim	85. tail	86. that	87. this	88. thou
Ganjule	tsˈɔɔlintˈe	šúcˈi	awwá	dakaada	naatsˈé	S sɛsa	haba	nɛna
Gidicho	tsˈɔlínte	šúcˈi	awwá	dakaadá	naatsˈé		haasa	nɛná
Kachama	tsˈɔlínte	šúcˈi; B paló	awwá	dakáado	naatsˈɛ́	S haŋk	yaa	néna
Koyra	tsˈɔɔlínte	šúcʿɛ	awwá; kˈɔ́sɛ	dakade H šarɛ	naatsˈɛ́ dupʿiya	H seˈɛssɪ	hayzi H haˈɛssɪ	nɛmba H nɛn(ɪ)
Chara	wóncˇeta	šučá	oyá	wátˈ-	sʿéra	sék(ä)n	háribe	néni
Bench	šayt	nyal čˈar *rock*	obar	mošt	unk	yinkuš	haši	ne(-na)
She	zat	bak	obər	aki	uŋk	-ɛn kuš	aŋaš	nena
Yemsa	baakura	šuˈa	awa	ičo	unyu	(h)aas	(h)an	ne
Bworo	kˈeena	šuttsa	aawa	waak-	dušša	(ekke)wi	(han)ni	nee
Anfillo	tsˈɪgɛːro	šʊːtso	aːbbo	waːnno	dʊːtso	B bɛni	hani	noː
Kafa	tˈodjeno	tˈaqqo	abe	wako	čʿero	arro; B ebi	hini	ne
Mocha	tˈoğğo	šut/tˈó	a·be	wà·na-	čʿe·ro *hair of tail*	òkkabi B ɛbe	haní	B ne
Dizi	biz	ñal-u	kʌi; čyaž	B bóro	čʿɪr-u	ɛke m. yɛk f.	ɛŋkɛ m. yɛŋk f.	B yetu
Shako	barin; B dɛrtn	šéˀì	sˈeasˈù	múšn̥	aːsú	yis	àːz	B yɛtá
Nayi	B šeʲca	B nʸelu	B šaž	B wokn	B šokn	Y neás m. Y neyin f.	B yes	B yeta
Mao	bɛ̀lgɛ́	šòːwɛ̀	ʾáwdóːgɛ̀	hà ɸáː ɲá	jɔ́ngʷɛ̀	jɛ̀čɛ́	nàšá	hìšɛ́
Seze$_{(1/2)}$	kɔ́zzì	waːsi/šɛ́wì	ʾáʷkálì	pʲáŋál pʲàŋá	ŋwiːni	híjàgɛ́nɛ̀	ʾ ɛ̀ːčɛ́/ʾétmá	hínšɛ́
Hozo	koza	watzi	abbi	paaŋgɛ	ʊmbulɛ	S zaŋaŋà	zaay	hɪŋga
Dime	bɛz	laːlɛ; F miːtu	ʾɪjɛ; B eːyo	B goːq; F dal	golan	ʾɪsɛno; F sah	ʾɪsɛno = 86	jaːe

Hamer	ʾe:zini	seni	(h)ai	war-	guli	aga	ka	ya
Banna	wončo	sini	ai	war-	guli; dubʌna	aga	ka	ya
Karo	kaco	suni	ai		dibini	ka; M ʾo:	ka	
Ari	manka	se:ni	ay	B wa:rri; zo:tima	gali H goyríy	aga	kona	B á(:)na
Ubamer	mo·ga	se·ni	hai	mıl-	go·li		kona	a·nə
Galila	bɛz	sé·ní	ʾa·i		go·li	okona	kona	yıná

	89. tongue	**90. tooth**	**91. tree**	**92. two**	**93. walk**	**94. warm**	**95. water**	**96. we**
Basketo	B ɪnts'ɨrs	ačči	B mıts	nəmʾi	A lukk-	B mıcin	B wa:tse	nu·na
Dokka	ɨrs'ɪns	ačči	mittse	nəmʾi	luk(k)-	so·lle	wa·tsi	nu·ni
Male	ʾɪndɪrsi	ʾači	mitsi	lamʾo	ʾa:dʿ-, dend-	ʾoidi; tanane:na	wa:tsi	nuni
Wolaita	int'arsa	acca A acc'a	mitta	naʾa Bk láha	ham-ett-	hoʾʾ-uwa A homb-iya	hatta	nu(-ni/-na)
Kullo	ins'arsa	acc'a	barzap'-iya	laʾʾa	be-edda	hopʿ-uwa	hatsa	nuna
Cancha	insʿarsa	accʿa	mits	nammʾa	b-ires	hoʾo	haats	nuni
Malo	ʾɪrɪ´nts	ʾáčʰə	mɪ´ts	nʌ́mʾe	tʰɔhérʌ	waré	ʾátsə	núni
Gofa	intsarsa	acc'a	mitsa	nammʾa	bi-is	hoʾo	hatse	nuni
Zala	int'arsa	ačča	mitsa	naʾa	ba	mičče	hatsa	nu
Gamu	ins'arsa	acc'a	mitsa	nammʾa	b-ires	hoʾo	hatse	nuni
Dache	ɪntsɛrs	acé	šara	nəmə	b-	hoʾo	hatse	nu:na
Dorze	ins'arsa	acc'a	mits	nammʾa	b-ires	hoʾo	haats	nuni
Oyda	iláns	ʾač, PL. o·či	mɪns'a	namʾi	kʿɪtʿ-	wuʾo	haytsi	nu·ni

(*Continued*)

Table 12. (Continued)

	89. tongue	90. tooth	91. tree	92. two	93. walk	94. warm	95. water	96. we
Zayse	intsˈérɛ	ʾacc'	mitsˈa	nammʾa	hám-	hɛɛra; S binna	wáatsˈi	nuíi
Zergulla	ʾınsəre	ʾacˈe	mintsa	nam	haŋgi	mičˈa	wa:tse	nuna
Ganjule	intsˈúrɛ	gaggo	mitsˈi	nammʾa	haŋg-	mittsˈa	waatsˈi	nuna
Gidicho	intsˈúrɛ	gaggo	mítsˈi	náʾu	haŋg-		wáatsˈi	nuna
Kachama	intsˈúrɛ	gaggo	mitsˈi	nammʾú	hantˈ-	S binna	wáatsˈi	nuná
Koyra	ʾúntsˈúrɛ	gaggo	mítsˈe Ce akka	lamʾe	háaŋg-	mittsˈá; H bınna	wáatsˈe	numba H nunı
Chara	ʾínsˈila	áčˈa	mítsa	näntá	sˈálm-; hám-	kˈés'-	á:sˈa	núni
Bench	eyts'	gaš; san	inč	nam	opt; ham	Ra bín	soʾ	nu(-na)
She	ets'	gaš	enc	nam	šila	bins	soʾ	nona
Yemsa	terma	aʾya	iʾo	hep/yep	kutu; ama	gumu; kiiču	aka	inno
Bworo	albeera	gaša	mitta	gitta	am-; šaʾ-	kˈeetsa	aatsa	noo
Anfillo	ɛrɪ:tso	ga:ššo	mɪ:tso	guto	nɨm	kˈe:tsˈo	yu:ro	nu:ši
Kafa	ečˈiyo	gašo	metˈo	guto	šaʾo; deno	qečˈe; game	ačˈo	no
Mocha	häčˈawo	gášo	mitˈó	gutto	täri-; šäʾa-; hàmmi-	qè·ččˈi- v.	à·čˈo	B no:
Dizi	yabıl	ažu	wıč	tˈagn	tɛg-	kez-ız	a:i	B inu
Shako	érb	á:čˈu	í:nču	tˈa:gn̥	í:n; tấg	B sɛrəb	á:y	nấta
Nayi	B yalb	B acu	B incus	B tagn	te:; B tego	B kˈetsua	B hai	B natoknu
Mao	ántsˈílɛ̀	à:tsˈɛ̀	ʾí:ntsɛ̀	jʊˊmbʊˊʰ	hà hɔ́jjà	kárɛ́	hà:tsɛ̀	hámbílɛ̀
Seze$_{(1/2)}$	jántsílɛ̀/ tˈagál	há:tsˈɛ́, ha:nsì	ʾinnsì	dɔ̀mbɔ́/ nɔ̀mbɛ́	šɔ́:mɔ̀; já:jà	kˈalkˈalá / kˈàlɛ́wì	há:nsˈì	dɔ̀lšɛ́

Hozo	S wìntə́lə	atˢ'i	S ʾí:nti	dombo	yɛ-ɪ; S šɔ́:ti	šɛši	haani	nunga
Dime	ʾɨdəm	F baŋgɪl; ɪts; kəsɪl	ʾaχe; B ha:γo	k'ʌt'ɨm BF k'astɛn	tɪŋgɛ F b-	F su:l-; di:ne	naχe B nə:γ-	wɔtu F watu
Hamer	ʾaďʾab	ʾats'	ak'-	lama	yɛʾ-	oid'i	noko	wɔsi/wɔdi
Banna	adʌb/adɪm	atsi	ɑhaka/ha:k'a	lama	rond-	šoli; tʊ̀ll	noko	wɑdi
Karo	attəp'	asi	aka	lama		oyddoʾ	nuk'o	
Ari	adim	atsi; B kasel geegi	ahaka	k'astɛn	kay- B bad-	kumʾɪla	noγa B nɔk'ɔ	H wɔ(:)tá
Ubamer	admi	atsi	aγa	k'astɛn ~ ʾ-	kay-	oyd'ima	luk'a, luγa	wətá
Galila	admi	ači	aγa/aháγa	k'astən,ºkən	ay-	oyd'-; so:l-	lu·γa/lo·γa	wʊtá

	97. what	**98. white**	**99. who**	**100. woman**	***101. five**	***102. four**	***103. three**	***104. ye**
Basketo	A ʾa B ɔnɔ	B bɔrs	A ʾo, ʾoni *whom*	A ʾasi; macc'	ɪšɪ´n	oyddi	aydzi	ɪntɛna
Dokka	abza	borse	o·no	maččʼa	ɪššɪ´ni	oyddi	həydzi	ɪntá
Male	ʾaigo	bore	ʾonʾa	la:lo; matso	dɔŋgo	ʾɔido	haitso	ʾɪnsi
Wolaita	aybaa	bootta	oo(-ni)	macc'a-asa	iccaša	oydda	heezza	hiinte(-na)
Kullo	ʾaybe	botsa	ʾoone	maʒʒawa	icc'aθa	ʾoydda	hezza	intena
Cancha	ʾazzi	buts	ʾooni	maʒʒas	icc'aθa	ʾoydda	hezza	inteni
Malo	ʾábba	bó:ts	ʾó:ni	máč	ʾɪ´čɪn	ʾɔ́jddɛ	hʌ́jtsɪ	jɪ´ntəna
Gofa	ʾaybe	botsa	ʾoone	maʒʒa	icc'es'a	ʾoydda	hedza	ʾenti
Zala	ay(be)	buotsa	oone	maččo	ičč'a:ša	oidda	hezza	inte

(*Continued*)

Table 12. (Continued)

	97. what	98. white	99. who	100. woman	*101. five	*102. four	*103. three	*104. ye
Gamu	ʾaybe	ʾOts	ʾooni	maʒ̈as	ic'ac'a	ʾoydda	hedza	eeti
Dache	aze	bots	oné	indo			hedza	ɪntena
Dorze	ʾaybe	buts	ʾooni	maʒ̈as	icc'aθa	ʾoydda	hezza	inteni
Oyda	aybe	boʊts	ó·ni	B ma:šo	ičɪ´n	oyddi	haydzi	hɪntɛni
Zayse	ʾáalma	bɔ́ts'i	ʾɔ́di	bíšɔ; S ma:č'ɪna	S ʾɪšik	S ʾoit	haydz S hais'	wútuna S ʾutini
Zergulla	ʾalmɔ	bo:ts	ʾɔde	ʾɪnda	ʾišidʒ	hoid	haits	hutuna
Ganjule	ʾáala	bɔ́ɔts'i	ʾɔ́ɔni	máac'o	S šič'	S ʾo:d	haydzi	ínnina
Gidicho	ʾáala	bɔ́ɔts'i	ʾɔ́ɔni	máac'i			haydzi	innina
Kachama	ʾáalamba	bɔ́ɔts'i	ʾɔ́ɔni	máac'i	S ʾišidʒ	S ʾo:idu	hadzi	innina
Koyra	abaw	bots'e	ʾɔ́ɔni	máac'ɔ L ɨndo	H ʾičɪče	H ʾoddɛ	háydzi	inumba H hɪnunɪ
Chara	á:ze	dälá	ó:ne	mašná	učá	obdá	e:zá	ínti
Bench	awu; hare	dal	on(-a)	mayn	uč,	od	kaz	yint(i)
She	hai-	dal	onə-	wano	CR uč'	CR od	kaz	B int-aikn
Yemsa	aanba	poro	oon	aśin	uuč	ačeeč	keez	nitto
Bworo	eebi; eege	nattsa	koonni	maattsa	uttsa	awdda	keeza	itti
Anfillo	a:nto	nɛ:ts'o	konne	ma:ce	G amitto	G auddó	ke:djo	ɪnta:ši
Kafa	abiči, amoči	nač'o	koni	mač'e	učo	awudo	kedje, kemo	itoš, itto
Mocha	ámo	näčč'o	ko	ma·čč́é; gänne	u·čč́o	awuddo	kä·ğğo	B itɔ:ši

Dizi	yɪri	godn-ɪs	igte	kwuɨčin	Al účú	Y kubm	B ka:du	B iti
Shako	B yɪr-	got'ns	ítí	bay	CR uču/ učem	kúb`m̥	kə̀du	ítì
Nayi	B iřa	B gotnz	B ɪtia	B mošk n	CR učum/učču	kubm	B kadu	Y itì-kis
Mao	kɔ̀mísìjà	ká:wɛ́	kí[:]já	múnts'è	k^{w}íssí	mɛ́:ts'ɛ́	tɛ́zɛ̀	hàwɛ̀lɛ̀
Seze$_{(1/2)}$	kínà:	kà:wè^{j}	kínà:/ hɪškɪnɛwá	šɛi/ša:j	k'úsɛ́/ k^{w}íssɛ́	bɛs'ɛ́	sija:zɛ́/ sì:zí	úkkɛ̀
Hozo	kinda	kaawɪ	haya	šɛɛ	S k^{w}ɪtsí	S bɛtsí:	siyazɪ	doŋga
Dime	B wɔyɔ:	F goyt' B gɔ:d'-	'aje, B a:yɔ	'amze	šənne	'ʊddʊ	mʌk^{h}ɨ m	jɛ:se, B yɛtP, F yɛse
Hamer	har	c'auli	'ʌi	ma, PL. maná	F donk'	F oydi	m'ak'an	yɛsi/yɛdi
Banna	are	č'auli	ayne	ma	donk'	oydi	mʊkkan	
Karo		c'auli		meɨ	doŋg	CR oydì	CR makàmm	
Ari	are	s'a-mi	ayi	ma-ɪnts	H dónq	H 'oydí	H makkán	H yetá
Ubamer	hare	s'a-mi	ay	mana	doŋk'/doŋg	oyddi	məkkən	yɛtá
Galila	hárá	č'amɪs	háyí	mana	doŋk'	oyddi	məkkən	yetá

Abbreviations of the authors: A Alemayehu Abebe; Al Allan; B Bender; Bk Beke; Ce Cerulli; CR Conti Rossini; F Fleming; G Grottanelli; H Hayward; S Siebert; Y Aklilu Yilma.

Note: The underlined words indicate borrowings.

5. Etymological comments

The Omotic reconstructions indicated by H, B, E follow the results of Hayward (1988), Bender (1988), Ehret (1995) respectively. The asterisked protoforms without any of these attributions were proposed by the present author. The Omotic lexical material was systematically compared with the Cushitic data. The comparisons from other branches of the Afroasiatic macrofamily demonstrate that the Omotic branch represents an integral part of this genetic unit, but they are not exhaustive.

1. **all**

1.1. Om.: (N) Gonga **bull-*; Nayi *bula* || (S) Hamer **wull* id. ||| Cush.: (C) Awngi *wullá*, Kunfäl *wela* id. || (E) Arbore *búlí* id. (D 190 **wu̯Vll-*: Om. + Awngi + Arbore). Appleyard (CAD 22) thinks that Awngi *wullá* and Kunfäl *wela* are of Omotic origin. The Ethio-Semitic source of the type Amhara *hullu* "all" is not too probable (in contrary to Dime *kull* - see #1.6.).

1.2. Om.: (N) Ometo **ʾubb-/*ʾub'-*; Kafa *uba* id. || (S) Dime *wu:f* id.

1.3. Om.: (S) Galila *tunni* id., cf. NOm: Yemsa *tuuma* = H *tunni* "full" ||| Cush.: (E) Rendille *tuummán* "all".

1.4. Om.: (N) Ometo **wur-*; perhaps different from #1.1.

1.5. Om.: (S) Ari *muda*, Ubamer *mudda* id. ||| Berb.: (E) Ghadames *imda* "tout": *əmdu* "achever, compléter" (Lanfry) || (S) Ahaggar *imdu* "compléter" < impf. **ya-mduH*, perf. **yu-mdaH* (Prasse 1975, 227). Takács (*Lingua Posnaniensis* 41, 1999, 202–03), the author of this comparison, adds EChad.: Dormo *moid*, Gabri *moid*, Chire *mōdo* "10" (Lukas) and Sem.: Akk. *miʾādu*(*m*), *maʾādu*, *mâdu*(*m*) "to be/become many, numerous", *maʾdu* "many"; Ugaritic *mid* "plenty; much" < **maʾd-*, *mảd* "plenty" < **maʾād-*; Hebrew *mĕʾōd* "very"; Arab. *maʾada* "to begin to grow".

1.6. Om.: (S) Dime *kull* id. < Ethio-Semitic: Geez *kʷəll-*, Tigrinya *kʷəll-u*, Harari *kull*, Amhara *hullu* id. (Leslau 1987, 281).

2. **ashes**

2.1. Om. **bud-*(*in-*) (B 145: **bend*) ||| Cush. (E) Rendille (Fleming) *ben* id. (D 241: Om. + Rendille). Related can be Eg (Med) *wbd* "brennen; Russ" (Wb. I, 297). More problematic are counterparts in Chadic, reconstructed as **buwuṭ* "ashes" by Stolbova (1996, 21–22), WChad. **b*[*u*]*wṭa* (Ead. 1987, 149) = Chad. **bətu* id. (Newman 1977, 22).

2.2. Om.: (N) Gonga **tull-* id.

3. **bark**

3.1. Om.: (N) Ometo **pok'-* id. ||| Cush.: ?(C) Awngi SLLE *pa:q* id. (or metathesized from usual *qap*) || Dahalo *pak'o* id. || (S) Qwadza *paʾuko* id.

3.2. Om.: (N) She *šan* ||| Cush.: (E) **ša'n-/*ši'n-*(Ehret 1990, #27) > Somali *saan* "pelle spessa di animale conciata"; ?Yaaku *çé'nó* "bark" || (S) Qwadza *tseno* "skin".

3.3. Om.: (N) Mocha *óqqo*; Mao *k'òk'éšè* id. || (S) Hamer *wok'umpa*, Dime *'oxom* id. ||| or < Cush.: (E) Dullay: Tsamay *q'aaq'-ay*, Dopasunte *q'aaq'-ičče*, Gollango *qóoq-e* id. (H & AMS).

4. **belly**

4.1. Om.: (N) Basketo *lippe* id. || (S) Galila *lip'a* id., further Ubamer *lIp'a/lib'a* "heart" and also Dizi *lIb*, if it is not an Ethio-Semitic borrowing ||| Cush.: (N) Beja *léew* "Seite" (Reinisch) || (E) **lab*'- > Afar *lab* "side of the body" (Parker & Hayward); Somali *laab* "chest" (Abraham); Oromo *lapp'-ee* "heart, chest"; Burji *lap'-ée* "side" (E 426, #882: SOm. + Cush. **łap'-* "rib") ||| Chad.: (C) Daba *lìbí*, Musgoy *lib* "belly" (JI_2, 21).

4.2. Om.: (N) Ometo **gub-* > Male *gobi* id.; Shako *gùb* "chest" || (S) Ari *gubi* "navel" ||| Cush.: (E) Afar *gublo* "lungs" (Parker & Hayward), Saho *gubuul* id. (Reinisch) ||| Chad.: (W) **gaba* "chest, breast; in front of" > Hausa *gàbaa* "frontal part (of body), Bolewa *bo-gawa*, Ngizim *bò-gàbà* "chest", Dwot *gup*, Seya *gəp* "breast" (Stolbova 1987, 214) ||| Sem.: Arab. *ǧibāl* "body" (E 179, #264 compares Ari *gubi* with Ometo **gaaw-* "belly" + Arab. + SCush. **gubayi* "trunk"). Maybe, the following Semitic parallels are better compatible: Akkadian *gabbu* "brain"(?) : Hebrew *gab* "arcade sourcilière", Arabic *ǧubbat* "os qui entoure l'oeil, orbite" and/or Syriac *gᵉbīnā* "sourcil", Arabic *ǧabīn* "côté du front, front, tempe" (Cohen 1970f, 94, 96).

4.3. Om.: (N) Ometo **ga:w-/*gaww-* id. ||| Chad. (E) Kera *gaw* "female breast" (Ebert) ||| Berb.: (N) Morocco Central *agawa* "cou long; gorge" (NZ III, 911) ||| Eg. Pyr *ḏ.t* "body" (Wb. V, 503) ||| Sem. **gaww-* "interior, inwards, centre" (Cohen 1970f, 106) = **gaw(w)i'-* "(front part of) body; chest, belly" (SED I, 92–93; Greenberg 1963, 53).

4.4. Om.: (N) Gimirra **šil* id. ||| Cush.: (E) Burji *salay* id., Harso *saalasiḫ-te* id. (Sasse 1982, 163) ||| Chad.: (W) North Bauchi: Jimbin *sala* "liver" (Skinner) || (E) Kera *sáḷḍé* "liver" (Ebert) (E 269, #492: Bench + Chad.) ||| Sem. **šily-at-* "placenta", **šalīl-* "embryo" (SED I, 216–17).

4.5. Om.: (N) Basketo-Dokka **mill-* id., cf. Kafa *mullo* "heart", maybe Mao *mä:le* "liver" ||| Chad.: (W) Bokkos *mulut* "liver" (Jungraithmayr); Burrum-Kir *mal* id. (Shimizu) || (E) Sokoro *mə̀l(ù)* "belly" (JI_2, 21); Sokoro *mēíldūm* "liver" (Lukas).

4.6. Om.: (N) Yemsa *kata* id. ||| Chad.: (W) Diri *kwàtə́n*; Tala *ka:du* id. || (C) Bura-Pela *kuta* id. (JI_2, 20; E 229, #399: Yemsa + Chad.).

4.7. Om.: (N) Ometo **'ul-* id. ||| Cush.: (E) Saho *ulu*ᶜ "intestine, belly", Afar *uleᶜi* "intestine" (Lamberti & Sottile 1997, 286: Ometo + Saho-Afar).

5. **big**

5.1. Om.: (N) Ometo **dar-* id.; cf. Bench *dorg* "fat, stout" ||| Cush.: (E) Afar-Saho *adar-* "be big" || ?(S) **dir* "big" (Kiessling & Mous 2003, 96) ||| Sem.: Sabaic *drr*

"plenty" > "harvest" (Biella 1982, 86), Arab. *darra* "to abund, yield in abundance" (E 135, #150: Ometo & Bench. + Arab.).

5.2. Om.: (S) Dime F *gad* "big" || (N) Zergulla *gudɛsa* "all" ||| Cush.: Beja *gwud* "many" || (C) Awngi *gud* "good" (Hetzron), cf. also Amhara *gud* "wonderful, marvelous" (CDA 76) || (E) **gud-/*gad-* "big" > Elmolo *guuta/guuda* "many", Arbore *guudá* id., *gudd'á* "big", Dasanech *guddu*, Oromo *guddaa*, Konso-Dirayta *kutt-* id. ||| Berb.: (N) Central Morocco *gudy* "être nombreux, en grand nombre, beaucoup" (NZ III, 737) ||| Chad. **gudV* "many": (W) Bolewa *gɔdɔŋ* "many" || (C) Padukwo *guda* "many" (Newman & Ma 1966, 237) ||| Sem.: Sabaic *gdd* "(the) great, great ones" (Biella 1982, 65), Arab. *ǧadda* "to be great", Tigre *gäddä* "être plus grand" (Cohen 1970f, 99; Greenberg 1963, 59, #48: Bolewa+Padukwo+Wandala *kwottya* id. + Beja + Oromo; E 180, #265: Dime + Cush. + Chad. **g-d-* + Arab.)

5.3. Om.: (N) Ometo **git-* id. (E 183, #275) ||| Cush.: (E) Oromo Harar *guutuu*, Borana *gutu* "full" ||| Berb.: (E) Sokna *uggut* "abondamment; beaucoup; très" | (N) Djerba *egget* "beaucoup; très"; Shawiya *igit* "être nombreux", Kabyle *gg^{w}et* "abonde", Shilha *igut* "abonder; être abondant, nombreux, en grande quantité, *tugut* "multitude; nombre; abondance" | (S) Ahaggar *iǧat* "être nombreux, abondant", *ăǧut* "grande quantité; la plupart" (NZ III, 908–09) ||| ?Chad.: (C) Wandala *kwottya* "many".

5.4. Om.: (N) Ometo **ʾad'-* > Zayse *adó*, Ganjule-Gidicho *arʾó*, Kachama-Koyra *arʾ* Q id. ||| Cush.: (S) **ᶜatl-* "be mature, fully grown" (Kiessling & Mous 2003, 62; Ehret 1980, 275) ||| Eg. OK *ᶜd* "Fett" (Wb. I, 239) ||| Sem.: Arab. *ᶜaZZ* "to be hard, firm, strong" (E 356, #703: Ometo + Cush. + Eg. + Arab.).

5.5. Om.: (N) Yemsa *innya*; Bworo *eena* id., cf. Mocha *eenn-* "to grow" (Lamberti 1993b, 267; E 467, #983: Yemsa + Cush. **wayn-* "to grow" + Chad. **w-n* "full").

5.6. Om.: (S) Hamer **geb-i* id. || (N) Mocha *gábino* "large" ||| Cush.: (N) Beja *goob* "melted fat" (R. Hudson) = "fat" (Bender) || (E) **gabʾ-/*gaʾb-* "fat" > Afar *gabuulu* "fat", Oromo *gabb-uu* "fat, thick" (Gregg), Konso-Dirayta *kapp-*; Gawwada *koop-i* "fat person"; Burji *gabb-óo* "fat"; Yaaku *keʾp-ey* "to become fat" (Sasse 1982, 74) || (S) **gab-* > Iraqw *gawit* "hard"; Maa *-gábi* "to become fat" (Ehret 1980, 234) ||| Chad.: (E) Dangla *góób'é* "to fill" (JI$_2$, 157) ||| Sem. **g-b* > Arab. *ǧabǧab* "abondante (sorce), nombreuse (armée)" (Cohen 1970f, 95; E 179, #262: Om. + SCush. + Arab. *ǧabl* "numerous", *ǧabbar* "great, mighty".).

5.7. Om.: (S) Basketo *b'ɛts*, Dokka *bɛče* id. ||| Cush.: (C) Khamir *bɨtšťχʾ* "many", Khamtanga *bɨčˈíq*, Waag F *bɨčɨwa* = Bk *bíčikh* "many". Appleyard (CDA 97) sees here the Ethio-Semitic origin, cf. Geez *bäzxa* "to be many".

6. **bird**

6.1. Om.: (N) **kap-(*E 216, #367: NOm. **k-* < Om. **x_1-*; B 145 connets it with SOm. **apt-i*).

6.2. Om. (S) **ʾapt-i* (B 145).

7. **bite**

7.1. Om.: (N) **sa:tts'-*(H 268, 280; B 146: **sac'*; E 164, #232: **sats'*-).

7.2. Om.: (S) **ga*'- id. ||| Cush.: (E) Saho *gaa*^{c} *a* id. (Reinisch); HECush. **ga'm-* id. (Hudson 1989, 27: **gam*'-).

8. **black**

8.1. Om. (N) **kar-* id. (E 217, #368: NOm. < Om. **x_1-) ||| Cush.: (C) **qir-/*qar-* "night" (CDA 105; D 206–07: NOm. + CCush.).

8.2. Om.: (N) Gimirra **ts'id* id., maybe Mao *tì:šindɛ̀* id. ||| Cush.: (S) **ts'ed* "red" (E 283, #529: Bench + SCush.).

8.3. Om.: (N) Dizoid **(t)s'ahn-* id. || (S) Dime *ts'a:no*, Karo *ts'ia*, Ari *s'ia* id. (E 169, #246: Dizoid + Dime + Yaaku *t'eeho* "charcoal").

8.4. Om.: (S) **č'ale-mi* id., cf. NOm. **č'il-* "green" (#35.1) ||| Cush.: (E) Gedeo *t'illo* "black", Sidamo *c'aale* "shadow" ||| Chad. **(n-)č'[i]l-* > (W) Angas *ǯil* "shadow"; Sha *čalâ* id. (Stolbova 1987, 194) || (C) Gisiga *nʒala* "west" (Stolbova 1996, 48); Buduma *tsillim* "black" ||| Sem. **ṯ-l-m* "black, darkness", **ṯ-l-l* "shadow" (E 293, #556: SOm. + ECush. **c'ilḥ-/*c'ulḥ-* "ember, charcoal" [see Sasse 1979, 22, 30] + Sem. + Chad.).

8.5. Om.: (N) Gonga **a:k'-* id. || (S) Hamer *'ɛk'* "green".

9. **blood**

9.1. Om.: (N) Ometo-Gimirra **suguts-* id. (cf. H 265, 279; E 165, #235 **sux*$_2$*ut(s)-*) ||| Cush.: (E) Dullay **saagan-/*sa'aan-* > Gollango *saakan-ko*, Gawwada *sakán-ko* "meat" (AMS) vs. Tsamay *sa'án-ko* id. (H); Yaaku *sógó* id.

9.2. Om.: (S) **zumb'-* id. || (N) Zayse *zonn-e* "pus" (H 265) ||| Chad. (W) **ʒyam* id. (Stolbova 1987, 140), besides Hausa *jini* id. < **ínyam?* || (C) Gudu *ajin* id. (Newman & Ma), besides Mwulyen *zaambo*, Bachama *zàmbäy* id., ?Daba *žèmi* "vein" (all Kraft) ||| Berb. (S) **z-n-H_1* "blood" > Ghat *azəni*, Awlemidden *azəni*, *ažni*, Ahaggar *ahni*, PL. *ihnîtän* (Prasse 1974, 169: **ā-zinīH*, PL. **ī-zinītan*; Militarev 1991, 253) ||| Eg. Pyr *znf* "blood" (Wb. III, 459) ||| Sem. **zam* "blood" > Mandaic *zma*; Geez *zam* id. (Leslau 1987, 638; SED I, 266–67).

9.3. Om.: (N) Seze-Hozo **hambile* id., cf. Bench *bal* "menses" ||| Cush.: (E) Afar *ᶜabal*, Saho *bíilo* "blood".

9.4. Om.: (N) Gonga **dam(m)-* < Ethio-Semitic **damm-*(E 132, #140 finds in the Gonga form a regular continuant of AA **dim-/*dam-* "blood" in his reconstruction).

9.5. Om.: (N) Dizoid **yerm* < Surma: Mesengo *ye:rm* id. (Bender), cf. Nilotic: Acholi *remo*, Bari *rema*, Kuku *rima*, Shilluk *Remo* id. (Fleming 1965, 645).

10. **bone**

10.1. Om. **k'os-* id. > (N) Dizoid **'u:s-* id.; Hozo *käši* "foot, leg" (Fleming) || (S) Dime *k'öss* (Bender) = *k'ʊs* "bone" (Fleming) ||| ?Cush.: (E) Dasanech *yás*, PL. *yàss-ù* "foot, leg" (Sasse) = *k'as* "Bein" (Haberland) ||| Chad. **k'as[i]* "bone" (Stolbova

1996, 65–66) > Hausa *ƙàshíí*; Daffo *kyâs*; Gerumai *ókashi*; Kariya *ƙáásù* || (C) Munjuk *keslke*; Logone *aaʂe* || (E) Dangla *kaaso*; Kera *kàskàŋ*; Sumray *guseŋ* id. (JI_2, 36–37) ||| (E) Ghadames *yess*, PL. *yāsen* (Lanfry) | Shilha of Tazerwalt *iḫs* (Stumme); Kabyle *iyes*(*s*) | (W) Zenaga *i'ssi*, PL. *a'ssən*^{h} (Nicolas) | (S) Ahaggar *eyäs*, PL. *iyäsân* "bone" < **ē-yasaH*, pl. **ī-yasāH-an* (Prasse 1974, 72) or < **a-qisi*, PL. **i-qisīy-un* (Vycichl 1978, 73) ||| Eg. (Pyr.) *qs* id., Copt. *kas* (Wb. V, 68) ||| ?Sem.: Arab. *qassa*, *qasqasa* "to pick bone clean and suck it out" (Greenberg 1963, 53, #11: Chad. + Eg. + Berb.; E 240, #428: Om. + Chad. + Arab.).

10.2. Om.: (N) **mik'*[*a*]*ts-* ~ **m*[*a*]*k'its-*(with the nominal suffix **-ts-*) > Dokka *mik'ɔts*, Malo *mɛk'ɬts'*, Bworo *mak'ɔttsa* (H 278 adds Male *mik'etsi*) ||| Cush.: (N) Beja *miik^{w}a* "long bone (femur, humerus, tibia)" || (E) **miq-* id. (Sasse 1979, 49). Maybe also (C) **ŋac-* id., although Appleyard (CDA 35) prefers its comparison with ECush. **moč'-* > Yaaku *moč'o* "bone".

10.3. Om.: (N) Chara *mertá*; Gimirra **mert*; Mao *má:lt'ɛ́* id. together with Dahalo *mitľo* id. can be of substratum origin, cf. Khoisan: Hatsa *mitľa & mi//a* (Ehret 1974, 67, 88).

10.4. Om.: (S) **lef-i* id. < Cush. (E) **laf-* id. (Sasse 1979, 21).

11. **breast**

11.1. Om.: (N) **d'iam-*(E 438, #915; B 145: **t'am*; he adds Hozo *ammi*, Seze *'ámmɛ̀* and SOm. **'ami*). The ECush. parallels, Konsoid **'amb'-* and Dullay: Harso *ᶜaa-micakkó* id., besides Tsamay *ᶜáb'ún-ko*, PL. *ᶜamb'-e*, can represent the Omotic loans or vice versa.

11.2. Om.: (N) Gonga **ge:t-* > Bworo *gerha & gerra* "Brustkorb", Mocha *gäto* id. (Leslau) = *geto* (Lamberti), Anfillo *geeto*, Kafa *geeto* id. || ?(S) Ari *geeti* "Schulter", if it is not borrowed from Oromo *gatii-tti* "Schulter" ||| Cush.: (E) Somali *gaad-do* "Brustkorb" (Lamberti 1993a, 314).

12. **burn**

12.1. Om. (N) **ts'ug-* id. ||| Cush.: Dahalo *t'ogg^{w}a* "smoke" (E 168, #242: NOm. + Dahalo).

12.2. Om.: (N) Ometo **'e:ts'-*; Kafa *at'io*, Mocha *'à·tt'a-*; Dizi *ats-* id. || (S) **a*(*t*)*ts'-* id. (B 145: Om. **at*; E 355, #699: Om. **as'-* + ECush. **ᶜac-*/**ᶜic-* "to shine, glow" with no examples; perhaps Afar-Saho *ᶜasa* "to be red"; Somali *ᶜas* "red" - see Black 1974, 203).

12.3. Om.: (N) Om.: (N) Bench *bunk'*; Dizoid **bonk'-* id. ||| Cush.: (N) Beja **(š-)b-k^{w}*: *bešáak^{w}* "gekocht sein, reifen", caus. *šíšbáak^{w}* "kochen" (Almkvist) || (E) Burji *bukk'-* "to burn", besides HECush. **bok-ees-* "to ferment", but cf. Amhara *bkk* "to ferment", compatible with Arab. *bkbk* "bouillonner" (Cohen 1970f, 64) || Dahalo *ɓakk-eed-* "to kindle" (Ehret 1980, 142) ||| Eg. (D XVIII) *bkbk* "ein Gebäck" (Wb. I, 430) ||| Chad. (W) **bak[k]-* "to burn" (Stolbova 1987, 152) > Hausa *bàbbakà*;

Bole *bokk-*; Tule *bàk*; Ngizim *baka* || (C) Muktele *bə́kâì* id. (JI_2, 54–55) || (E) Jegu (Jungraithmayr) *bíik* "rösten", Dangla (Fedry) *bíkè* "faire griller de la viande au dessus du feu" (EDE II, 339: HECush. + Chad. + Eg. + Sem.).

12.4. Om.: (N) Ometo-Gonga **miɕ'-*(cf. H 280) ||| Cush.: (S) **mits-* "to shine, glitter" (Kiessling & Mous 2003, 208)..

12.5. Om.: (S) Banna *ko:ka* id. ||| Cush.: Dahalo *kok-* id.

13. **claw**

13.1. Om.: (N) **ts'igum- ~ *ts'ugim-(*B **ts'ukum*) > Zergulla *ts'ugumi*, Seze *ts'úkùmi*, vs. Basketo *s'uŋgıts*, Malo *ts'ángúts*, Mocha *č'ungət'-o* id. (H 269: Ometo **ts'ugun-*; Id. 278 adds 'Pseudo-Male *ts'iguntsa*; Lamberti 1993a, 400 thinks about a relation with HECush. **t'ulung-* "claw") || (S) Dime *šukum*, Ubamer *šuk'(u)ma* id. ||| Cush.: (S) Asa *seŋget*, Qwadza *tse'uko* "claw" (Mukarovsky 1981, 203: Om. + SCush.). But cf. also NS: Nara *šokna* "claw, foot", Kuliak **sak/*sag* "claw" (Ehret 1981, 274).

13.2. Om.: (N) Male *dapo* (Ehret), Gofa *dap'o*; Yemsa *difna* (E 133, #145).

13.3. Om.: (S) **guš-(*E 187, #288: SOm. **guš*$_1$- + Sem.: Arab. *ǧassa* "to touch, feel").

13.4. Om.: (S) Ari *šuqlí* id.; cf. perhaps NOm.: Kachama *sukule* "piede" (Conti Rossini) ||| Cush.: (E) Konso *šoloqloq-*; Harso *šulšúlle*; Yaaku *ségíl* "claw" ||| Sem. **s/šuql-* > Arab. *suql* "flancs, hypocondres", Soqotri *suqal* "pied" (Leslau 1938, 290; SED I, 213–14; Tigre *ɕ̌ənqəlit* "coude"; Amhara *ɕ̌əqəll* "elbow", Zway *ɕ̌iqile* "forearm" resemble the ECush. counterparts of the type **c'ikl-* "elbow" (Sasse 1979, 30 & 1982, 105, 176). Siwa (EBerb.) *siqāl*, PL. *suaqīl* "jambe" has been explained as the Arabic loan. On the semantic dispersion "claw" vs. "foot" – cf. Slavic **nogъtь* "claw" vs. **noga* "foot"

14. **cloud**

14.1. Om.: (N): Ometo-Yemsa **šaar-*, cf. She *šarr* "smoke", Bench *ṣ̌art* "to be foggy" (H 270; E 269, #493) ||| Cush.: (N) Beja *šaay*, PL. *šay* id. || (E) Elmolo *séér* id., cf. Rendille *seréy* "sky", *saraat* "of the sky" (D 193: NOm. + Beja).

14.2. Om.: (N) Mocha *gu:ppo* "cloud, fog"; Shako *gìbu* "cloud" ||| Berb.: (W) Zenaga *agayaf* "vent frais après pluie", *agafey* "Nord" (NZ III, 920, 750) ||| Eg. (Pyr.) *ỉgp* "Regenwolke, Gewitterwolke, Gewölk" (Wb. I, 140) ||| Sem. **g-f-y*: Jibbali *gofe'* "shadow", Mehri *həgfō* "to throw a shadow" (Cohen 1970f, 169); further Tigre *gíf^{c}e* "Wolke" > Bilin *gifí^{c}* id. (Reinisch 1887, 424, 143; Cohen 1947, 13: Eg. + Berb. + Bilin; E 185, #282: Mocha + Eg.).

14.3. Om.: (S) Dime *č'i:č'-* "cloud, fog" || (N) Male *č'i:š-i* "cold weather" (H 266).

14.4. Om.: (N) Ometo **d'um-*(E 437, #913) > Ganjule-Koyra *d'umá*, Kachama B *d'uma* id., ?Oyda *d'one*, maybe also Wolayta *t'uma* "darkness" = (Ce) *tu:mo:* "smoke" ||| Cush.: (E) **d'um-* > Saho *d'um-* "to become dark"; Konso *d'um-* "to set (of

sun)" ||| Chad.: (W) Pero *ḍamḍam* "darkness" || (E) Mubi *ḍèḍém* "night" (Takács, *Lingua Posnaniensis* 38, 1996, 61–62: Ometo + Chad.).

14.5. Om.: (N) Dache *damɛna*; Bworo *dawna* id. < CCush. (E 133, #143): Bilin *damna*, Khamir *dimna*, Khamtanga *dɨ́mána*, Awngi *damminí* "cloud" (CDA 46), borrowed into Saho Re *damäna* id. and Ethio-Semitic: Geez || (E) **duman-/*duban-* > Rendille *dubbat*, Bayso *dumbo*; Oromo Borana *dumansa*; Hadiyya Ce *duubanco* id., Burji *duubanci* "fog" ||| Chad.: (W) Hausa *daamunaa* "rainy season", Ngizim *dəmán* id., Bade *dəmànən* id. || (C) Kotoko: Logone, Kuseri, Gulfei *deman* "rain" (D 51: NOm. + Cush. + Chad.).

14.6. Om.: (N) Wolaita *guma* id. < Amhara *gum* "fog, mist", *gime* "cloud", cf. also Geez *gime & gum* "fog, cloud, dampness, mist, vapour", Gurage *gūwä*, Tigre *gim* "fog, cloud", Tigrinya *gime*, *gəmä* "cloud, mist". The word was also adapted by various Cushitic languages: Beja *giim & guum*; Awngi *gum* "cloud"; Sidamo *gomicco* "fog" (Leslau 1987, 193).

15. **cold**

15.1. Om.: (N) Ometo **za:(-)k-* id. || Ari *za:z-* "to feel cold" (H 265: plus Male *za:k-* "cold").

15.2. Om.: (N) Kafa *qewo*, Mocha *qäwo* "cold", cf. Mao (B) *kawɛ-* "wet" (?) ||| Cush.: (C) **känb-* "cold" (CDA 46–47) || (E) **qab-* id. (Sasse 1979, 49) ||| Eg. Pyr. *qb* "cold", Copt. *kba* id. (Wb. V, 22; E 233, #408: Mocha + ECush. + Eg.).

15.3. Om.: ?(N) Yemsa *koočo*; Shako *k'et',s* id. || (S) **k'a(:)ž-* id.; cf. NOm.: Mocha *qà·ǧǧi-* "to be wet, green"; Dizi *k'ɛžiž* "wet" (E 236, #418: SOm. & Dizi) ||| Cush.: (C) Bilin *qajqaj-* "to be cold", Khamir *qazqaz-*, Qwara *xexez* (all Reinisch), Kemant *ḫašḫaš-* (Conti Rossini), Khamtanga *qäzqázâw* id. (Appleyard), Awngi *kazkazzi* "cold" (Hetzron) vs. Amhara *k'äzäk'käzä* id. (CDA 47) || (E) HECush. **k'iiz-* "cold" (Leslau 1980, 120; Hudson 1989, 43: **kiida*, but Alaba-Qabenna *qiizá* indicate **-z-*).

15.4. Om.: (N) Gonga **'akk'-* id. ||| Cush.: ?(N) Beja *eeki* "mist" (R. Hudson) || (C) Awngi *ạyúmi* (Beke) = *ayomíi* (Conti Rossini) = *ɨyumi* (Bender), Kunfäl *ome* id. (D 263 & CDA 47: Gonga + Awngi) ||| Chad. (W) Zaar *yàkŋ* "cold" || (C) Glavda *áákwhyà* id. (JI_2, 79).

16. **come**

16.1. Om. (N) **yi'-* id. || (S) Hamer *yɛ'-* "to go" ||| Cush.: (N) Beja *yi'e* (R. Hudson) = *'i* "to come" (Roper) || (E) Somali *yaa'* "to run away"; Gedeo *e'y-*, Sidamo *e'-* "to enter" (Hudson 1989, 58) ||| Chad. **ya* "come!" (Stolbova 1996, 91: **ya'-/*yaw-* "go") ||| Berb.: (N) Kabyle *ɔyya* "come!" ||| Eg. (Old) *ỉj*, *ỉw* "to come", Copt. *ei* (Wb. I, 37, 44; D 184 and E 477, #1007: NOm. + Cush. + Chad. + Eg.).

16.2. Om.: (N) **w-* > Chara *woy-*; Bench *w-*, She *wu*; Yemsa *wu!*; Gonga **wa-* id. (B 146: **wa*) ||| Cush.: (C) Awngi *áw* "come" (CDA 48) ||| Eg. (MK) *w3j* "to come" (Wb. I, 246) ||| Chad. **wa'/y-*(Stolbova 1996, 91: Chad. + Eg. + Om.; E 469, #987:

NOm. **w-* & Ari *wo*'- "to stand up" + Cush. **wa:*'- "to rise, go out" + Chad. **w-* "to pass by, go out, come, go").

16.3. Om.: (S) Hamer *ni*'- ||| Chad.: (W) Ngizim *nì* "to come, go", *nài* "to come" || (C) Musgu *na* "aller", Masa *nàná* "to go" (Stolbova 2005, 96) ||| Eg. (old) *nᶜj* "to travel (by ship)" (Wb. II, 206; E 323, #627: SOm. + Cush. **naaᶜ-* "to go" [without examples] + Chad. + Eg.) ||| Sem.: Hebrew *nāᶜ* "to move", Geez *naᶜā* "come!", *naᶜawa* "to hunt" (Leslau 1987, 382; Cohen 1947, #447: Eg. + Sem.; EDE I, 351).

16.4. Om.: (N) Male *ad-* id. (Fleming) || Ari **'a(:)d-* id. ||| Cush.: (E) Saho *-ad-/-ed-* "to go", Somali *aad-* id. || (S) Asa *adi* ||| Berb.: (N) Izdeg *addu d*, Zayan *əddu d* "to come", Senhaja *addu* "to go" || (W) Zenaga *ėdda* id. (Militarev 1991, 255, 262) ||| ?Sem. **ᶜ-d-y/w* "to pass, cross, run" (Leslau 1987, 56–57; D 242: Male + Ari + ECush. + Sem.; Militarev 1991, 255: Male + ECush.+Asa + Sem. + Berb + WChad.: Sha *du-* "to come").

16.5. Om.: (N) Ometo **hay-* id. Perhaps borrowed from NS: Koman: Kwama *hɔyɔ*, Komo *hawi* id.

17. **die**

17.1. Om.: (N) Ometo **hayk'-*; Kafa *aqqo*; Mao *'à hék'"* id. ||| Chad.: (W) Ron: Daffo *hyek*, Fyer *'ek* id. (JI_2, 212).

17.2. Om.: (N) Yemsa *kitu*; Gonga **k'it-* id. (E 240, #430: Yemsa & Gonga) ||| Cush.: (C) **kət-* id. > Bilin *kər-*, Khamtanga *kər-/kət-*, Kemant *ki-*, Awngi *kət-/kər-* id. (CDA 54; D 245: CCush. + NOm.; Dolgopolsky also added Yaaku *akehet* id. by Greenberg, but it has to be *-kɛhɛ* by Heine) ||| Berb.: (S) Awlemidden *k̲ətəyyət*, verbal noun *ak̲ətəyyi* "marcher, se déplacer de côté, s'en aller, disparaître, mourir" (Alojaly), Ahaggar *kətiyət*, perf. *iktäyyät* "mourir" < **yikitiyyit*, perf. **yikitayyat* (Prasse 1973, 229; Militarev 1991, 247, 255: Kafa + Ahhagar).

17.3. Om. (S) Hamer **di-*, Ari **de*'- id. ||| Cush.: (E) Oromo Harar *du'a*, Konso-Dirayta *tow-* ||| Chad.: (W) Bolewa *du* "to kill", Angas-Sura *tu* id. || (E) Jegu *dá*, Mubi *dii*, Sokoro *dee* id. (Illič-Svityč 1971, 225, #6: Oromo + Chad.).

18. **dog**

18.1. Om.: (N) Ometo-Yemsa **kan-*, Gonga **kunaan-*, Gimira-Dizoid **kyan-* id., Mao & Ganza *kana* (Bender 1990, 602) || (S) Dime *kɛnɛ*, Galila *kani* id. ||| Chadic: (West) Fyer *kweéŋ*; Bade *wuunaayá*; (Central) Gamergu [Benton] *kěnee*; (East): Jegu *kany*, East Dangla *kaanyà*, Migama *kânnyà* "dog" (JI_2, 106–107 and JI_1, 49; let us mention that the authors derive these forms from the skeleton **k-g-n*; they think about a Saharan origin). Another cognate can be found in Guanche: Gran Canaria *cuna* "dog", Tenerife *cancha & cuncha* "dog, puppy" (Militarev 1991, 256).

18.2. Om.: (S) Hamer *k'aski*, Banna *qaski*, Ari *'áksi*, Bako *aksi*, *aksɛn*, Ubamer *aksi*, PL. *aksin-kɛti*, Galila *akši* id. ||| Cush. (C) **gəzəŋ* / **gəzäŋ* "dog" (CDA 56) > Bilin *gədəŋ*, PL. *gəšən*, Khamir & Kemant *gəzəŋ*, Awngi *gséŋ*, Kunfäl [Cowley]

kassaŋ (Appleyard 1984: 43, 44) ||| Chadic: (W) Ngizim *gà:sâ* || (C) Musgoy *h<u>á</u>z<u>a</u>*, Daba *xìzà* etc. (JI_2, 106–07) ||| Berb: Augila *aγzin*, Beni MenaZer *aqzīn*, Messaud *aqžaw*, Kabyle Irjen *aqžun* "dog", Rif *aqzin*, Seghrushen *aqqzin* "puppy" (Militarev 1991, 255: Berb. + CCush. + Eg.) ||| Eg. (MK) *<u>t</u>zm* "hound" (Wb. V, 409; Militarev 1991, 255; Appleyard (p.c.) pays attention to the fact that CCush. **ŋ* reflects Cush./AA **m*).

18.3. Seze *wiš*[*š*]*i* id. ||| Cush.: (E) HECush. **waša* id. (Hudson 1989, 52 thinks about a borrowing from Amhara *wáša* id., but it is possible to accept the opposite direction too) ||| Chad. (E) Bidiya *ʾùsù* "dog" (Alio) ||| ?Eg. (NK) PL. *ỉš* "Art Hunde [neben: Schakalen] die das Sonnenschiff ziehen" (Wb. I, 134) ||| Sem.: Arab. *ʾaws* "loup, chacal", cf. Geez *ʾawst/ᶜawst* "vautour" (Cohen 1970f, 13). Cf. also NS parallels: Nara *wos*; Taman: Merarir *wiis* "dog" (Ehret 2000, 583, # 1436).

19. **drink**

19.1. Om.: (N) **ʾus*(*s*)–(H 282; E 342, #665: NOm. **uš*$_3$-; B 146: **uš*).

19.2. Om.: (S) **wučʿ*- id. (E 455, #953; B 146: **wocʿ*).

20. **dry**

20.1. Om.: (N) Ometo **mel*- id. ||| Cush.: (E) Kambatta-Sidamo *moola* id., *mool*- "to dry" (Hudson 1989, 54; Bender 1975, 160: Wolaita + Sidamo) ||| Sem.: Arab. *mall* "live coal", *mallat* "hot ashes" (Steingass 1988, 1050).

20.2. Om.: (N) Chara *kó:la*; Gimirra *kol*; Dizi *kal*-, Nayi *kola* id. ||| Cush.: (E) Somali *kulayl*, Rendille *kuléel*, Arbore *kuld'á*, Elmolo *kúla*, Dasanech *kul* "warm/hot" || (S) **xal*- "dry" > Qwadza *kalahayi*, Alagwa *xala*ʾ, Maʾa *-halaʾe* "dry" (E 212, #359: NOm. **kol*- < Om. **x*$_1$*ol*-).

20.3. Om.: (S) **wuč*- id. (E 454, #951: SOm. **woč*- & Yemsa *ičma* ~ *išma* id. < Om. **wiš*$_3$-).

20.4. Om.: (N) Gonga **šu:kk'*- id. (Haberland & Lamberti 1988, 138 compared it with Bayso *sokwat*- "to get thirsty", derivable from ECush. **s-ugg*- "dry, thirsty", and Bilin *səkʿʷana* "thirst", Awngi *canqúna* id.; Appleyard in CDA 135 reconstructs **t-saqʷən*-, compatible with Awni *soγen* "wet").

21. **ear**

21.1. Om.: (N) Ometo **h*$_2$*ayts*-(E 524, 1015) = #46.1. "leaf".

21.2. Om.: (S) **k'a:m*- id., **k'a:m-s*- "to hear" (E 246, #447) ||| or > Cush.: (E) Dullay **qaam*- "ear" (AMS) ||| ?Chad.: (W) **k/kʿūma* "ear" (Stolbova 1987, 209; JI_2, 114, 184). At least some of the Chadic forms can be derived from **kV-sum*-, cf. Kera *kósɔ́ŋ* "ear" (JI_2, 115).

22. **earth**

22.1. Om.: (N) **gad*- > Ometo **gad*- id.; Mocha *gadó* "clod of earth" (Leslau 1959, 30) ||| Cush.: Dahalo *gùδδe* "land, country" || ?(S) Qwadza *gundiku* "bush,

bush country" (Ehret 1980, 239: **gud-*) ||| Berber: (E) Siwa *iždi* "earth" | (S) Ahaggar *éğêdé*, PL. *iğîdân* "dune de sable" (Prasse 1974, 243–44: **ē-gīdiH*, PL. **ī-gīdāHan*), Awlemidden *agidi* "dune de sable vivant", Kel-Ui *ažedi* "sand" | (N) Šilha *igidi & igudi* "sand", Ntifa *igidi*, Mzab-Wargla *ižədi*, Nefusa *židi* "sand", etc. (NZ III, 721–22; E 190, #296; he quotes Chad. **g^{w}id-/*g^{w}ud-* "place" without any examples).

22.2. Om.: **tu:r-* > (N) Dizoid **tu:r-* id., cf. Bench *tor* "down" || (S) Karo *tore* "earth" ||| Cush.: (E) Yaaku *tirri* "earth" || (S) Iraqw *têeri* "dust" ||| Chad.: (W) Dera *turo*, Karekare *tarà* "farm, fields" (Kraft) ||| Eg. (Pyr.) *t3* "earth, land, ground", Copt. *to* (Wb. V, 212; D 57: Nayi + Yaaku + Iraqw; E 144, #178: NOm. + Cush. + Eg. + Chad. < AA **ter-/*tor-*).

22.3. Om.: (N) Gamu, Oyda *sa'a*, Ganjule *sɛxa:*, Koyra *saha* (Lamberti 1993a, 384 adds Wolaita *sa'a* "Land", Kachama *saaha* id., Zayse *saka* "Fussboden, Staub", Dache *saka* "Fussboden"); ?Gonga **šaaw-/*šaww-* id. (but cf. #71.5.) ||| Cush.: (C) Kemant *sēḫā* "prairie" (Conti Rossini) || (S) Qwadza *sa'amuko* "earth" ||| Chad. (C) Hwona *čəḫwēra*; Logone *sχē* "Acker, Feld", Buduma *čúi* id. (Lukas) ||| Eg. Pyr. *sḫ.t* "Feld, Gefilde" (Wb. IV, 229–31) ||| Sem.: Akk. *saḫḫu* "meadow, productive meadow land"; Arab. *saḫwā'* "plain with sweet soil" (EDE I, 196: Kemant + CChad. + Eg. + Sem.).

22.4. Om.: (N) Zayse *'ala*, Kachama *'alla*; Dizi *yellu* id. || (S) Dime F *hilə*, B *yi:l* id. ||| Cush.: (E) Somali *hal* "Ort, Platz" (Reinisch) and/or HECush. **ull-a* id. (Hudson 1989, 423) ||| Chad.: (W) Angas, Ankwa, Sura *yil*; Karekare *'eli*, Tangale *yellì*, Pero *illìy*, Galambu *yíl*; Guruntum *yil* "earth" (Stolbova 1987, 240: WChad. **'/yaŝi*) ||| Eg. OK *i3.t* "Ort, Stätte" (Wb. I, 26) ||| Berb.: (N) Mzab *al* "place" | (W) Zenaga *al*(*l*) "place, pays" (D 195; EDE I, 78: Om. + HECush. + WChad. + Eg. + Berb.).

22.5. Om.: (N) Bworo *gaawa*; Dizi *gob* id. ||| Cush.: (N) Beja *gwāb* "broad open flat ground with little or no vegetation" (Roper) || (E) Dullay **gabb-e* "earth" || Dahalo *gúβa* "plains" (Ehret 1980, 238) ||| Eg. (Pyr.) *Gbb* "Erdgott" (Wb. V, 164) ||| Sem.: Akk. *gabību* "sorte de terrede pâture", Arab. *ğabūb* "sol, terrain, terre friable" (Cohen 1970f, 94).

22.6. Om.: (N) Ometo **biitta* is probably borrowed from Oromo *bii-ti*, subj. case of *biyy-a* "earth". Cf. CCush. **bət-a*, discussed by Appleyard (CDA 59).

23. **eat**

23.1. Om.: (N) **mV'-* id. ||| Cush.: (N) Beja *'aam* "to eat, devour" (Roper) || Dahalo *ᶜaameemit-* "to eat continuously" (Tosco) || (S) Qwadza *am-* "to chew" ||| Chad.: (W) Pa'a *'m̩ma* "to eat" || (E) Sumray *'ə̀m*, Sokoro *ə̀ymɛ̀*, Dangla *ɛ̄mɛ*, Mokilko *'óòmí*, Migama *'âymé* id. (JI$_2$, 119, 121) ||| Eg (OK) *ᶜm* "verschlucken" (Wb. I, 183; Takács, *Afrikanistische Arbeitspapiere* 61, 2000, 95: N+SCush + Chad. + Eg.). Ehret (E 314, #606) compared NOm. with Afar *maaᶜo* "food", *maaᶜiyo* "canine tooth"+ Yakku "mouth" + Chad. "mouth". Together with the Berber counterpart

they represent the exclusive isogloss "mouth": Cush.: (E) Yaaku *mɛ:'* "mouth" ||| Chad.: (W) Ngizim *miya*, Bade *mnya* | (C) Tera *me*; Gude, Bata *ma*; Mafa *mà*, Mofu *méi*, Gisiga *me*; Musgoy *mạ*, Daba *mà*, Kola *mâ*; Gidar *mọ*; Musgu *mu* "mouth" (JI_2, 244–45) ||| (E) Siwa *ambu*, Augila *am*, Fodjaha *imi*, Ghadames *ami* | Ayr-Awlemidden *imi*, Ahaggar *emi*, PL. *imawän* < **ē-miw*, PL. **ī-miHHāwan* < **ē-miHiw*, PL. **ī-miHiHāwan* (Prasse 1974, 130) | (N) Ntifa, Senhaja, Shenwa, Mzab, Wargla, Nefusa, Kabyle *imi* | (W) Zenaga *immi*, PL. *ammun* (R. Basset; see Militarev p.c.).

24. **egg**

24.1. Om.: (N) **bu[l]bul-* id. || Hamer-Banna *bʊla* id. ||| Cush.: (E) Yaaku *bolboli* id. (E 116, #109: NOm.+Yaaku)

24.2. Om.: (N) Yemsa *kewa*, Anfillo *ke:nno* id. ||| Cush.: (N) Beja *kwáhi*, *kúhi* id. (Roper) || (E) Somali *ugaḥ*, PL. *ugḥan*, Rendille *ukáḥ* (Heine); Arbore PL. *hakkó*; Oromo *okhokaan*, Konso *húkukkaa*; Harso PL. *okáaaḥ-e*, Tsamay *ugaḥ-íte* id. || Dahalo *ḳawe* "eggshell" (Elderkin), or better *ᶜogóhi* "egg" (Tosco) || (S) Iraqw *qanḥi*, Alagwa *qana'i*; Maa *ikokóha* id. (E 247: pCush **kʷan-*) ||| Chad.: (W) Hausa *k'wái* "egg"; Sura *kwɛ̀ɛ* "hen"; Diri *ákyùm* "egg"; Ngizim *àgwái* id. (Stolbova 1987, 214; JI_2, 122; E 247, #449: Cush.+Chad.).

24.3. Om.: (N) Kafa *k'abit'o* & *gabet'o*, Mocha *gabat'o* id. ||| Cush.: (E) Oromo Guji *quupa*, HECush. **k'uupp'-ee* id. (D 82: Gonga+HECush.).

24.4. Om.: (N) Chara *mó:la*; Gimirra *mul*; Nayi *moala*, Dizi *mialgu*, Shako *meaku* || (S) Dime *mɔlu* ||| Cush. (C) Khamir *mílā* "testiculi" (Reinisch) || Saho *mī́lā* "Hoden, testiculi" (Reinisch) ||| Chad.: (C) Hina *milti* "Eier" (Strümpell). Takács (*Rocznik orientalistyczny* 57/2, 2004, 56–57) rejects any relation to Berb. **tā-millāl-t*, PL. **tī-millāl-an* "egg", PL. frequently "testicles", apparently a derivative of **m-l-l* "to be white" (cf. CChad.: Sukur *míłán* "white" – see JI_2, 345). But the same semantic motivation is applicable for the Cushitic & Omotic words, although the primary color name is not attested here.

25. **eye**

25.1. Om. **ʾa:p-*, cf. also Ari *ạ́:f-* "to find" (E 345, #672) ||| Cush.: (E) HECush. **af-* "to know, find, get" (Hudson 1989, 69, 87) ||| Berb.: (N) Shilha, Wargla, Kabyle, Sened, Djerba *af* "trouver" | (E) Ghadames, Sokna, Fogjaha *af*, Awjila *uf*, Siwa *if* id. (ZN III, 513–14) ||| ?Sem.: Eblaite *íb-íb-ba-tum* "pupil of the eye", Ugar. *ᶜpᶜp-m* "eyes", Hebrew *ᶜapᶜappayim* du. "eye-lashes", Arab. *ᶜafā'* "tache blanche dans la prunelle de 'l oeil" (SED I, 19; Fleming 1974, 88: Om. + HECush. + Ugaritic).

26. **fat**

26.1. Om.: (N) Gonga **k'o:tts-*; Seze-Hozo **k'ots-* || (S) Dime *kʰuštʰu* id. ||| Chad. (W) Hausa *kítsèè* id. > Guruntum *kitsè* id. || (C) Muktele *àkwàs*, Logone *kàysá* id. || (E) Kwang *kăysə*, Sokoro *káisi* id. (JI_2, 132–33).

26.2. Om.: (N) Chara *kúwa*; Gimirra **ko*'; Dizoid **kow-*.
26.3. Om.: (N) Gonga **'aww-* id., cf. Yemsa *àw* "great, big" (E 365, #734).
26.4. Om.: (N) Anfillo *še:o* id. could be related with or borrowed from CCush. **saγ^{w}-a* "fat" (CDA 64) > Bilin *sax^{w}a*, Khamtanga *säwa*, Khamir *səwa*, Kemant *saγ^{w}a*, Awngi SLLE *sa:χ^{w}i:*, Kunfäl *säwi* id. (D 111: Anfillo + Khamta).
26.5. Om.: (N) Ometo **mold'-*, besides Male *ma:li*; Mao *málɛ́* id. ||| Chad.: (C) Kilba *mal*, Margi *maèl*; Gidar *méélē*, Musgu *àmel* id. (JI$_2$, 132–33).
26.6. Om: (N) Male *mo:rɛsi* || (S) Hamer *mɔr* id. < Amhara *mora* or ECush. **moor-* id. (Sasse 1982, 147).

27. **feather**
27.1. Om.: (N) Zayse *sɛ:lo* id. || (S) Banna *sIle*, Ari *siľa*, Ubamer *silla* id. ||| Cush.: (E) Gollango (AMS) *sole*, Tsamay *siliti* id.
27.2. Om.: (N) Ometo **kep-* "wing" (but Ometo **k'ep-* "feather") || (S) Ari *ka:f-i* id., Galila (Fleming) *kaaf-i* "feather; bird" (E 199, #324: Ometo & Aroid). Related is apparently NOm. **kaap-* "bird" (see #6.1.). Let us add Sem. **kapp-* "palm, flap of hand or foot" (SED I, 134) ||| Chad.: (W) Sura *cáap* "wing, feather" || (C) Tera *kopax* "wing" (Newman). Ehret (E 199, #324) also compares Sem. **kanap-* "wing" (SED I, 130–31) and 'Cush. **kanp-/kinp-*'. Semitic **kanap-* "wing" has promising cognates in ECush.: Arbore *kunúf* "claw, nail, hoof" (Hayward), Elmolo *kúnuf* "fingernail" (Heine), Dasenech *konof*, PL. *konfu* "hoof; finger" (Tosco), probably also Rendille *góffan* "nail, fingernail, toenail" (Pillinger & Galboran)., while CCush. **kanf/b-/*känf/b-* "wing" is of Ethio-Semitic origin, cf. Geez *kənf* (see CDA 148).
27.3. Om.: (N) Zayse *paange* (Lamberti), Koyra *φaŋgɛ*; Bworo *banga*; Seze *báŋkì*, Hozo *baŋka* id.
27.4. Om.: (N) Chara *kúsna*; Yemsa *koso*; Mocha *g/qošo* id.
27.5. Om.: (N) Ometo **ball-/*baal-* id. < Cush. (E) **baal-* id. (Sasse 1982, 31).
27.6. Om.: (S) Dime *k^{h}ɛlɛ* id. < Cush.: (E) Oromo Guji *koola*, Borana *kola*; Tsamay *kool-o*, Harso-Dobase *hóolo* id. || (N) Beja *keláay* "bird", *kil* "to fly" (Reinisch).

28. **fire**
28.1. Om. (N) **tam-* id., cf. Seze-Hozo **tant-* "red" (E 135, #160).
28.2. Om.: (N) Dizi *al-u* id. ||| Cush.: (C) **'aläl-* > Kemant *aläl* "être en chaleur", Khamir *ielel*, cf. Amhara *allälä* (Conti Rossini) | ?(E) Bayso *elen* "fire" || (S) **a:ɬ-* > Iraqw, Burunge, Alagwa *ahla*, Ma'a *mw-ahla* "fire" (Bender 1975, 163; Ehret 1987, #341: C+E+SCush; E 418, #62: Dizi + SCush.).
28.3. Om.: (S) **luh/'-* id. (E 328, #640: SOm. **now-* + Eg. CT, Med *nwḫ* "verbrannt werden, versengt werden", Wb. I, 224) ||| Cush.: (C) **läx-* "fire" > Bilin *läxa*, Khamtanga *liya*, Awngi *leg* (CDA 68). The other Cush. parallels indicate **-^{c}-*: (N)

Beja *n'e* "fire" || (E) Saho *la^{c}-* "to be hot", Afar *lac^{c}-* "to be warm", *lo^{c}o* "daytime", Saho *la^{c}-* "to be hot", *lelle^{c}* "day" || (S) Iraqw *lô^{c}a* "sun, god". Takács (EDE I, 141) adds Eg. *r^{c}* "sun".

28.4. Om.: (N) Yemsa *geeyà* id. ||| Cush.: (C) **gi-* "to be ripe, cooked"; caus. **gi-s-* "to cook" (CDA 48, 116: C+E+SCush.) || (N) Dasanech *Gíété* (Sasse) = *ɟ'ieti* (Tosco) "fire" || Dahalo *gii'-* "to scald" (Ehret 1980, 238) || (S) coll. **gi'ilu* "embers" (Kiessling & Mous 2003, 116).

29. **fish**

29.1. Om.: (N) Ometo **mol-* id. || (S) Galila *mola*, Ubamer *molta* id. ||| Cush. (E) Somali, Jiddu *mallay*, May *mallaallay*, Bayso *moole*; Mossiya *moole* (Lamberti & Sottile 1997, 461) id., if it is not borrowed from Omotic (E 306, #587: Ometo **mol-* > Macro-Somali **mallay*) ||| Chad.: (W) Mbaru *mwalaŋ*, Guruntum *mollaŋ* (Shimizu) = *molḷaŋ* "fish" (JI$_2$, 140) ||| ?Eg. (Dyn. XVIII Mag.) *mr.t* 'ein Tier (zwischen Fischen genannt)' (Wb. II, 105); Takács (*Linguistica* 37, 1997, 53–54) adds the name of the king *N^{r}r-mr* of the 1st (or "zero") dynasty. The first component depicted as the "catfish" is naturally "catfish". Takács offers to interpret the second component as the "fish".

29.2. Om.: (N) **wa:r-* > Bench *oru*, She *woři*; Kafa *ollo*; Shako *o:rú*, Dizi *urgit* || (S) Dime *'ɔrɔxo* id., cf. SOm. **wa(:)r-* "to swim" (E 463, #975).

29.3. Om.: (N) Kafa *haašoo*, Anfillo, Mocha *ha:šo*; Nayi *aša* id. < Ethio-Semitic: Geez *^{c}āśā*, Tigre *^{c}asa*, Amhara *asa*, plus other areal parallels as Beja *aaši* (Roper) || Bilin *'/^{c}asa*, Kemant *asa*, Awngi *asi* vs. Khamtanga *aza* id. (CDA 68) || Saho *^{c}aasa* id.

29.4. Om.: (S) Hamer *kara*, Banna *ka:ra* id. ||| Cush.: ?(N) Beja **kware* in *ašoob kwaremn̥a* "fisher" (R. Hudson) || (E) Dasanech *káara* id. (Haberland); Dullay **xaar-* id. (AMS) ||| Chad.: (W) Hausa *kíífíí*; Kulere *kirif*; Siri *kerfei* || (C) Nzangi *húrfẹ*; Zime-Batna *kérfé* id. (JI$_2$, 140–41; Stolbova 1996, 62–63: **kirop-*) ||| ?Sem.: Official Aramaic *kwr* "fish", Judeo-Aramaic *kawrā*; Soqotri *kúwerhor* "sp. fish" or *kér* "sp. shark" (SED II, 171).

30. **fly n.**

30.1. Om.: (N) **wuts'in/l-*.

30.2. Om.: (N) **gudunts-*.

30.3. Om.: (N) Yemsa *kurà* id. ||| Cush.: (E) Oromo *kurupp'isa* "Heuschrecke"; Gedeo *koro'nisa* id., Sidamo *koró & koricco*, Kambatta *koruta* "flea", Hadiyya *kora* id. (Lamberti 1993b, 355; Hudson 1989, 65: HECush. **kora* "flea").

30.4. Om.: (N) Gonga **zabunz-*. Maybe an adaptation of Ethio-Semitic forms: Amhara *zəmb*, Harari *zəmbi* id. (Lamberti 1993a, 408).

30.5. Om.: (N) Maoid **š(i)aŋg-i* || (S) Galila *šuŋgula* id.

30.6. Om.: (S) Dime-Hamer **kosuub-* id.

31. **foot/leg**

31.1. Om.: (N) **tuk-* > Male *toki*, Zayse, *túki*, Kachama *túke*, Seze *tʊggÝ* (E 140, #168: **tok-*; B 147: **tok-*), cf. Malo *t^{h}ɔh-* to go, walk" : *t^{h}ɔhɔ* "foot"; Dizoid **teg-* "to go, walk" ||| Cush.: (E) Somali *teg* "to go, walk" (E 140, #168) ||| Chad. (C) Glavda, Gava *takwaya*, Bokwa *tókwoya* "Bein unt. Teil" (Büchner), Muzgu *túgu* "foot" (Lukas).

31.2. Om.: (N) Ometo **gad-* id., cf. Kafa *gidoo* "knee, force" (Cerulli), Kachama *tihee geddo* "sole" (Conti Rossini) ||| Chad. (C) Nzangi *gedaatji* "leg", by Kraft *gɨddʿìgaci* "thigh", Holma *gedgaači* "Bein", Gude *gedehán* "leg" (Strümpell) = *gudʿxan*, Fali Bwagira *gudʿugan* "thigh" (both Kraft) || (E) Sumray *gídam* "foot" (Lukas).

31.3. Om.: (N) Dizoid **ʾaš-*; maybe also Yemsa *očo* id. ||| Cush. (E) Dullay **ʾašš-* "to go"; cf. Burji *acc-* "to start, begin; try" (Sasse 1982, 23) ||| Chad. **asə* "foot/leg", cf. **(a)sə* "to come" (Newman 1977, 29, 24; JI$_2$, 220–21 & 82–83) ||| Berb.: East Numidian *y-s-ʾ* "to come" || (E) Siwa *as əd*, Sokna *yus-əd*, Ghadames *as id* | (N) Kabyle, Mzab, Wargla, etc. *as d* | (W) Zenaga *šed* (R. Basset) | (S) Ahaggar *as d* etc. (Prasse 1973, 122–23: IMPF. **yaHsiH* vs. PERF. posit. **yuHsaH*; Militarev 1991, 255) ||| ?Eg. (MK) *3s* "eilen" (Wb. I, 20) ||| ?Sem. **ʾ-y-š*: Arab. *ʾayisa* "to despair of", Geez *ʾesa* "to waste away" (Leslau 1987, 50; E 364, #731: Dizoid + Dullay + Chad.; Takács, *Lingua Posnaniensis* 38, 1996, 48: ECush. + Chad. Berb. + Eg. + Geez.).

31.4. Om.: (N) Shako *ša:nu* "foot" (if it is not an extension of the preceding form) ||| Cush.: (E) **saʾn-* "footprint, heel" (Sasse 1979, 52). A borrowing is not excluded too. Cf. also Eg. (Dyn. 20) *sn.wj* "zwei Füsse (alles, was geht auf zwei und vier Füssen)" (Wb. IV, 148), if it is not derived from the numeral *sn.wj* "two" (ibid.).

31.5. Om.: (N) Mocha *gano* "foot / leg" ||| Chad.: (C) Bata *gwone* "leg" (Mouchet) || (E) Nancere *guánung*, Dormo *dé-ganu* "Bein" (Lukas).

31.6. Om.: (N) Seze *dàkká* id. ||| Cush.: Dahalo *dakaᶜa* id. || (S) Iraqw *deqᶜ* "joint".

31.7. Om.: (S) **du*[*r*]*t-* ||| Chad. (E) Sumray *dùrí* "thigh", Ndam *dí:*, Tumak *gun dùrò* id. (JI$_2$, 325; E 135, #151; E 442, #926 reconstructs SOm. **dowt-*).

31.8. Om.: (N) Anfillo *tʿammo*; Mao *čɛ́:mɛ́* id. < Amhara *čʿamma* "foot; shoe" < CCush. **canb-*/**canf-*(CDA 122).

31.9. Om.: (N) Kafa *batto* id. < Amhara *bat* "calf of leg"; Geez *baᶜāt* id. was probably reconstructed from the Amhara word (Leslau 1987, 85).

32. **full**

32.1. Om.: (N) Ometo **kum-* id., besides Ometo **kum-*; Gonga **hum-* "1000" (D 78–79) ||| Cush.: (E) Somali *koon*, PL. *koomo* "multitude, a lot of", Boni **komoom* "all" (Heine), if they are not borrowed from Arab. *kūm* "heap"; Bayso *kamogani* "much, many" (Bender); Dullay **kum-* > Tsamay *kum-*, Gollango *hum-* "to finish" (AMS); besides ECush. **kum-* "1000" (Sasse 1979, 25) || (S) **kuma*, PL. **kumamee*

"1000" (Kiessling & Mous 2003, 181) ||| Eg. (MK) *km* "vollständig machen" (Wb. V, 128) ||| Sem.: Arab. *kamma* "to assemble" (E 198, #322: Ometo & Yemsa *kùmá* "hill"+SCush.+Arab. *kamma*; Takács 1996, 141: Ometo "1000"+E+SCush.+Eg.+CChad.: Ga'anda, Boka *kùm*, Chibak *kuma* "10", although their Bantu origin is more than probable).

32.2. Om.: (N) **č'e:n-/*č'o:n-/*č'u:n-* > Mocha *č'e:nno* id., Bench *c'onc'-* "to fill", Seze *ts'ó:nì/ts'unewi* (H 270, 273; he adds Yemsa *tuunu* "full").

32.3. Om.: (N) **č'o(o)č'-* > Chara *s'os'á*; Dizoid **s'os'-* || (S)**č'o(o)č'-*. These forms could reflect the root **č'o(:)n-* plus *-s/š-*, cf. Kafa *č'een(n)-* "to become full" vs. *č'eečč-* "to fill" (tr.) < **č'een-s/š-*(H 282).

33. **give**

33.1. Om.**'im-(*H 282) ||| Eg. (OK) *ỉmj* "gib!, setze, lege!", Copt. *ma-*(Wb. I, 76; E 349, #683: Om. + Eg.).

33.2. Om.: (N) Dizoid **'ats-* "to give" ||| Cush.: (E) Kambatta *aass-* "to give", if it is not borrowed from Omotic; further Somali *sii* "to give", Boni & Rendille *síi* id. (Heine); Arbore *sihis-*, Elmolo *sííse*, Dasanech *šii-š* id.; Yaaku *-isɛ'ɛ* id. (Heine) ||| ?Berb.: (S) Ahaggar *aws* "payer un impôt" (Foucauld) ||| Eg. (old) *ỉsw* "compensation, salary, reward", Copt. *asu* "price" (Wb. I, 131) ||| Sem.: ?Ugaritic *šy*, Hebrew *šay* "gift"; cf. Arab. *s-w-y* "to cost" (E 356, #702: Dizoid **ats'-* + Eg.; Takács, *Discussions in Egyptology* 38, 1997, 97: ECush.+Berb.+Eg.+Sem.).

33.3. Om.: (N) Seze-Hozo **neen-* id. ||| Cush.: (N) Beja *nuun* id. (Roper). Cf. CCush. **nay-* "to give here" < **naakk-*(CDA 74).

34. **good**

34.1. Om. **lak'-* > (N) Ometo **lo*[*k'*]- > Wolaita-Kullo *lo'-uwa*, Gofa, Zala, etc. *lo'o*, Gidicho *lQ:kP* || (S) Ari *laya-mi*, Ubamer *lay-mi* id. (E 397, #806: NOm. **lo'-*) ||| Berb.: (S) Ghat *iulay-ən* id., Ahaggar *aləy* "to be good" < **w-l-y* (Militarev 1991, 262).

34.2. Om.: (N) Bworo *še:ŋga* id. ||| Cush. (C) Awngi *cənkút* id. || ?(E) Somali *san* "to become good"; Oromo *šagga'* "good"; Yaaku *-šɛn*, PL. *-šeme* "sweet" (Lamberti 1993a, 377: Bworo + Oromo; E 280, 519: Bworo + Awngi + Somali+Yaaku).

34.3. Om.: (N) Basketo-Dokka-Male **koš(š)-* id. ||| Cush.: (S) Ma'a *kuhlo* id.

34.4. Om.: (N) Gonga **do'-*. Cerulli (1951, 423) saw a source in Shilluk *dooc* "good".

35. **green**

35.1. Om.: **č'il-* > (N) **č'il-(~ *č'al-)* id. || (S) Dime *c'Ilo* id., cf. Ari *č'el-* "raw" (H 266, 270) ||| Cush.: (E) Oromo *t'ilo* "green" (Tutschek) || (S) Qwadza *tsalam-* id.

35.2. Om.: (N) Kafa *gare* id. ||| Cush.: (E) Somali *ʿagaar* "green"; Hadiya *ajaara* "yellow". Cf. also CCush: Kemant *arägʷina* "green", vs. Amhara *aräng*^{w}*ade* id. (CDA 77).

36. **hair**

36.1. Om.: (N) Dizoid **tsic-* id.; Mao *tis'i* id. || (S) **sit'-i* id. ||| Chad. (C) Bura, Chibak, Ngwaxi, Wamdiu *šiši*, Margi *šÄši*, West Margi *cici*, *citi* id. (all Kraft). There are interesting NS parallels: Kuliak: Ik *sitsʿ*ᵃ, So *sɨj-at* "hair" (Lamberti 1998, 114: Om. + Kuliak).

36.2. Om.: (N) Yemsa *soma* id.; She *som* "hair, head" (Montadon) ||| Cush.: (E) Afar *sammo* "pubic region"; Burji *šómi* "pubic hair" (Sasse 1982, 174) || (S) **seʾem-* "hair" (Ehret 1980, 350) ||| Chad.: (W) Hausa *suma* "growth of hair", Ngamo *sṓṏm* "hair" (Kraft) || (C) Baza *simthífi* (Lukas), Fali Jilbu *sɨmcɨn*, Fali Mucela *šïmki*, Banana *simità* (all Kraft), Sukur *šúmbùt* "hair" ||| Eg. (Pyr) *sm3* "hair" (Wb. IV, 122) ||| ?Sem. **š-m-y* > Mehri *šdmɜy* "fine hair shed by camel".

36.3. Om: (N) Ometo **binaan-/*binnan-* id. || (S) Dime *bande* id. ||| ?Sem.: Akk. *bibēnu* "temple, side of head".

36.4. Om.: (N) Chara *kúsna*ʾ; She *kust*, Bench *kʊ:st* id. (E 215, #365: NOm. **ku:ts-* < *Om. **x_1-*).

36.5. Om.: (N) Dizi *sar* id. ||| Eg. (MK, Med) *sr* id. (Wb. IV, 191).

36.6. Om.: (N) Male *gami* in *to:ki gami* "hair" (*to:ki* = "head"), cf. Basketo-Dokka *gəmma* "skin" || (S) Ari *gʌmma* "skin" ||| Cush. (C) Kemant *gama*, Khamir *gami* "mane" || (E) **gamm-* "mane, stuff" (Sasse 1982, 77) > Amhara *gamma*, Harari *gaamma* "mane" (Leslau 1963, 72) ||| Chad. (W) **gVmV* "beard, chin" > Hausa *geemu* "beard" (Stolbova 1987, 217) ||| Eg. (CT) *gm.t* "Locke oder Flechte; Schläfe" (Wb. V, 171) ||| Sem.: Arab. *ǧummat/ǧammat* "big lock of hair", Tigre *gəmmat* "long hair" (SED I, 75–76; D 216–17).

36.7. Om.: (N) Gonga **tsʿigar-* id. < Ethio-Semitic: Geez *ṣagʷr*, Tigre *č̣əgär*, Tigrinya *ṣägʷri*, Amhara *č̣ägär*, Gurage *č̣əgär*, *ṭəgär* id.; hence Qwara *tʿagur*, Bilin *šagar* id. (Leslau 1987, 550); but cf. ECush.: Afar *daguru*, Saho *dagar*, Somali Re *dʿagur*, Ab *dʿogor*, Boni *d'ȍȍr*, Bayso *ogorro* (D 119: Gonga + ECush. + Ethio-Semitic).

37. **hand**

37.1. Om.: (N) **kuč-(*H 283; E 208, #347: NOm. **kuš$_2$-/*kiš$_2$-* + W+CChad. **k-s* "to take" + Sem. **k-c* "to grasp"). Perhaps better Cush.: (C) **käs-* "shoulder" (CDA 122) || ?(E) Gawada *xašito* id.; Burji *káččoo* id. ||| Chad.: (W) Ankwe *kɨθšat* "shoulder" (Kraft) || (C) Gulfei *geŝe* "(fore-)arm" || (E) Kera *kasi* (Ebert), Kwang *kósí* "shoulder", Kabalai *kabi kaasi* "arm" : *kobí* "arm" (Lukas).

37.2. Om.: (S) **ʾa:n-i* id. Fleming (1976, 318) saw cognates in Bworo *aan* "to give back" and Dizi *hani* "to put, place". Later Fleming (1983, 447) and Lamberti (1988, 114) mentioned the Nilo-Saharan parallels, esp. Kuliak **an* "hand(s)".

38. **head**

38.1. Om.: (N) Ometo **kʿumm-* id. || (S) Ari *xüma* "top of head" (Bender) ||| Cush.: (C) Bilin *qʷəmiš* "face, cheek" (Reinisch) || ?(E) Oromo *qoma* "chest" ||| Chad.: (W)

Hausa *kumci* "face, cheek", North Bauci **ḥa-kʿum-* "cheek" (Skinner 1977, 15) || (C) Chibak *kumà* "face", *kumi* "cheeks", Fali Mucella & Fali Gili *kuma* "face", Higi Nkafa *kuma* id. (all Kraft) || Sem.: Arab. *qimma* "top of head".

38.2. Om.: (N) **kʿell-/*kʿe:l-* > Bench *qe:l*; Kafa *qello:* (both Cerulli), Mocha *kʿällo* id. (Leslau 1959, 46), Bworo *kʿellá* "face" (Lamberti) ||| Chad.: (W) Hausa *kʿoli* "top, summit", *kʿwalluwa* "skull", *kʿwakʿwalwaa* "brain", Sura *kwnlbiyap* (Kraft) Geji *kulɨ-gɨ* id. : *gɨ* "head" (Shimizu) || (C) Logone *kúlum* "forehead" (Lukas), Lame *kwukwálwà* "brain" (Kraft) || (E) Kera *kə́kə́l* "brain" (Ebert) ||| Berber: (N) Sus *aqəllal* "head, jug", Beni Snus *aqelqūl* "head, skull" (Destaing) ||| Sem.: Arab. *qulla* "(top of) head".

38.3. Om.: (N) **to:k-/*tokk-* > Male *to:ki*; Chara *to:yá*; She *to*; Bworo *toka*, Anfillo *to:kko* id., Kafa *tukkoo* "upper side", Mocha *toʾo* "hair" (Lamberti); Maoid **to(:)k-i* "head" ||| ?Cush.: (E) Konso *tuq-tá* "back part of neck"; Yaaku *tocʿono* "neck" (Lamberti) ||| Chad. (W) pAngas **tVwVk/*tVyVk* "neck, nape" (Stolbova 1987, 241); Dira *tax gàmi* "hair" : *gàmi* "head", Buli *tɨk* "hair" (all Kraft) ||| Sem.: Akk. *tikku*, *tíku* "neck".

38.4. Om.: (N) Koyra *kʿínne*, Kachama *kʿino* id. (Conti Rossini) ||| Chad.: (W) Angas *kiin* "front" (Foulkes); Kirfi *kún-tí-kà* "head" (Schuh) || (C) Margi-Gwara *kən*, Vizik *xan*, Lamang *ghan*(*a*) id. (all Wolff), Fali Kiria *yùnù* id. (Kraft) || (E) Sokoro *góndu*, Gabri *cain* id. (Lukas) ||| Sem.: Arab. *qunna* "top" (Müller 1975, 64: Koyra + Chad.).

38.5. Bench *deb* id. ||| ?Cush.: (E) **dabb-* "hair" > Oromo Maca *dabbasa* "hair" ||| Chad.: (W) Angas *tàbùr* (Kraft), Sura (Jungraithmayr) *tə̀bùk* "brain" (*t-* < Chad./AA **d-*, see Stolbova 1987, 41) || (C) Tera *dəbʿ äɟin* "skull" : *ɟin* "head" (Newman) ||| Eg. (NK) *db.t* "head", late *dbn* id., *dbn.t* "Haarflechte" (Wb. V, 434, 437, 438) ||| ?Sem.: Arab. *dabba*ʾ "hairy", *dabab*, *dubb*, *dababān* "duvet, poil aux joues" (Cohen 1970f, 205; E 125, #120: Bench + Cush. + Eg . *dbn.t* "Haarflechte" + Arab.).

38.6. Om.: (N) Dizi *geli* = Ce *gaylli*, Shako *gärí* = Ce *gari* id. ||| Chad.: (C) Muktele *gə̀l* "head"; Musgu *gaˊley*, Gidar *gaḷi*, Masa *gályaḷ* "joue" (Mouchet) || (E) Kwang *gol* "head", Sokoro *geltim* "brain" (Lukas) ||| (E) Augila *t*(*e*)*-gili* "head" (NZ III, 771) ||| Eg. (OK) *d̲3d̲3* "head" (Wb. V, 530) ||| Sem. **gulgul-*(*at-*) ~ **galgal-*(*at-*) "skull" (SED I, 74; Takács, *General Linguistics* 36, 1998, 154: Dizoid + Chad. + Eg. + Sem.).

38.7. Om.: (N) Ometo **ʾu:mm-a* id. > HECush.: Sidamo *umo* "head", Gedeo *umo* "hair" (Hudson 1989, 77) ||| Cush.: (N) Beja *huum*, PL. *hum/him* "Mark, Gehirn, Hirn"; *hámo* "Kopfhaar" (Reinisch) || (E) Saho *amoo* "head", Afar *amo* "head, summit, top, hair, intelligence"; HECush. **iima* "top" (Hudson 1989, 156) ||| Chad.: (W) NBauchi: Tsagu *ààmài*, Paʾa *ʾàm-á*, *hama*, Diri *àmá*$^{(h)}$, besides Wariya *yám-áy*, Kariya & Miya *yàm*, Jimbin *gàmá* "head" (Stolbova 1987, 226 reconstructs WChad. ** ḥama*) || (C) Masa *yám*, Banana *yambà* "head" (JI$_2$, 182–83) ||| Berb.: (S) Ahaggar *imme*, PL. *immewän*, augmentative of *timme*, PL. *timmawîn* "front"

< **Himmīw*, PL. **Himmīw-an* & *t-Himmīw-t*, PL. **t-Himmāw-īn* (Prasse 1974, 171–72) || (N) Babor, Takitunt, Mentano, Kalawn *timmi* "front", Ntifa, Zayan, Menacer *timmi*, Semlal *timiut*, Warain PL.? *timmau*, Shenwa *hamiu* "sourcil" (A. Basset, *Études de géographie linguistique en Kabylie*, Paris 1929, 22, 59) ||| Sem.: Arab. *hāmat*, PL. *hāmāt & hām* "head; chief" (Steingass 1988, 1161; Takács, *Lingua Posnaniensis* 38, 1996, 54: NOm.+N+ECush.+W+CChad.+ Berb.+ Arab.+ Eg. Gr. *hmhm* "Name der Krone", Wb. II, 491).

38.8. Om.: (N) Ometo **hu:p'-* id. (E 384, #785: **$h_1$up'-*) ||| ?Eg. (NK) *ḫf.t* "Gesicht" (Wb. III, 274).

38.9. Om.: (N) Nayi *mot*, Shako F *mo:tu*, *mu:tu* id. || (S) **mat-a* id. ||| or < Cush.: (N) Beja *mat* "top of head" (Reinisch) || (E) **matḥ-* "head" (Sasse 1979, 5, 8, 10, 36, 37, 53, 57) ||| Chad.: (C) Vulum *măt* id. (Tourneux), Muzgu *mada*, Padukwo *mudạra* id. (both Mouchet) ||| ?Sem.: Akk. *muttu* "front, forehead".

38.10. Om.: (N) Yemsa *teeta id.* < **ts'eer-* "point, tip" (H 270).

39. **hear**

39.1. Om.: (N) **siy-* > NOm. **siy-* || (S) Hamer *sI:k-(*H 265).

39.2. Om.: (N) Male *waiz-*, Yemsa *weeso*; Kafa *wayo* id., cf. Ometo **$h_2$aits-* "ear" ||| (C) **was-* > Bilin, Khamir, Kemant *was-*, Khamtanga *wâšu* id. (CDA 82 & D 187–88: NOm. + CCush.).

39.3. Om.: (N) Mocha *qäbbi-*; Mao *k'ə́wá*, Seze *k'^{w}a:k'^{w}a* id. ||| Cush.: (E) **k'ab-* > Dullay **qabaq-* id.; Burji *akkab-* id. (Sasse 1982, 24). Mukarovsky (1981, 205) added Mussiye *koporo* "ear " (SLLE), but *k-* < **g-*.

39.4. Om.: (S) Dime *k'ams-*, Hamer *k'ans-* id., cf. SOm. **k'a(:)m-i* "ear" ||| or > Cush.: (E) Dullay **qaam-* "ear" (AMS) ||| Chad. (W) Tangale *kụmẹ*, Bolewa Bolewa *kùmé*; Kir, Buli *kum*, Zaar *ka:m*; Ngizim *kūmū* id. (JI$_2$, 184).

39.5. Om.: (N) Yemsa *odo* id. ||| Eg. (Old) *ỉdn* "ear"; MK lit., Med., BD *ỉdj* "to be deaf" (Wb. I, 154; 151) ||| Sem. **'u$\underline{d}$n-* "ear" (SED I, 6).

40. **heart**

40.1. Om.: (N) Basketo-Dokka **bu(:)d-a* id. || (S) Dime *budə*, Galila *bu:da* id. ||| Chad.: (C) Gudu Kr *móbùd* id., Daba *məbbùra* "chest" (Kraft), Uzam *mạbruv*, Mada *burov* "heart" (both Mouchet) ||| ?Berb.: (S) Ahaggar *ăbâda*, PL. *ibâdân* "sein; pied des pentes" (Prasse 1974, 315: **ā-bādāH*, PL. **ī-bādāHan*), Ayr-Awlemidden *ăbada* id. (NZ I, 17).

40.2. Om.: (N) Male *'ina* id.; Chara *'í:na = hi:na* (Fleming) = *yina:* "belly" (Cerulli); Nayi *yi:nu* id. ||| Eg. (Dyn. 22) *hn* "belly" (Wb. II, 492).

40.3. Om.: (N) Ometo **muts'ur-/*muts'ir-/*muts'ar-* id., cf. also Ganjule *məs'uro* "lungs" (Fleming) ⟨or⟩ Cush.: (E) Dirayta *mút'uraa*, Mashile *musura*, Musiye *musuro* "heart", Oromo *mač'ure* "bowels, intestines", borrowed into Ethio-Semitic: Harari, Gurage *märäč'i* id.

40.4. Om.: (N) She *še:n* "belly" (Conti Rossini); Dizi *čwonu* (Bender), Shako *ʂon*, Nayi *šunus* "heart"; Seze *ši:nɪ* id. (Bender), Mao *šin* "belly" (Fleming) || (S) Dime *ši:ne* "belly", maybe borrowed from NS, cf. Surma: Meqan *šɪni* "belly" (Fleming 1965, 649). There are also interesting Semitic parallels: Akk. *sūnu* "lap, crotch"; Arab. *ṯunnat* "bas-ventre, abdomen (entre le nombril et le pubis)"; Harari *šān* "groin" (SED II, 334: **ṯVn(n)-*).

40.5. Om.: (N) Ometo **wazan-* < ECush. **wazn-*(Sasse 1979, 20, 58; Id. 1982, 187; E 468, #985: Ometo **wizn-* < ECush.) || (C) Bilin Re *wådän*, PL. *wåzän* "belly, heart", Khamir *əzän*, Kailiña *əzan* "heart" (CDA 82 & D 187, 295: Om. + C + ECush.) ||| Chad.: (C) Daba *wènji* "intestines" (Kraft) || (E) Sumray *wódom* "heart" (Lukas).

40.6. Om.: (N) Gonga **nibb-*, Seze-Hozo **nibb-a* id. < Ethio-Semitic: Geez, Tigre, Amhara, Gurage *ləbb* "heart", Tigrinya *ləbbi* id. (Leslau 1959, 43; 1987, 305).

40.7. Om.: (N) Male *saza* id., cf. Shako *s'ezga* "breast" || (S) Ari *saza*, Banna *sada* "heart" < Cush.: (E) **saz^{ʿ}-* "heart" (Sasse 1976, 138).

41. **horn**

41.1. Om.**k'uš-(im/b-)* > (N) Basketo *ʊšɪnš*, Zayse *uššúme*, Dizi *ušum*, Shako *uʂń*, Mao *'i:šìmɛ̀* id. || (S) Dime *' ʊšʊm*, Hamer-Banna *k'ušu(m)ba*, Ari *k'ošma* id. (E 355, #700: Om. **uš$_1$(um)-*) ||| Chad. (C) Logone *káāŝú* id. (Lukas) ||| Berb.: (N) Matmata *qiš* (Destaing), Harawa *kīššu* (R. Basset) id.

41.2. Om.: (S) Galila *ba:li*, Dime *baltu* id. (Fleming) ||| Chad.: (W) Montol Jg *bulu* (Jungraithmayr), Gerka *b'ɛl*; Tangale *bọl*, Kirfi *balla*, Bolewa *b'alum* id. (JI$_2$, 192).

41.3. Om.: Shako *fera*, maybe also Bench *pel*, She *fäl* id. (Fleming) ||| Chad. (W) Kofyar *feer*; Warji *farai/párái*, Karya *pár*, Miya *àpár* || (C) Daba *fàlām* id. (JI$_2$, 192–93).

41.4. Om.: (N) Gonga **k'ar-*; Seze-Hozo **k'ar-* id. are probably borrowed from Ethio-Semitic: Gurage, Harari, Tigre *qär*, Tigrinya *qärni*, Geez *qarn* id. (Leslau 1959, 48; 1987, 442). Ehret compares it with Om. (S) ?Ari *k'ari* "tusk", *k'ar-mi* "sharp" ||| Cush.: (E) **k'ar-* > Somali *qar* "hill higher than *kur*", Oromo *qar-r-ee* "peak", cf. also Konso *qaar-ta*, Dirayta *k'ar-a*; Hadiyya *k'aar-a* "sharpness" (Sasse 1979, 48–49) ||| Sem. **qarn-* "horn" (SED I, 151; E 238, #424: Gonga & Ari + Cush. + Sem.).

42. **I**

42.1. Dizoid **'in-* ||| Cush. **'an-a, -i, -u* > (N) Beja *aní* (Reinisch) || (C) **'an/*'än* (CDA 87) || (E) **'ani/*'anu* (Sasse 1982, 26) || Dahalo *'anyi* (Ehret), *ana/ani* (Damman) || (S) **'ana* (Kiessling & Mous 2003, 56) ||| Chad. **['a]n-a, -i* > (W) Hausa **ni-i*; Sura-Gerka **-'an*; Ron **yin* || (C) Higi **'yina*; Bata **Hun[i]*; Gidar *na*; Musgu obj. **ana*; Kotoko obj. **-na/*-ni*; Masa **n-anu*, obj. *-an* || (E) Kwang-Kera **(-)Vn*; Sumray obj. **an*; Sokoro *na*; Jegu *nóò* (Blažek 1995, 40) ||| Sem. **'an-ī, -ā* (Blažek 1995, 46–47; E 362, #724: Dizoid + Cush. + Chad. + Sem.).

42.2. Om.: (N) Ometo-Gimirra-Yemsa-Gonga abs. **ta*, subj. **ta-ni*, obj. **ta-na*; Mao *tíjá*, Ganza *ti*. Zaborski (2004, 181–82) discusses the Nilo-Saharan parallels: Teda-Daza

ta; Majang *eet*. On the other hand, Dolgopolsky proposed another solution, seeing here the derivative of the verb of existence of the type **t-*, attested e.g., in Oromo *ta*'-, Sidamo *-te*, Hadiyya *et-* (D 20–21).

42.3. Om.: (S) **'inta*. Zaborski (2004, 181) sees here the structure analogous as in Olam (Surmic) *an-ee-ta* "I" : *inee-ta* "thou" and Ragreg (North Nilotic) *anyit* "I". Dolgopolsky derives **'inta* from the construction **'VnI tVH-'V* "I being-I".

43. **kill**

43.1. Om.: (N) **wod'-* id. (E 465, #979).

44. **knee**

44.1. Om.: (N) **bu(n)k'-* id., cf. Shako *bòka* "thigh" || (S) **buk'-* id. ||| Chad.: (W) Ron: Fyer *fúŋ*, Daffo *mbɔ̀ŋ* id., Sha *'abwóŋ* "shin-bone" (all Jungraithmayr) || (C) Bata Garwa *bongtongče:* "elbow" (Strümpell); Bura *b'unji*, Chibak *b'wunài*, Ngwaxi *wɨ́ngì*; ?Banana *fwökìvà* "knee" (all Kraft).

44.2. Om.: Chara *k'uma*; Bench *k'um*; Shako *k'umu*; Hozo *k'ʊʊmɪ*, Mao *túl-k'ùmɛ̀* id.

44.3. Om.: (N) **gulbat-* < Ethio-Semitic: Amhara *gulbät*, Chaha *g^{w}urbät*, Gyeto *g^{w}ərmät* id. (Leslau 1963, 71) < Cush.: (E) **gulb-(*Sasse 1979, 5, 6, 18; E 191, #301).

44.4. Om.: (N) Zala *iquorsa* ||| Cush.: (E) Arbore *kuršat*, Dasanech *kúr*, PL. *kurram* id., Rendille *kursánte* "knee-cap".

45. **know**

45.1. Om.: (N) **'ar-* > Ometo **'er(r)-*; Chara *ár-*; Gimira **'er-*; Gonga **'ar(r)i-*; Hozo *arɛti*, Seze, Mao *'àl-* id. ||| Cush.: (E) Dullay**'ar-* id. || (S) **'ar-* "to see" (Ehret 1980, 286; E 364, #729).

45.2. Om.: (N) Dizoid **t'us-* id. || (S) **d'es-* id. (Bender 1975, 171: N+SOm.)..

46. **leaf**

46.1. Om.: (N) **h$_2$ayts-* = "ear" || ?(S) Galila *haca*, Hamer *aca* "leaf" ||| Cush.: (C) **xaš-* "leaf" (CDA 91) || (E) Dullay: Konso *haša* id.; Tsamay H *áaš-e* id.; Gedeo *hoccó* "ensete leaf" ||| Sem. **ḫawṣ-* "leaf" > Epigraphic Aramaic *ḫwṣn* "palm leaves", Aramaic *ḥūṣā* "palm leaf", Syriac *ḥ^{e}wåṣe* > Arabic *ḫūṣ* id., Arabic of Hadramaut *ḫawṣ* "leaf", Soqotri *ḫeṣ* "leaf" (Leslau 1938, 168; D 157: Om. + C+ECush. + Sem.).

46.2. Om.: (N) Gimirra **kar-*; Shako *kə̀ra* id.

46.3. Om.: (S) **k'alb'-* id. Mukarovsky (1981, 207) compared it with NS: Saharan: Kanuri-Kanembu *kálu*, Tibbu *kolú*, Teda-Daza *kolu* "leaf".

47. **lie**

47.1. Om.: (N) Ometo **zin'-* "to lie, sleep", cf. Mocha (Leslau) *yɪ:n-* "to spend the day" (E 151, #199: NOm. + ?Sem.: Arab. *zan'a* "to stick to a place; adhere to the ground").

47.2. Om.: (N) Yemsa *kuna* id. ||| Sem. **k-w-n* > Akk. *kânu* "to be firmly in place"; Ugaritic *kn* "to be", Hebrew *k-w-n* to be firmly established"; Syriac *kān* "to be, exist"; Arabic *kāna* "to be, exist"; Sabaic *kwn* "to be, occur"; Mehri *kān* "to be";

Geez *kona* "to be, become, exist, happen, occur, take place, there is" (Leslau 1987, 299–300). Ehret (E 220, #377) compares it with (S) Ari *gin*'- "to fall asleep" ||| Cush. (E) **g'iin-/*g'aan-* "to stay in place" ||| Chad. **xʷən-* "to lie down" (Newman 1977, 29) ||| Sem.: Arab. *yanana* "to stay, remain; be at a place".

48. **liver**

48.1. Om.: (N) Ometo **mayz-*; She *mai* id., Bench *may* "heart"; Hozo *mɛ̀'i*, Seze *bèì* "liver" ||| Eg. *mjz.t* "liver" (Wb. II, 44).

48.2. Om.: (N) Gonga **'apa:r-* id. ||| Cush.: (E) Sidamo *affale*, Burji *affála* id. (Sasse 1982, 24), cf. Afar *afᶜado*, Saho Re *afᶜadoo* "heart" || ?(S) Iraqw *afetlo* "waist", Asa *'ofolunk* "navel", Qwadza *ofu'uko* id. (Ehret 1980, 377) ||| Chad.: (W) Ankwe *pàlàk*, Montol *filìk* id. || (C) Hwona *kub'ar-fara* "heart", Fali Bwagira *farɨn* "liver" (Kraft), Gisiga *me-vel* id. (Lukas) ||| ?Sem.: Soqotri *ᶜefel-oti* "stomach" (Leslau 1945, 241 compared it with Arab. *ᶜafala* "to fat on the thigh of the animal").

48.3. Om.: (N) Ometo **tir-* || (S) **tur-* id. ||| or < Cush.: (E) **tir-* id. (D 203) || (S) Ma'a *itiráo* id. (Ehret 1980, 225) ||| Chad.: (W) **ḷa-tirsa* "kidney" (Stolbova 1987, 166; E 144, #179: Ometo+Ari *tirá* < ECush.; Cush.+Chad.).

48.4. Om.: (N) Malo *kilahó*, Koyra *kille* id. (Cerulli) || (S) Ari *käla* "kidney" (Bender), cf. Koyra *kalátte* id. (Hayward), Wolayta *killahuwa*, Gamu-Dache *kilaho* id. (Lamberti & Sottile 1997, 410), related with or borrowed from Ethio-Semitic: Gurage *kiläyo*, Harari *kuläy* "kidney" or Cush.: (E) **kal-* "kidney" (Sasse 1979, 12) || Dahalo *kalle* "kidney" ||| ?Chad.: (C) Hwona *kulìša*, Bura-Ngwaxi *kulši*, Ga'anda *kùlkùlarà* "kidney" (all Kraft) ||| Sem.: **kuly-*(*at-*)/**kaly-*(*at-*) "kidney" (SED I, 141)..

48.5. Om.: (N) Dizi-Nayi *bow* id., Shako *bo* "belly" (E 80, #3: NOm. **bu:p* "chest") .

48.6. Om.: (N) Yemsa *zogara* id. (Bender); cf. Ganjule *zu'ano* "chest" (Fleming) ||| Chad.: Hausa *zukàatáa* "heart"; Bolewa *zogo* "heart, liver, life"; Warji *zugwai* "soul", Diri *ajùkwá* "heart" (Skinner 1977, 25) || (C) Masa *duka* "liver" ||| Berb.: Guanche *a-ta-zayka-te* "gran cuore" (Woelfel 1965, 413–14: Guanche + Hausa).

48.7. Om.: (N) Oyda *šɪnfɪlɪ* id. < HECush.: Kambatta & Tembaro *šinfilla* or Ethio-Semitic: Tigre, Tigrinya, Amhara *šənfəlla*, Zway *šənfəlla*, Gogot, Selti *šənfəl* "small stomach of ruminants"; cf. also Bilin *šanfilla* and Kafa *šippiloo* (Leslau 1979, 581).

49. **long**

49.1. Om.: (N) **gallal*(*l*)- id. || (S) Galila *g'al-* id. ||| Cush.: ?(S) Qwadza *gala* "very" (Ehret 1980, 235 otherwise) ||| Chad.: (C) Hitkalanci *gùlù* "big"; Masa *gal* id. (Lukas) || (E) Mokilko *gòolé* "big" (Jungraithmayr) ||| Berb.: (S) Ayr *aglu* "surpasser", *megla* "surcroît, augmentation" (NZ III, 770) ||| Sem.: Arab. *ǧalla* "to be big, elevated"; Tigre *gällälä* "honorer, révérer" (Cohen 1970f, 126; Müller 1975, 68: CChad.+Arab.)

49.2. Om.: (N) Gonga **ge(:)nǯ* - id. || Ari-Ubamer *gažmi* id., maybe also Dime *gudum*, Hamer *gʊdʊb*, Karo *gutup* id. ||| Cush.: (N) Beja *gúmad* "long" (Roper) || (C) Bilin *g$^{(w)}$ənd-* adj. "fat" (CDA 64). Cf. E 181, #269: Ari+Chad.: (W) Ngizim *gâzb‹r* "tall, long, deep" + Sem.: Arab. *ǧaḏba* "to stretch, extend"

49.3. Om.: (N) **p'ad-* -(E 113, #102).

49.4. Om.: (N) Zergulla *di:tso*; Yemsa *diča* id.; derived from the verb attested in Yemsa *dič-* "to grow", Gofa *dič'-*, Dache, Wolaita, Gamu *dičč'-* id. (Lamberti 1993b, 337).

49.5. Om.: (S) Banna *orma* id. ⟨or⟩ Cush.: (E) Tsamay *'orma* "long" (Fleming); cf. also NS: Ongota *'orma/urma* "long, far".

50. **louse**

50.1. Om.: (N) *č'ugu*(*č*)*č*$^{(ˌ)}$–(H 270, 284). Lamberti (1993a, 401) and Lamberti & Sottile (1997, 328) compared it with ECush. **d'uud'-/*d'udd'-/*d'und'-* "ant" > Saho *d'u:d'e* "ant"; Jiddu *d'udd'u*, Bayso *tunce* "red ant", Dirayta & Mossiya *d'ud'eta* "ant"; Burji *c'uc'e*, Hadiyya *c'u:ma* id.

50.2. Om.: (N) Mao *k'íšè* id. || (S) **k'as-a* id.

50.3. Om.: (S) Dime *garse* id. ||| Chad. (E) Mokilko *gérsè* (Jungraithmayr), Sokoro *ŋgírsa* (Lukas), Tumak *ngírsā* (Caprile), Lele *ngírsa*, Kwang *ngərsà* id.

51. **man**

51.1. Om.: **'ats-*(H 283) = #65.

51.2. Om.: (N) Ometo **'u(:)r-* > Zala *u:ra*, Gofa *ura* "man", Wolaita *issi-'ura* "someone" (Lamberti & Sottile 1997, 292); Gonga **wur-* "male" (Cerulli 1951, 403, 515; Leslau 1959, 57; Lamberti 1993a, 271) ||| Cush. (E) Somali *war* "man" (Cerulli 1951, 403) || (S) **war-* "mature young person" (Ehret 1980, 311–12; E 463, #974: Gonga + SCush.).

51.3. Om.: (N) Kafa *anamo* "man, male", cf. Mocha *nà:mo* "son", Bench *nam* "daughter", *nans* "son" ||| Cush.: (E) **nam-/*nim-/*num-* "person" > Saho-Afar *num*; Somali *nin*, PL. *nim-an*; Oromo, Konso, Dirayta *nam-a*; besides **'anam-* "son, boy" (Sasse 1979, 24) || Chad.: (W) Barang *nyam*; Ngizim *nin* "person" || (E) Tumak PL. *ném-nám* id.; Ndam *nòm-tō* "woman"; Sumray *nàmde* "people" (JI_2, 267; Stolbova 2005, 143) ||| Sem.: Arab. *nummā* "somebody, anybody" (E 320, #621: Kafa + ECush. + Chad. + Arab.).

52. **many**

52.1. Om.: (N) Ometo **la:g-* id. ||| Cush.: (E) Konso *-leka* (Lamberti), Dirayta *-leki*, Mussiye *-lekka* id. (both SLLE) (E 420, #868: Ometo + Cush. **łag-* "much" + Sem. **ŝ-g* "to be large in number, size").

52.2. Om.: (N) Zergulla *mač'o*; Kafa *meeti*, Mocha *mẹ:tto* id. ||| ?Cush. (C) Awngi *ménč*, Kunfäl *menči* id.

52.3. Ometo **dar-*: see #5.1. **dar-* "big".

52.4. Om.: (S) **bed-i* "many, much" ||| Cush.: (E) Saho *bad-* "to be rich"; Somali *badi* "to make numerous; increase", *badan* "many", Jiddu *badne* id. ||| ?Sem.: Arab. *badin* "fat and strong", *bada^{c}a* "to grow fat" (Cohen 1970f, 46; E 113, #102).

53. **meat**

53.1. Om.: (N) **ʾač-(*H 297, fn. 32) ||| Cush. (C) Awngi *əšši* id., Kunfäl *eš* (Appleyard, CDA 98, separates it from NCCush. **səx-* and speculates about the Omotic influence; E 366, #735: Om. **a:yš$_2$-*, including Ari *a:yzi* "skin" + Awngi + Eg. *3js* "viscera").

54. **moon**

54.1. Om.: (N) Bench *ʾyarp*, She *erf* id. || (S) **ʾarp-i* id. ||| Cush.: (C) **ʾärb-/*ʾärf-* id. > Bilin *ʾarbá*, Qwara *arfa*, Kailiña *arfa*, Khamir *arβa*, Kemant CR *arfa*, Khamtanga *árba*, Awngi *árfá*, Kunfäl *arfa* "moon" (CDA 100; E 354, #696: Om.+CCush.). There are interesting parallels in NS: Kuliak: Nyangatom *aragwan*, Ik *aragwan^{a}*, So *awan* "moon" (Lamberti 1988, 117).

54.2. Om.: (N) **ʾam(t)si-/*ʾatsim-* > Yemsa *assa*; Bworo *aašittsa*; Dizi *atsɪm*; Mao *ʾá:nsé*, Seze *ʾémsí* id. (E 362, #722: Mao **a:ms-*) ||| Cush.: (E) Afar-Saho *alsa* "moon/month". An attractive parallel occurs in NS: Kuliak: Ik *asan* "moon" (Lamberti 1988, 117).

54.3. Om.: (N) Ometo **ʾaguun-/*ʾagunn-* id.; Kafa *agene*, Mocha *ʾagane*; Shako *a^{y}kn* id. ||| or > Cush.: (E) HECush. **agan-a* id., but Burji *agun-co* (D 49; Sasse 1982, 24; Hudson 1989, 404) ||| Chad.: (W) Ron: Sha *ʾaŋwén-* "star" (Jungraithmayr); NBauchi: Siri *ìgání* "sky" || (E) Sokoro *ígindī* "Morgen" (Lukas) ||| Eg. (MK) *gnḫ.t* "Stern des Himmels", PL. "Gestirne" (Wb. V, 177) ||| Berb.: (S) Ahaggar *ağənna*, PL. *iğənnawän* "ciel" < **ā-ginnāw*, PL. **ī-ginnāwan*, cf. *tağnawt* "petite pluie", preserving the 3rd radical **w* | (E) Ghadames *ajenna* "ciel" | (N) Shilha *ignna*, Kabyle *igenni* | (W) Zenaga PL. *gunuen* id. (NZ III, 837–38; Takács 1996, 149 & EDE I, 173: NOm. + HECush. + Berb. + Chad. + Eg.)

54.4. Om.: (N) Anfillo *wəro* id. < Ethio-Semitic: Gurage, Amhara *wär*, Geez *warḫ* "moon, month", cf. further Mehri *warḫ* "month", Sabaic *wrḫ* "moon", etc. (Leslau 1987, 617). Ehret is apparently mistaken, comparing Anfillo + Arab. *warya* "to burn, blaze" (E 462, #973).

55. **mountain**

55.1. Om.: (N) **dar-* id., cf. Gamu *dare* "country" ||| Cush.: (N) Beja *dar* "edge, bank of khor" (R. Hudson) || (E) Mussiye (SLLE) = Bussa (Bender) *tarra* "mountain", Konso *tarra*; Gawwada *taraara* id. (Lamberti 1993a, 298) ||| Chad.: (W) Ngamo *diri* "hill" (Meek) ||| Berb: Guanche *adara* "monte" (Woelfel 1965, 594–95) || (E) Siwa *ədrär*, PL. *idrärən*, Ghadames *adurar*, PL. *durāren* id.; cf. Ancient Libyan Δύρεν "the Atlas mountains" | (N) Nefusi *drar*, Kabyle, Shenua, Rif, Shilha

adrar | (S) Ahaggar *adrar*, PL. *idrârän* "mountain" < **ā-dirār*, PL. **ī-dirāran* (Prasse 1974, 183; NZ II, 371) ||| Sem.: Akk. *midru* "eine Art Land"; Hebrew *mɛḏɛr* "clod of earth"; Syriac *meḏrā* "clod, soil"; Arab. *madar* "clods of earth, mud", Sabaic *mdr* "territory, ground"; Geez *mədr* "earth, ground, soil, field, country, land", perhaps the *nomen loci* formed by the prefix **ma-/*mi-* from the original root **d-r-w*, attested in Jewish Aramaic *dᵉrā* "élever, soutenir"; Harsusi *derō* "atteinde le sommet" (Leslau 1987, 330; Cohen 1970f, 312; Bender 1975, 176: Shilha + Wolaita).

55.2. Om.: (N) Male *d'uka* "stone"; Kafa B *t'aak'o* "stone" < ECush.? || (S) Hamer *d'uk'a*, Banna *d'uka* "mountain" ||| Cush.: (E) **d'ag-* "stone" (Sasse 1982, 61) ||| Berber: (E) Siwa *a-dyaɣ*, *tyoɣ-ət*, Sokna *t-yaɣi-t* | (N) Gurara, Zwara, etc. *adyaɣ*, Wargla, Shawiya, Sened, Qabyle, Nefusi *adyaɣ* "stone" (NZ II, 359–60).

55.3. Om.: (N) Bworo *gura*; Shako-Nayi *ge:ra*; Hozo *gura*, Seze *gʊ̀lì* id. ||| Cush.: (C) Bilin *gira* "mountain", Qwara *gara* "rock" (Reinisch) || (E) Somali *guro* "punto, cima, vertice"; Oromo *gaara* "mountain"; Dullay: Tsamay *gaar-ko* "wood", Dopasunte *kaar-ko*, Harso, Gollango *kaar-kó* id. || (S) **gada* > Iraqw-Gorowa *gara* "virgin forest", Alagwa *Gara* "name of a mountain" (Kiessling & Mous 2003, 112; D 61: NOm. + Cush.) ||| Chad.. (C) Masa *gɔrhina*, Musseye *gɔira* "stone" (Kraft).

55.4. Om.: (N) Ometo **buč'-*.

55.5. Om.: (N) Ometo **zum-*.

55.6. Om.: (N) Nayi *mai* ||| Chad.: (C) Kilba *mâ'*; Wamdiu *məw*; Bachama *mwɛy* "mountain" (Kraft).

55.7. Om. (S) Ari *balá*, Ubamer *bəlá* id. ||| Berber: (S) Ghat, Kel-Ui *ablal*, Ahaggar *a-blal*, PL. *i-blâl-än*, Awlemidden *a-blal*, PL. *i-blal-ăn* etc., "stone", is derivable from pBerb **ā-bilāl*, PL. **ī-bilāl-an* (Prasse 1974, 183: F.VII.A.1; NZ I, 54) ||| Sem.: Akkadian *billu* "ein Stein" (AHw 126) or Mehri *aybel* "silex, briquet" (Cohen 1970f, 3–4). Note: Cohen (1947, #79) compared Tuareg *ablal* and Arabic *'aᶜbalu* "rocher blanc", Soqotri *ᶜebhaléten* "pierres pointues" (Leslau 1938, 293).

56. **mouth**

56.1. Om.: (N) **no:n-ᵕ(*Lamberti 1993a, 369; E 326, #636: NOm. + Arab. *namm* "breath, breeze") || (S) **na(:)m-i* "name"; cf. HECush. **suume* "mouth, lip" vs. **sum'a* "name" (Hudson 1989, 92–93).

56.2. Om.: (S) **'ap-* id., cf. Yemsa *a:fa* "language" (E 100, #65), which is most probably borrowed from ECush. (Oromo?) ||| Cush.: (N) Beja *yaf* "mouth" (Roper) || Cush: (C) **'äf-* > Bilin *'äb*, PL. *'äfəf* "mouth" (CDA 102; Qwara *af* & Falasha *af* can be borrowed from Amhara *af* id.) || (E) **'af-* > Saho-Afar *af*; Som, Boni, Rend *af*; Dasenech *'af-u*; Or *af-aani*, Konso *af-aa*; Burji *afay*, Sid, Alaba, Kamb *af-oo*, Gedeo *af-aʔo* id., Had *af-oʔo* "hole" (Sasse 1982, 23; Hudson 1989, 102) || Dahalo

'áfo id. (To) || (S) **'af-* > Iraqw, Burungi, Alagwa *afa*; Asa *afok*; Qwadza *afuko* id.; Ma'a *aféta* "doorpost" (Ehret 1980, 281; Id. 1987, #191: N+C+E+SCush + Dahalo) ||| Sem. **pay-* "mouth" > Akk. *pū*; Ugaritic *p*, Hebrew *pe*, constr. *pī*, Syriac *pummō*, Arab. *fam*, constr. *fū*, Sabaic *f* (SED I, 195–96, 126); Ethio-Sem. **'af* id. (Leslau 1987, 9) is rather of Cushitic (C or E) origin. Rabin (1981, 27, #24) adds the following WChad. forms: pAngas **p*^{w}*a* > Sura *pwɔ̀ɔ̀*, Chip *pwɔ*, Angas *po*, Ankwe *pu* "mouth"; Ron: Fyer *fo*, Daffo *fo*(*h*) id. (JI$_2$, 244–45), while Stolbova (1987, 152) derived them from pWChad **ba-ki*. An alternative (or another?) Chadic cognate may be identified in WChad **'aap-* "to open (mouth)" > Hausa *áfà* "to throw in mouth"; Chip, Montol *'ɛp*, Ankwe *ep* "to open", Sura *ap*, Angas *ep* "to yawn"; Karekare *ʔaf-*, Tangale *ʔap-*, Pero *ápp* etc. (Stolbova 1987, 230), plus Eg. (Pyr) **wpy* /*ỉpy* "to open, separate" (Wb. I, 298; EDE I, 109–10).

57. **name**

57.1. Om.: (N) **sum-* id. (H 270: **sun-*; E 160, #220: **sum*(*ts*)-) ||| Cush. **sim-*/**sum-* id. > (N) Beja *sim* id. || (C) **šəŋ*^{w} id. (CDA 103) || HECush. **sum'-a* id. (Hudson 1989, 103) ||| Chad. **sumi* id. (Stolbova 1996, 55–56) > Hausa *súúnáá*; Sura, Daffo *súm*; Kirfi *šimmi*; Pa'a *sun*, *sɪm*; Zaar *sûm* || (C) Tera *łim*, Gidar *łum*; Muktele *tlúm*; Yedina *hemú*, Logone *ŝémii*; Masa *sém-ná* || (E) Kera *sám*, Sumray *súmí*; Migama *sémè*; Mokilko *sùmá*; Mubi *sàmì* (JI$_2$, 248–49) ||| Eg. MK *smj* "berichten, anzeigen", Copt. *smmɛ* "accuser" (Wb. IV, 127–28; Vycichl 1983, 189) ||| Sem. **šum-*~(Leslau 1987, 504; E 160, #220: NOm. + Cush. + Chad. + Eg. + Sem.). The word for "name" in most of the Berber idioms resemble Arab. *ism*, e.g., (E) Ghadames *ism* | (N) Izdeg, Senhaja, Snus *ism* | Zenaga *ašm* (Basset), *ešəm* (Nicolas) | (S) Kel-Ui *asam*, Awlemidden *əsəm*, Ghat, Adghaq *isəm*, Ahaggar *isəm*, PL. *ismawän* id. < **Hisim*, PL. **Hisimāwan* (Prasse 1974, 124–25), but such forms as Zenaga or Kel-Ui indicate that these forms are independently inherited. Vycichl (1955, 307, fn. 5) mentioned, that in the case of borrowing from Arabic, the Zenaga word should preserve Arabic *s*. He reconstructs proto-Berber **a-sim*.

57.2. Om.: (S) **na*(:)*m-i* id. (E 315, #610: SOm. **na:b-* + Mocha *nàb-* + Chad. **n-b* "to count" + Sem. **n-b* "to call, cry"); cf. MOm. **no:n-* "mouth" – see #56.1.

58. **neck**

58.1. Om.: (N) Ometo **k'od'-* id., cf. Mocha *qo:t'-* "to swallow" (E 249, #452; Lamberti & Sottile 1997, 436), but cf. Hamer-Banna *k'orči* "neck" ||| Cush.: (N) Beja *kwod'ad'* ~ *kad'ad'* "base of skull, foot of mount" (R. Hudson) = *kad'at* "scruff of neck" (Roper) || (E) **k'uc'*-(*t*-) > Jiddu *quj'* "neck" (Lamberti); Oromo *quc'-e* "back of the neck" (> Gedeo *k'uc'-e*), Kambatta *k'utta-ta*, Burji *k'uc'-oo* "nape of neck" (Hudson 1989, 104);

58.2. Om.: (N) Zergulla *wollo*, Zayse *wɔ́:llɔ*; She (Montadon) *wul*, Bench *ul* id., perhaps from the verb of the type Yemsa *wòl-* "to return" (E 460, #968: Yemsa + Cush. **wal-/*wel-* "to go around" > (C) Khamir *wulɔs-* "herumrühren", Awngi *wol* "neighborhood" || (E) Harso *ollas-uy-* "to roll" || (S) **wal-/*wel-* "to go round, revolve" – see Ehret 1987, #576).

58.3. Om.: (N) Nayi *feli*, Shako *fö:ri* id. ||| ?Cush.: (E) Saho-Afar *filla* id. (Bender) ||| Chad. (C) Hildi *fulàŋoo* id. (Kraft), Bura *fulam*, Margi *fə́láŋ* "throat, swallow" (Hoffmann).

59. **new**

59.1. Om.: (N) Ometo **kill-/*kiil-*; ?Seze *ʾɔ́:lé*, Hozo *hɛlli* || (S) **kill-/?*(*k*)*all-*.

59.2. Om.: (N) Ometo **ʾor-ats-* ||| ?Cush.: (E) **haar-* "new (thing)" (Sasse 1982, 87–88). Note: Lamberti (1988, 118) connected ECush. + Ik (Kuliak)u *er~~u~~-ts*[a] "new".

60. **night** -

60.1. Om.: (N) Ometo **č'el-*(*m-*) > Zayse, Koyra *č'emmo*, Zergulla *č'e:mo* id. || Ari *č'el-mi* "black", Hamer-Banna *č'alɛ-mi* (H 266). See #8.4 "black". This inner Omotic comparison probably excludes Chad. * *ʗama* "night, dark(ness)" (Stolbova 1996, 47) as a cognate, but Stolbova (l.c.) reconstructs pChad. * *ʗilVm-* "to be dark", compatible with both Sem. * *ṯ -l-m* "to be dark" and Om. **č'el-*(*m-*).

60.2. Om. (N) Bench *t'uman* "at night", Gonga **t'um-* "night"; cf. Wolaita *t'uma* "darkness", Gamu & Dache *d'uma* id. || (S) Ari *d'um* id. ||| Cush.: (C) Khamtanga, Khamir, Kemant *təm-a*, Qwara *təm* "darkness" (CDA 52) reflect **təm-*, compatible with ECush. **d'um-* "to be dark"; it is not necessary to propose an irregular cognate in **dVm-an-* "cloud" (so Appleyard) || (E) **d'um-* > Saho *d'um-* "to become dark"; Konso & Dirayta *d'um-* "to set (of sun)" (Lamberti 1993a, 396: Om. + Konsoid); Sidamo, Hadiyya, Kambatta *tuns-* "to become dark" (Hudson 1989, 47), Sidamo *tuma* "darkness" (Moreno), but Hadiyya *t'umo* "darkness", *t'uns-* "to become dark" by Plazikowski-Brauner (D 53–54: Gonga + HECush.) ||| ?Chad.: (W) Pero *ḍamḍam* "darkness" || (E) Mubi *ḍèḍém*, Mokilko *dìḍḍó* "night" (JI_2, 257) ||| ?Eg. *ḥtmtm* "?" (Wb. III, 199), Copt. Sahidic *htomtm*, Bohairic *thomtem* "obscurir, s'obscurir" (Vycichl 1983, 316 compares it with Arabic *ḥutma* "couleur noire"; Id. 1934, 43, fn. 2: Copt. + CCush.). The emphatic dental is indicated by Arab. *taʾaṭṭum* "to be very dark" (Steingass 1988, 67). Cf. also Takács, *Lingua Posnaniensis* 38, 1996, 61–62: Ometo + Pero+Mubi.

60.3. Om.: (N) She *dutan*; ?Anfillo *dihro* "night" ||| Cush.: ?(N) Beja *títa* "Finsterniss, Dunkelheit; finstere Nacht" (Reinisch) || (E) Afar *data* "to be black", *diti* "blackness" (Parker & Hayward), Saho *data* "black" (Welmers); Oromo Waata *doti* id. (Heine).

60.4. Om.: (N) Ometo **k'amm-* id. ||| Cush.: (C) **k*[w]*əm-* "evening" (CDA 62).

60.5. Om.: (S) **soyt-i* id. ||| Cush.: (N) Beja *sóotay, súutay, sóoday* "dunkelfärbig, dunkelgrün, -braun, -grau" (Reinisch).

61. **nose**

61.1. Om.: (N) **si:d'-(*H 269) = **sint'-(*B 147) = **sind'-(*E 161, #224; Dullay **sind'-* < NOm.) ||| Cush: (C) **'əsan-/*'əsən-* id. (CDA 105–06) || (E) **san-/*sin-/*sun-* id. (Sasse 1982, 169) || Dahalo *sina* id. ||| Chad. **sunə* "to smell" (Newman 1977, 31) ||| Eg. (Pyr) *sn* "to smell" (Wb. IV, 153).

61.2. Om.: (N) Ometo **kunk-* id. (E 230, #401) < **nuknuk-*? || (S) **nuk-i* id. ||| Cush.: (S) Qwadza *ningwato*, Ma'a *nuŋa* id. (Ehret) ||| Chad.: (W) Burrum *nyuŋsi* (Jungraithmayr), Tule *ŋgisəʰn* id. (Shimizu) || (E) Kera *nīnə* id. (Ebert).

61.3. Om.: (N) Chara *mudda*; Kafa *muddo* id. (Cerulli) ||| ?Cush.: (N) Beja *miida, midala* "tongue" (Reinisch) || (E) Oromo *madde* "cheek" (Gregg) ||| Chad.: (C) Mafa *muddẹy* "front, forehead" (Mouchet).

62. **not**

62.1. Om.: (N) **ba* (E 80, #2; B 1988, 152: **b'a*) ||| Cush.: (N) Beja *baa-/bii-* "not" || (C) **b-* "to lack, not have" (CDA 90) || (E) Hadiyya *be'e* "to not be present", Gedeo *-ba-/-bo-* "negative infix", Kambatta *-ba-* "suffix of negative perfect & imperfect", Burji *abóonn*(*i*) "no" (Sasse 1982, 22) || Dahalo *ɓa*(*:*)- "part of several negative forms" || (S) Burunge *-bei* "not" ||| Chad. **ba* "negative marker" > (W) Hausa *bà*(*a*)- .. *-bá*(*a*) "nicht"; Mupun *bà*; Daffo-Butura *ba'*; Karekare *bái*, Ngamo *bu*; Zumbul *bi*; Ngizim *bái* || (C) Bura *bi*; Gudur *bá*; Masa *bay* || (E) Kera *bà*; Jegu *bâa-* ||| Berb.: (E) Sokna *a-bû* "ne pas" | (S) Ahaggar *bobo* "non!", *a-ba* "ne pas y avoir de, ne plus y avoir de", Awlemidden-Ayr *i-ba* "ne pas y avoir de" (NZ I, 1) ||| Eg. (NK) *bw* "nicht" (Wb. I, 453) ||| Sem.: Punic *by* "without"; Soqotri *be* "sans; avant que"; Tigre *bäy bäy* "no", Gurage Ennemor *ba'ay*, Eža *bay* "no" (D 39: NOm. + N+C+ECush. + Chad. + Eg.; EDE II, 176–78: NOm. + Cush. + Chad. + Eg. + Berb. + Sem.).

62.2. Om.: (N) Yemsa *ánè* "not" ||| Cush.: (E) Somali *an* "not"; Oromo *hin* "not", Konso *an-* "prefix of negation" (Lamberti 1993b, 318) ||| Sem. **'-n*(*-y*) > Akk. *ya'nu/a*, *yānu*, Ugaritic *ỉn*, Hebrew *'ayin*, *'ēn* "il n'y a pas", Arab. **'in* in *'illā* "ne .. que", Geez **'ən-* in *'ənbəya* "je n'ai pas" (Cohen 1970f, 24).

62.3. Om.: (S) Hamer *-ma* id. ||| Cush.: (E) **ma'* > Afar *ma*, Somali *ma'*, Dasanech *ma* "negative particle" (Sasse 1979, 52) ||| Berb.: (S) Ayr *ma* "ne .. pas" (Alojaly) < **mā* (Prasse 1972, 247) | (N) Shawiya *ma* id., if it is not borrowed from Arab. *mā* "not" (Prasse, l.c.) ||| Eg. (Pyr.) *m* "nicht" (Wb. II, 3–4: derived from *ỉmj* "nicht sein") ||| Sem.: Arab. *mā* "not" (Brockelmann 1908, 500); Harari *mē'* "no!" (Leslau 1963, 102)..

63. **old**

63.1. Om.: (N) Chara *gänä*, Gonga **ge(:)n-* id. ||| Cush.: (C) **gän-* "be old, grow old" (CDA 106) || (E) **gaan-/*gann-* "to grow; be big" (Sasse 1982, 73, 78) || Dahalo

gààno "big, large" (Ehret) || ?(S) Iraqw *gaŋ* "very (much)" (Ehret 1980, 236) ||| Eg. (Pyr.) *gn* "mächtig sein; der grosser Machthaber" (Wb. V, 173; D 211: NOm.+C+ECush. + Eg; E 185, #281: Mocha + Cush.).

63.2. Om.: (N) Ometo **gal'-a* || (S) Ari *galta* id.

62.3. Om.: (S) Hamer *gec*(*c*)*o*, Dime-Galila *gɛš* id. ||| or ⟨or⟩ Cush.: (E) Dullay: Tsamay *gecc-ate/-ayke*, Gollango *g'eeš-*, Gawwada of Dalpena *keeš-akko* id.

64. **one**

64.1. Om.: (N) **ʾis-* > Ometo attributive **ʾissi*(*-na*), singulative **ʾis-ta*; Yemsa *isa*; Nayi *isn*; Seze *ìšílɛ̀*, Mao *íškɛ̀*; maybe also Gonga **ʾikka*, if it is derivable from **ʾis-ka* (E 363, #726: NOm. + ECush. **ʾis-* "self" + Sem.: Arab. *'ins* "person", but it is doubtful) ||| Sem. **ᶜaštay-* "1".

64.2. Om.: (N) Ometo **pet*(*t*)- id. ||| Chad.: (C) Margi *'pʰāːtū'* (JI_2, 262).

64.3. Om.: (N) Gimirra **mat'-* id.; cf. Kafa *mittoo* "sole" (Cerulli) ||| or < Cush.: (E) **mat-/*mit-* "1" (Sasse 1982, 143) ||| Chad.: (W) Kirfi *modi*, Bolewa *moḍi* "1" (JI_2, 262) ||| (C) Mefele *məətá* "1". Takács (1996, 136–37) offers a convincing hypothesis, assuming the derivation from the appellative of the type ECush. **matḥ* "head" (Sasse 1979, 36) || Beja *mat* "(top of) head" ||| Chad.: (C) Musgu *mada*, Vulum *măt* "head" ||| Eg. (OK) *mtj* "Vorsteher einer Priestergilde" (Wb. II, 168) ||| Sem.: Akk. *muttu* "front, forehead". See also #38.8.

64.4. Om.: (S) Dime *wokkil* vs. Aroid **wəl*(*l*)*ak'-* id.; cf. NOm.: Koyra *wolakko* "one of two" (Cerulli).

64.5. Om.: (S) Hamer-Banna-Karo **kala* id. There are NS parallels of the type West Nilotic **kɛl* "1"; Central Sudanic **kala* "1"; Taman **kara* "1" etc.

65. **person**

65.1. Om.: (N) **ʾats-(*H 283; E 365, #733: NOm. + Cush. **ʾats-/*ʾits-/*ʾuts-* "torso").

65.2. Om.: (S) **ʾed-i* id. || NOm. **ʾad-* "man, male" > Malo F *adde*, Kachama B *əde*, Gamu *adde*, Dache *ad*(*d*)*e*, Zayse *ade*, Koyra *ade:*, Yemsa *adka* (Lamberti & Sottile 1997, 301) ||| Cush.: (E) Rendille *ét* "person" || (S) Asa *'iduk*, Burunge *hedi*, Alagwa *heru*, Iraqw *he* id. (E 343, #668: Aroid + Cush. **ᶜad-/*ᶜid-* "body of a person") ||| Chad. (C) Musgoy *híddi*, Kola *hìdí*, Daba *hìd* "person (JI_2, 267) ||| Berb.: (N) Sened *ida* "les gens de" in the tribal names, e.g., *Ida Ugarsmukt*, *Ida Ultit* (Pr 117); Ksurs *idu* "people" (R. Basset), Shilha *id*(*a*) "people, nation" (Stumme) ||| Sem: Geez *ᶜəd* "man, male, husband", PL. *-aw* "men, people", Tigre *ᶜad* "family, people", Tigray *ᶜaddi* "country, village, people"; cf. also Hebrew *ᶜędy-ęk* "the prime" (Leslau 1987, 56).

AA: ? Ch: (C) Gidar *di* "people" : *dəf* "man", Musgu *day* : *dif* id., Musgoy *híddi* "man", Mboku *du* ~ *d'* id., *dąąhąy* "people" (Mouchet) ||| Om: (N) Zayse *ade*, Dache, Gamu *adde*, Koyra *adee* "man", Yemsa *ad-kā* "male being" (Lamberti & Sottile 1997, 301)

65.3. Om.: (N) Ometo **ge(:)r-* id. ||| Cush.: (C) **gərw-* "man, male" (CDA 96) || (E) **gor-/*ger-* "people" > Elmolo *gura*; Oromo *jara*; Dobase *koro*; Burji *jóora* "family" (Sasse 1982, 111) ||| Chad.: (W) Sura *gùrùm*, Tal *gwìym* "person"; Bole *gɔrzo* "man" || (C) Gude *gūrá* "man"; Glavda *ghwálvà*; Mafa *ŋgúrà'à* id. || (E) Mubi *njò-gòrók* id. (JI_2, 230–31, 266–67).

65.4. Om.: Seze-Hozo **mo:* id. ||| or ⟨or⟩ Cush.: (E) Arbore *mó*(*h*), Elmolo *mó*, Dasanech *mú* id. ||| Chad.: (W) Tangale *muu*, Dera *múù* "person" (JI_2, 266).

66. **rain**

66.1. Om.: (N) Ometo **'i(:)r-/*'irr-*, Yemsa *iro*; Dizoid **i(:)r-u*, cf. also Kafa *iro* "cloud" (Cerulli) and Ometo **'irts'-* "wet" ||| Cush.: (C) Awngi *ərri* (Lamberti) = *ɔ́rí* "rain" (Hetzron), Kunfäl *rintu* = **əri yintuy[a]* "rain has come" (CDA 113 || (E) **'ir-/ᶜir-* > Somali *ᶜiiro* "fog, smoke"; Arbore *'iri*, Elmolo *iri*, Dasanech *'ir* "rain", perhaps also Dullay: Tsamay *'εεrro* (SLLE); Gollango *irrawo*, Harso-Dobase *irawwo*, Gawwada *errawo* id. (Lamberti & Sottile 1997, 292: NOm. + C+ECush.) ||| Chad. (W) Ron: Sha *'arè*, Bokkos *ri'*, Daffo *ɽí* "cloud" (Jungraithmayr) || (C) WMargi *'yar*, Kilba *'ɛr* "smoke" (E 353, #695: NOm. "rain" & Ometo "wet" + Cush.).

66.2. Om.: (N) Yemsa *yeešu*; Gonga **'amiš-*; Seze *'ins$^{(ɔ)}$í*, Hozo *ɪmsi* id. ||| Cush.: (N) Beja *yam* PL. "water" ||| Chad. "water": (W) Sura *àm*; Fyer *ham*, Kulere *'aàm*; Tangale *am*, Kirfi *amma*; Pa'a *ambi*, Jimbin *ìmbí*; ?Boghom, Kir *yip*; Ngizim *âm* || (C) Tera *'yim*; Margi *'ímí*; Fali-Kiria *jiami*; Nzangi *mbii*; Laamang *íimí*; Wandala *jàwè*; Sukur *yâm*; Gisiga *yam*; Daba *yɪm*; Buduma *āmái*; Musgu *yim*; Zime-Batna *b̰ì* || (E) Kwang *káam*; Kabalai *kaamə̀*; Sumray PL. *nímī*; Sokoro *mbo* id., Mawa *amî* "water; rain"; Dangla *ámáy*, Migama *àmmì* id., Bidiya *'àmay* water; rain"; Mubi *'àm* "water" (JI_2 340–41) ||| Berb.: "water": (E) Siwa *aman* PL. (Laoust), Ghadames *āman* m. PL. (Lanfry) || (S) Ahaggar PL. *âman* < **HamāHan* (Prasse 1974, 410), Ayr & Awlemidden PL. *aṃan* (Alojaly) || (W) Zenaga *aman* (Basset) || (N) Shilha of Tazerwalt *aman* PL. tantum (Stumme) ||| Guanche *a*(*h*)*emon* "water" (Woelfel 1965: 513; Vycichl, *WZKM* 52[1955]: 314 and *Mémorial Basset* [1957]: 143, reconstructed an unattested SG. as *$^{+}$im* or *$^{+}$ymi* resp., while the plural should reflect **i-imē-ĕn* or **i-ymi*(*yu*)*n* resp.) ||| Sem. **yammɪ* "sea" > Ugaritic *ym*, Hebrew *yām*, PL. *yammī* "sea, lake; reservoir, large basin", Aram *yammā*, Ar *yamm* "sea" (Aistleitner 1965: 129; Klein 1987: 259). A Semitic source is evident for Eg. (18th Dyn) *ym* "Meer" (Wb I: 78), Demotic *ym* id., Copt Sahidic *eiom*, Bohairic *iom*, Ahminic & Fayyumic *iam* "mer; pressoir", Bohairic PL. *amaiu* < **yammīy-u* (Vycichl 1983: 63; E 300, #569: Gonga **am-* + Chad. **am* "water").

66.3. Om.: (N) **deeb-/*doob-* > Dizi *dɪɛb* "to rain" || (S) Dime *dɪ:bɛ*, Hamer *dubi*, Ari *do:bi* "rain" ||| Chad. (C) Gidar *dúbbya* "saison des pluies" (Mouchet) || (E) Kera of Fianga *dubueni* "rain" (Lukas). (E 444, #928: SOm. **de:b-/*do:b-*).

66.4. Om: (N) Ometo: Koyra-Oyda *buk-* "to rain", Wolaita *buk'*-(Cerulli), Kachama *bok'-(*Conti Rossini) id. ||| Cush.: (C) **bək^{w}-äna* (CDA 46) > Bilin *bək^{w}ána*, Qwara *bekåna* "cloud", Kemant *buk^{w}ana* "fog" (Conti Rossini) || (E) Somali Hawiya *bokkod*, Oromo Wellega *bokkaa*, Borana *bokaha* "rain" ||| Chad.: (E) Mubi *'bok* "regnen", Jegu *boŋ* id., Bidiya *ḅok* "ooze" (Stolbova 1996, 126: EChad. + Arab. *b-q-q* "to rain") ||| Berb.: (N) Snus *tbika & tbica*, Figig *tbica* "pluie" (NZ I, 101; D 270 & E 93, #12: Ometo + Cush.; Bender 1975, 180: Oromo + Mubi).

66.5. Om.: (N) Gimirra **wol-* id. ||| Cush.: (C) Kunfäl *wel*, Awngi SLLE *wʊl* "cloud" (CDA 46) || (E) Burji *wáalla* "cloud, fog" (Sasse 1982, 186 compares it with Yemsa *waalla* "night" by Bender = *waalà* "Abend & Nacht" by Lamberti).

67. **red**

67.1. Om.: (N) **zuk'-* > Ometo **zok'-/*zuk'-*, Chara *zúwa*; Bench *zok'*; Mao *zàŋkítɛ̀* || (S) Dime *zu*:-, Hamer-Banna *za:wu* etc. (H 265) ||| Chad.: (W) Hausa *já*(*á*), PL. *jáájààyéé* "red"; Ron: Bokkos *júwi* id., *ji jyu* "to become red"; North Bauchi *jiyu* "yellow" (Skinner 1996, 119) || (C) Wandala *jágana*; Padukwo *žágana*; Kotoko: Logone *zey* "rouge" (Mouchet). ||| Berb: Guanche of Palma *azuquahe* "moreno o negro" (Woelfel 1965, 425–26) || (E) Siwa *a-zəggəy*, PL. *i-ən* "rouge" (Laoust 1931, 290), Augila *zuây*, PL. *-en* "rosso" (Paradisi), Ghadames *zaggay* id. (Motylinski) | (S) Ahaggar *ezäggäy*, PL. *izäggäyän* "rouge" < **ē-zawway*, PL. **ī-zawway-an* (Prasse 1974, 148) vs. *ihway* "être rouge" (Foucauld) = Ayr-Awlemidden *izway*, PERF. *zăggay* id. (Alojaly) | (N) Nefusi *zuggáy* "rosso", determ. *azuggây* "rosso" (Beguinot), Iznacen & Senhaja *azugg^{w}ay* "rouge" (Renisio), Zayan *azegg^{w}ay* id. (Loubignac), Kabyle *azegg^{w}ay* id. : *ezwey/izwiy/uzwuy* "rougir" (Dallet), Shilha *zwiy* "rot werden" : *zgg^{w}åy* "rot sein" (Stumme) | (W) Zenaga *jobba* "rouge" (R. Basset) Takács (*BSOAS* 63, 2000, 268–69: Om. + Berb. + Chad.) proposes an attractive hypothesis, adding here Eg. *zš* "to write; paint" (Wb. III, 475), derivable via **zẖ* from **zḫ*, compatible with the Berber counterparts. The semantic development has a good analogy in Indo-European languages: Slavic *pъsati* "to write" : Old Indic *piśáŋga-* "rötlich, rotbraun".

67.2. Om.: (N) Bworo *bira* id. ||| or < Cush.: (E) **bu'r-* "red, brown, yellow" > Dasanech *bur* "red", Arbore *burr-i-d'a* id. etc. (Sasse 1982, 39). Cf. further Beja *boy* "blood" (Roper) || CCush. **bər-* "blood" (CDA 34) ||| Chad.: (W) Jimbin *bàrámá*; Jimi *biram* || (C) Gidar *bēli*; Musgu *fẹ̀l* || (E) Lele *kùbàrò*; Sumray *bàrē*, Dangla *báári*; Jegu *bar* "blood" (JI$_2$, 30–31; Newman 1977, 22: **bar-*; Stolbova 1996, 19: ** ḥa-puram*).

68. **road**

68.1. Om.: (N) Basketo-Dokka-Male **goyts-i* id.; Anfillo *gɔ:tto* || (S) Hamer-Banna **goyt-i*, ?Dime *gaš* id.

68.2. Om.: (N) Zayse *goge*, ?Dizoid **ko:k-* id. || (S) Ari-Ubamer-Galila **go:gi* ||| Cush.: (N) Beja *giig-* "to go away" || (C) Bilin *gug*, Khamta *g^{w}əg* id.; Agaw > Geez *gogawa* "to wander" (CDA 116) || ?(E) Hadiyya *googo*, Kambatta *goggoo* "road" (Hudson 1989, 124 derives them from pHECush. **doogo*, regarding Sidamo *doogo*, PL. *doogga*, Burji *dawwa* id.) ||| Chad.: (W) Bole *goggo*, Dera *gokò* (Kraft) "road" (D 256: Dizi+Ari + N+C+ECush; E 222, #380: Dizoid & Ari + Chad.).

68.3. Om.: (N) Gonga **wor-et-* id. ||| Cush.: (E) Gedeo *ora* "road" (Hudson 1989, 124) || (S) Gorowa *urûwa* "road, way" (Whiteley) ||| Chad.: (W) Ankwe *war* "road" (Stolbova 1987, 231 adds Sura *ar*, Angas *aar*; Kulere *'araw* "road"; Tangale *ware* "to come"; Tala *war*, Zem *wàri* id.) || (E) Mokilko *'ûrzí*, Bidiya *'óorà*, Kajakse *ìrí*; Mawa *ər* "road" (Robertson) ||| Eg. (OK) *w3.t* "Weg" (Wb. I, 246–48) ||| ?Sem. **'urḫ-* > Akk. *urḫ* "road", Aramaic *'urḥ*, Hebrew *'ōra* id. : Soqotri *'era* "venir, arriver" (Cohen 1970f, 32; Müller 1975, 66: Kafa + Eg. + Ankwe; Gidar *úrą*, quoted by Greenberg 1963, 61, does not mean "road", but "brousse", i.e., "bush", see J. Mouchet, *Études camerounaises*, Tome III[1950], No. 29–30, 16).

68.4. Om.: (N) Gimirra **bod*; Nayi *buřun* id. ||| Cush.: (N) Beja *bid'* "to make way" (Roper) || ?(E) Konso *póra* "road" (Black), Dirayta & Mossiya *pora* "place" (Lamberti 1993a, 405) < **b-*||| Chad.: (E) Mubi *bò̱dò̱l*, Dangla *bótól* "road" ||| Berb.: (N) Nefusi *brīd*, PL. *ibrîden* "via di campagna" | (E) Ghadames *abrīd*, PL. *berdan & bridawen* "chemin" | (S) Ahaggar *ăbarid* "chemin", Ghat *abarid*, PL. *ibariden* "chemin, route" < **ā-bārīd*, PL. **ī-bārīd-an* (Prasse 1974, 299: F.XVIII.A.1; Kossmann 1999, 128) (Bender 1975, 181: Tamazight + Mubi).

69. **root**

69.1. Om. **ts'ab'-* > Om. (N) **ts'ab'-* id. > HECush.: Alaba *t'ap'a*, Kambatta *tap'a* id. (Lamberti 1993a, 400) || (S) Ari *ts'ab'-i* "blood vessel" (H 265, 267; **ts'ayp'-* ||| Cush.: (E) Yaaku *t'eebe* "sinew, limb" (Ehret 1987, #115; E 173, #260: NOm. + Yaaku).

70. **round**

70.1. Om: (N) Male *mu·mme* id. < **mul-m*^{o}? (> Bayso *muumee* id.) || (S) Dime *mılu*, Galila *mulla* id.

70.2. Om.: (N) Bench *kar* "to be round" ||| ?Cush.: (E) **kir-/*kor-/*kur-* "circular formation" (Sasse 1979, 5, 9) || (S) **kar-* > Qwadza *kangal-* "to turn around"; Maa *kikarara* "ring" (Ehret 1980, 242) ||| Berb.: (E) Ghadames *kəwwər* "to be round" (Lanfry) | (W) Zenaga *kōrər* id. (Nicolas) | (N) Shenwa *akərnənnay*, Kabyle *ak^{w}ərnənnay* "round" (E 200, #328: Bench + SCush. + EChad. **k-r* "to dance" + Arab. *karra* "to unroll itself").

71. **sand**

71.1. Om.: (N) Ometo: Dokka *k'ač'e*; She *k'ac'*, Bench *k'amč'*; Kafo *qač'o*; Shako *k'amč'u* id. || (S) Galila *k'ač'* in *p'ič' k'ač'*, where *p'ič'a* means "earth".

71.2. Om.: (N) Cancha-Gamu-Dorze *anc'o*, Chara *ámša* id. Is it a variant of #71.1.?
71.3. Om.: (N) Dizi *harč* || (S) Ubamer *ha:ča* id. A variant of #71.2.?
71.4. Om.: (N) Ometo **šap'-*; Anfillo *ša:po*, while Mao *šá:ʷà* id., together with Hozo *šakiwi*, Seze *šà:kɛ́wí* id. probably represent another etymon. Borrowed into (or from?) HECush. **šaafa* "sand". It is tempting to compare the Berber parallels: (S) Ahaggar *asuf* "valley" (Prasse 1974, 213: **ā-sūff*), Ghat *asif*, PL. *isaffen* id. | (N) Sened *suf*, PL. *iseffen* "river", Shawiya *suf*, PL. *isaffen* id., Kabyle, Demnat, Shilha *assif* id. (Woelfel 1955, 156).
71.5. Om.: (N) Gonga **ša(:)w-/*šaww-* > Mocha *šaawo* id., *šawwe* "land", Bworo *šawa* "(Fuss)boden", Kafa *šawo & šowo* "land" (Lamberti 1993a, 384) || (D) Dime *šayy* (Fleming) ||| Chad.: (W) South Bauchi: Boghom *šey*, Zeem *aŝi* "sand" (Shimizu) || (C) Kotoko: Affade *se* id. (Sölken) ||| Eg. (Pyr) *šᶜ* "sand" (Wb. IV, 419–20; EDE I, 382: Dime + Chad. + Eg.).
71.6. Om.: (N) Gonga **šiiš-* > Bworo *šiya*, Mocha *ší·šo* id. ||| Cush.: ?(C) Bilin *ᵊsäs* "dust" || (E) Afar *siisa*, Somali *siiso* "dust" (Lamberti 1993a, 385: Gonga + Bilin + ECush.).
71.7. Om.: (S) *ša(:)m-i* id. ||| or ⟨or⟩ Cush.: (E) Dullay: Tsamay *šumáḥ-tó* id. (Hayward); cf. also NS: Ongota F *šumaha / šumaxa* id. ||| ?Berb.: (N) Rif Iboq *smiḏ* id.

72. **say**
72.1. Om.: (N) **ga[h]-/*gi[h]-* > Basketo *ga:-*, Male *gɛ'-*, Gofa *gi-*, Dorze *gay-*, Bench *gah*, Shako *ge:* id. || (S) Hamer *gi-*, Ari *ga-*, Dime *geem-* id. ||| or > or < Cush.: (E) Tsamay (Bender) *gæhæ* id. (E 183, #274: Bench + SCush. **geh-* "to speak" + CChad. **g-* "mouth").
72.2. Om.: (N) **y-* > Dokka *iyya*, Zayse *yi-*, Zergulla *y-*; Yemsa *i* id. ||| Cush.: (C) **y[ə]-* "to say" (CDA 118) || (E) **-y[y]-*(Sasse 1982, 118) || Dahalo *joom-* "to say, speak" || (S) **yo-/*'o-* id. (Ehret 1980, 318) ||| Eg. Pyr. *j* id., *jj* "Spruch" (Wb. I, 36; D 184–85; E 471, #991: NOm. + Cush. + Eg.; EDE I, 79: Cush. + Eg.).

73. **see**
73.1. Om.: (N) **bek'-* id. ||| Cush.: (E) **beeq-* > Somali *beeq- & beeg-* "sorgfältige Ausschau wie etwas ins Werk zu setzen sei, Versuch, Probe" (Reinisch), "to measure" (Luling); Oromo *beek-* "to know" > Gedeo *beek-*, Burji *beeh-* id. (Sasse 1982, 35) ||| (W) Hausa *bííƙì* "devoting attention to"; Boghom *ḅayoŋ* "to look for", Mangas *bùk* id. || (C) Mofu-Gudur *bák* "regarder (un jeune homme)"; Gisiga *bok* "zeigen" || (E) Kera *ḅégé* "bekanntgeben, informieren" ||| Eg. (CT) *bq* "erblicken, überblicken, beschauen" (Wb. I, 426) ||| Sem. **b-q-w* > Aramaic *bᵉqā* "examiner, essayer", Arab. *baqā* "regarder qn., attendre qc." (Cohen 1970f, 78; D 264: NOm. + Somali+Oromo + Arab.; EDE II, 325–26: NOm. + ECush. + Chad. + Eg. + Sem.).

73.2. Om. **zag-* > (N) Male *zag-* ; Dizoid **sag* id. || (S) Hamer *zag-* "to seek", Ari *ziig-* id. (H 265)
73.3. Om.: (N) **ts'il-(*Lamberti 1993a, 398–99; E 170, #250: NOm. + Sem.: Arab. *ṭilāᶜa* "to look at attentively, see; show" – see Steingass 1988, 641).
73.4. Om.: (S) **šed-* ||| or > or < HECush.: Hadiyya *siid-* "to look & find" (Hudson 1989, 94).

74. **seed**
74.1. Om.: (N) Ometo **zer-ets-*; Bench *zar*; Yemsa *zala*; Gonga **yar-*; Shako *zara*, Nayi *yaru* id. || (S) Ari *zer-* seed, sow" (H 265) < Ethio-Semitic: Geez *zarʾ* "seed", Tigre *zärʾa* "seed, sow", Amhara, Gurage *zärra* id. (Leslau 1987, 642).
74.2. Om.: (N) Oyda *budíts*, ?Zergulla *bič'ɛtta*; Dizi *butknɪs* || ?(S) **bet-a /*pʾet-a* id. ||| ?Cush.: (E) Arbore *báaḍo* "seed" (H); Gawwada *pod'aḥe* (AMS), Tsamay *booraḥ* (Savà) ||| ?Eg. (Pyr) *bd.t* "Emmer, Spelt" (Wb. I, 486).
74.3. Om.: (N) Bworo *šooka* id., Bworo & Kafa *šok* "to sow" (Lamberti 1993a, 374–75 adds Gedeo *sok'-* "säen, mahlen"), Mocha *šókki* "to seed"; Maoid **šok-* "seed" ||| Chad.: (W) Hausa *shuka* "to sow" || (C) Vulum *súki* "faire le trou avant de semer", Mbara *čok* "to sow" (Tourneux et al.) ||| Eg. (MK) *skj* "Samen, Nachkommenschaft vertilgen" (Wb. IV, 312–13: connected with *skj* "vernichten, zu Grunde richten"; Blažek & Boisson 1992, 26: Gonga + Chad. + Eg.).
74.4. Om.: (S) Dime *mʌšitʰ*, Hamer *maš-*, Galila *məšá* id. ||| Chad.: (W) Ron: Bokkos *musút* "Saatkorn", Daffo-Butura *masút* "Samen" (Jungraithmayr).

75. **sit**
75.1. Om.: (N) Basketo *dɔʾa-*, Dokka *duʾ-*, Male *dɛʾ-*; Yemsa *di / duwu* id. ||| Cush.: (N) Beja *diʾ* "tun, (fertig) machen", caus. *daa-s & dadaa-s* "legen, stellen" (Reinisch) || ?(C) Bilin *diw-* "to stay, wait" (CDA 129) || (E) Saho *dehe* "to give back" (Welmers) = *dahay & dihay* "zurückgeben, -stellen" (Reinisch) ||| Chad. (C) Logone *ḍə* "stellen" (Lukas); ?Musgu *da* "tun, machen" (Lukas) ||| Sem.: Arab. *wadaᶜa* "poser", Soqotri *daḥ* id. (Leslau 1938, 125; D 186: Saho + CChad. + Sem.).
75.2. Om.: (N) Ometo **ʾut-* id. (E 355, #701: Ometo + Cush. **ᶜat-/*ᶜit-* "to stop")
75.3. Om.: (N) **bet-* > (E 90, #38 compares it with Cush. **baᶜ-* "to lie down")
75.4. Om.: (N) Anfillo *t'ɛːpa* id. ||| Chad. (W) Ngizim *dláb'(a)-* "to sit, remain, stay, residue, be, become" (E 415, #852: Anfillo + Ngizim).

76. **skin**
76.1. Om.: (N) Wolaita *goga*, Chara *gongá*; Yemsa *googo*; Bworo *gook'a*, Kafa *goqo*, Anfillo *gɔŋk'o*; Hozo *goŋk'a* id. < Ethio-Semitic: Gurage *goga*, Harari *googa* < Cush. (E) **gog-* id., perhaps derived from **gog-* "dry" ||| Chad.: (C) Wandala Mch. *gọ́gwa* "peau" (E 211, #354: NOm. < Cush.)

76.2. Om.: (N) Shako *ba:c'i* id. || (S) **bič-* id. ||| Cush.: (E) **biš-* "body" > Tsamay *biš-ko*, Dopasunte *piš-ko*, Harso-Dobase *piš-kó* (E 87, #25: SOm. + ECush.); maybe of NS origin: Gumuz (B): Sai *bıca*, Sese *bıc'a*, Kokit *bıc'a* "flesh, meat"; Songhay: Gao *baši* id.; Ongota F *bíša* "body".

76.3. Om.: (N) Dizi *kut*; Mao *kúté*. There is a tempting parallel in NS: Kuliak: So *kʉs/ kʉt* (Lamberti 1988, 121).

76.4. Om.: (N) Yemsa *kooda*; Bworo *k'edda*. Probably borrowed from Amhara *k'oda* "skin".

77. **sleep**

77.1. Om.: (N) Ometo **geh-* id. ||| Cush.: (C) Khamtanga *g^{w}äyy-*, Kailiña *g^{w}äq̄i* - "to sit" (CDA 124) || (E) Burji *goh-* "to sleep" (Sasse 1982, 82); Yaaku *-gwɛ'ɛ-* "to live" (Heine) || Dahalo *gwaḥ* "to sit" || (S) Qwadza *gwahas-* "to rest" (Ehret 1980, 264; Ehret 1995, 190, #298 adds Sem. **g--*, but without any material support).

77.2. Om.: (N) Gonga: Bworo *k'ey-* = B *k'eu*, Kafa B *k'ɛbe*, Mocha *qè·yĭ-* = B *k'ɛb^{e}* id. ||| Cush. **k'ab-/*k'ib-* "to be quiet" > (C) Awngi *qəpp-* id. || (E) Yaaku *-qap-* id. || (S) Burunge *qab-* id. (Ehret 1987, 42, #146) ||| ?Eg. (late) *qbb* "ruhig, langsam", usually connected with *qb*(*b*) "kühl, kalt" (Wb. V, 22–23; E 233, #410).

77.3. Om.: (N) Ometo **d'isk-*(H 268) > Gofa *d'isk-*, Kullo *disk-*(Borelli), Wolaita H *t'iskk-* = L *t'esk-* id. ||| Cush. (E) Oromo *c'iis-* "to lie down, rest, stretch oneself", besides Burji *dosk-* "to sleep" and Ethio-Semitic: *dəqqasa* "to be sleepy", Tigray, Amhara *däqqäsä* "to sleep" (Lamberti & Sottile 1997, 537–38) ||| Berb.: (E) Siwa *əṭṭrəs*; (S) Ahaggar, Awlemidden *ərrəs*; (N) Senhaja *əṭṭras*, Iznacen, Zkara *yiṭṭəs*, Nefusa *əṭṭəs* "to sleep" (Militarev).

77.4. Om.: (N) Dizoid **sok$^{(,)}$-* id. ||| Cush.: (E) Somali *sug-* "to wait"; Oromo *fokokaa-* "to move sitting down" || (S) Iraqw *suknunu'at-* "to squat" (Ehret 1980, 326, 350) ||| Sem. **s-k-n* "to reside", **s-k-b* "to lie down"; Arab. *saka^{ᶜ}a* "to sleep, fall asleep" (E 157, #214: Dizoid + Cush. + Sem. < AA **suk-* "to stay").

78. **small**

78.1. Om.: (N) **gu:tts-* id. (H 284) ||| Cush: (E) Arbore *gičč'á* id.; Hadiyya *gotto'o* id. || Dahalo *g^{w}íttsa* id. (E 193, #305: AA **-g^{w}îts-* > NOm. **gu:tts-/*gi:tts-* + Dahalo; Lamberti 1993a, 317: Om. + Hadiyya).

78.2. Om.: (S) Dime *lek'-*; Hamer-Banna *lıkka* id. ||| Chad. (W) Angas *lok* "small", Ankwe *tior-lok*, Montol *šwor-lok*; Ron: Fyer *lókò* "young, small"; Tangale Kr *làkey* "small" (Stolbova 1987, 237: WChad. **lV[ḳ]V*; Ead. 2005, 75: **lVKV & *lVḳV*: Chad. + Arab.) ||| Eg. (MK) *nkt* "etwas von, ein wenig .." (Wb. II, 347) ||| Sem.: Arab. *lukluk* "court, petit" (Takács 1996, 141: WChad. + Eg.)

78.3. Om.: (S) Dime *c'ɛkk'-* id. ||| Cush.: (C) **cəg^{w}-* > Bilin *šəg^{w}*, Qwara *šigw*, Khamir *ciq^{w}-*, Kemant *šəgg^{w}-* id. (CDA 125) || (E) ?Konso (Black) *šakk-* = (Fleming) *šəka*

id. (D 119) || (S) Burunge *coko* id. ||| Chad.: (C) Wandala *cukwà* id. (Kraft) ||| Berb.: (E) Siwa *aḥkīk & ḥakîk* "petit", Sokna *məʒ̣ʒ̣î*, PL. *měuʒ̣ʒ̣kət*, f. *məʒ̣ʒ̣okiyət*, PL. *məʒ̣ʒ̣úknat* (Laoust 1931, 275), Augila *meššék* "I am small", Fodjaha *məẓâi* "piccolo; figlio" (Paradisi) | (N) Nefusi *meššék*, PL. *meššûket* "piccolo" (Beguinot); Senhaja *məẓẓəy* (Renisio), Zayan *amәẓẓyan* : v. *meẓẓi* (Loubignac) etc. | (W) Zenaga *iməzzigən* (R. Basset), *məẓẓūg* id. (Nicolas).

78.4. Om.: (N) Male *d'aka* id. || ?(S) **toko-* id. ||| Cush.: (E) **dikk'-/*d'ikk'-* "small" (Sasse 1982, 47).

78.5. Om.: (N) Ometo **'er-ats-*; Bench *yars* id. ||| (E) **yar-* > Somali *yar* "small", Rendille *yeryer* "thin", Bayso *yer-aat-* "to decrease"; "Oromo *yar-aa* weak"; Burji *yér-aa* "bad (thing); ugly; dirty" (Sasse 1982, 193; Haberland & Lamberti 1988, 148–49: Zayse + ECush.).

79. **smoke**

79.1. Om. (N) **č'ugw-* id., cf. the velar responses in Oyda *č'uggə*, Anfillo *s'u:ggo* (H 266, 270: **č'ub-*; E 290, #546: Om. + ECush. **c'aab-* "to be afire") || (S) **č'ub-* "smoke" ||| Cush.: (C) Wag *č'əga* "smoke" || Dahalo *t'uggwa* "smoke" (Tosco). Cf. #12.1. "burn".

79.2. Om.: (N) She *šarr* id. – see #14.1. "cloud".

80. **stand**

80.1. Om.: (N) Ometo **ekk'* – id. ||| Cush. (N) Beja *yak* "aufstehen" (Reinisch).

80.2. Om: (N) Bench *yit'* & She *ita*; Gonga **(y)eed*$^{(.)}$- || ?(S) Karo *wedi*, Galila *wə*'- id. ||| Cush.: (E) Somali *heed'-* "to remain over"; Sidamo *heed'-*, Hadiyya *hee'-*, Burji *yed'-* "to live, be in place" (Sasse 1982, 193).

80.3. Om.: (N) Chara *aš-ne*; Dizoid **aš-* id. ||| ?Cush.: **'əs-* "to do, make" (CDA 55–56) || (E) **'as-/*'is-* > Saho *is-* ~ *iš-* "to do, make"; Boni *as-* "to prepare, make"; Sidamo *ass-*, Kambatta *ass-*, Hadiyya *iss-*, Burji *iss-* "to do, act, make" (Sasse 1982, 107; D 151).

81. **star**

81.1. Om.: (N) Ometo **ts'olint-* id. > HECush.: Burji *t'oolintee*, °*dee*, °*gee* (Sasse 1982, 182; Hudson 1989, 142; E 170, #251: Ometo + Gonga **če:ll-* "red" ||| Cush. **c'ilḥ/*c'ulḥ* "coals"). Cf. the suggestive parallel in NS: Kuliak: Ik *'doleat*^{a} "star" (Lamberti 1988, 122).

81.2. Om.: (N) Gonga **t'oǯǯ-* id. (E 440, #920); perhaps related to Ometo **ts'olint-*.

81.3. Om.: (N) Yemsa *ba:kura*, Mao *bɛ̀lgɛ́* id. || (S) Dime *bɪkɪl* "Venus, morning or evening star" (Bender 1994, 161) ||| Cush.: (E) **bakkal-* "morning star", cf. Sidamo *bakkal-* "yellow" (Sasse 1982, 32; Hudson 1989, 142) ||| Eg. (Pyr) *bk3* "der zweite Tag, das Morgen" (Wb. I, 481) ||| Sem. **b-k-r* > Akk. *bakrum* "früh", Hebrew **b-k-r* "frühe sein", hence "frische Früchte tragen; zum Erstgeborenen machen", Arab. *bukrat* "cras; mane" (EDE II, 336; E 83, #11).

81.4. Om.: (N) ?Dokka *bɛrse*; Dizi *biz* || (S) Dime *bɛz*, Galila *bɛz* id. ||| or > or < Cush.: (E) HECush. **beʒʒeekko* "star" > Burji *biziik'o*, Kambatta *bezzeeccu*, PL. *bezzee-bezzaa*, Alaba *beezeta*, Sidamo *beddakko*, PL. *beddahe* (Hudson 1989, 142; Sasse 1976, 137). Cf. also CCush.: Bilin *bídā & bīdárā* "Morgenstern" (Reinisch). The root also passed into Ethio-Semitic: Geez *bez*, *biz*, *bezā* "shining star, morning star" (Leslau 1987, 117).

81.5. Om.: (N) ?Male *wuts'atsi*; Chara *wóńčeta* || (S) Banna *wončo* id.

81.6. Om.: (S) Hamer *'e:zini* id. < ECush.: Dasanech *hizi-n-tti* id. (Sasse 1976, 138).

82. **stone**

82.1. Om.: (N) **šu(:)čč*$^{(.)}$–(H 269, 270), maybe borrowed into Cush.: (E) Garre-Tunni *šị̣ịt*, May-Digil *šiid*, Jiddu *šiit* id.; Oromo *sida'* "heavy and compact stone" (Lamberti 1993a, 382–83; he adds Bilin *šäša* "stone").

82.2. Om.: (N) Zayse-Zergulla **mal-o* id. ||| Eg. (Pyr) *mnw* id., D19/20 *mn*(*.t*) "mountain" (Wb. II, 72, 64, 69).

82.3. Om.: (N) Bench *nyal*; Dizi *ñal-u*, Nayi *nʸel-u* id. || (S) Dime *la:lɛ* id. ||| Chad.: (W) Sura *lar* "Fels, Stein", Angas *ler* "flat rock", Mupun *lāar* "boulder, stone" (Stolbova 1987, 243; Ead. 2005, 95: WChad. + Akk.) ||| Eg. (OK) *ỉnr* id. (Wb. I, 97) ||| Sem.: Akk. *narû* "stone monument, boundary stone".

82.4. Om.: (S) **sun-i* id. ||| Eg. (MK) *sn.t* "block of stone" (Wb. IV, 152).

82.5. Om.: (N) Kafa *t'aqqo* id. < Cush.: (E) **d'ag-* id. (Sasse 1982, 61).

83. **sun**

83.1. Om.: (N) Ometo **'aw*(*w*)-, cf. also Koyra *'aw-* "to shine", but Male *'abi*, Chara *oyá* "sun"; Bench *ab* "day" : *obar* "sun"; Yemsa *awa*; Gonga **'aabb-*; Hozo *abbi*, Mao & Seze **'aw-* id. || (S) Dime *'ɪjɛ*; Aroid **'ay-* id. ||| Cush.: (N) Beja *'aawi* "noon; zenith" (Roper) || (C) Awngi *áwá*, Kunfäl *awi* "sun" (CDA 130 thinks about Omotic origin; the Omotic source is proposed by Leslau for the Gurage forms: Gyeto "*wayä*) ||| Chad.: (W) Ron: Daffo-Butura & Bokkos *ḅwè* "sun" (Jungraithmayr) || (C) Dghwede *ḅiya* "to shine" ||| Eg. (Dyn. XVIII) *ᶜbᶜb* "von den Strahlen der Sonne, die auf Gesicht scheinen" (Wb. I, 178) ||| Sem.: Arab. *ᶜab*(*b*) "sunlight", *ᶜabwa* "to shine, be bright" (E 341, #663: NOm. **'a:b-* + Cush. **ᶜab-* "to burn, shine" + Arab.; Takács 1996, 129: NOm. + HECush. **iibba* "hot, warm" + SCush. **ᶜabakʷ-* "white" + Chad. + Eg. + Arab.; Blažek 2005, 376: Om. + N + CCush.).

83.2. Om.: (N) Zayse *k'ɔ́ɔs*, Koyra *k'ɔ́sɛ* id., cf. Hozo *koza*, Seze *kɔ̀zzì* "star" || (S) Karo *kaco* "star" ||| Cush. (C) **kəs-* "to become dawn" (CDA 52) || (E) Mossiya *kooso* "sun" is probably of Ometo origin. On the other hand, there are NS parallels: Nara *kɔs* "sun" and Ongota *'ak'ac'u*/(*h*)*ak'ac'o*/*akaco*/*xác'o* "sun" (Fleming).

84. **swim**

84.1. Om.: (N) Ometo **wad'-* id. || (S) Hamer *war-*, Ari *wa:rri* ||| Cush.: (C) Kemant *war-* id. (CDA 131). Cf. also HECush.: Hadiyya-Kambatta *waacc-* "to swim" (Hudson 1989, 147) and Gurage: Selti *wäčča* id. (Leslau).

84.2. Om.: (N) Gonga **wak-*; Nayi *wokn* id. The velar extension of the preceding root?

84.3. Om.: (N) Dokka *piŋ-*; Maoid **paaŋ-* id.

84.4. Om.: (N) Dache *lim-* id. ||| or ⟨or⟩ Cush.: (E) Harso *liima*^{c}- id. (AMS). Lamberti (1988, 123) mentioned NS: Kuliak: Ik *il~~u~~l~~u~~m~~u~~-* "to swim" (Lamberti 1988, 123).

84.5. Om.: (N) Bench *mošt*; Shako *mú*$\underset{\circ}{\text{š}}\underset{\sim}{\text{d}}$.

84.6. Om.: Ometo **dak-aad-* id. < HECush. **daak-ad'-* id. < Oromo *dak-* id. < ECush. **zak-* id. (Sasse 1976, 140; Id. 1982, 52; Hudson 1989, 147). SOm.: Ari *zo:tima* can be borrowed from some of the East Cushitic *z*-dialects.

85. **tail**

85.1. Om. **guly-*(*an-*) > (N) Basketo-Dokka **golisi*, Ometo **go*[*l*]*ina* id. || (S) Dime *golan*; Aroid *guli* id. ||| Berber **ā-g*[*u*]*lāl*: (N) Shilha *ag*^{w}*lal* "queue (de lézard)" (NZ III, 772), Seghrushen *ažlal*, Shenwa *hažəlalc*, Menacer *ažlāl* "queue" | ?(S) Ahaggar *ârğal*, PL. *ârğalän* "queue, membre viril", *târğalt*, PL. *târğalîn* "anus" (Foucauld; Prasse 1974, 318: **ā-rgāl*, PL. **ī-rgāl-an*), Ayr & Awlemidden *argal*, PL. *irgalăn* "verge" (Alojaly), Ghat *ardjal*, PL. *irdjalen* id. (Nehlil) < **ā-lgāl* < **ā-glāl*?

85.2. Om.: (N) ?Koyra *dup'iya* id. || (S) Hamer **dub-an/in-* id. ||| Cush: (E) **dub-/*dib-* id. (Sasse 1982, 57) ||| Chad.: (W) Ankwe *däp* (Kraft), Sura *dúp* "penis" (Jungraithmayr) || (C) Gisiga *duba* (Lukas), Mofu *dubạ*, Gidar *dúbo*, Muzgu *dəbạ* "back" (Mouchet) ||| Sem.: Mandaic *dibra* "back, tail"; Arab. *dubr* "backside; neck", Amhara *žəbär* "backbone" (SED I, 44; Cohen 1970f, 212; E 125, #119; elsewhere Ehret [E 134, #146] connects the Koyra word with Eg *dp.t* "loins").

85.3. Om.: (N) Male *'uŋki*; Bench *unk*, She *uŋk*; Yemsa *unyu*; Mao *jɔ̀ng*^{w}*è*, Seze *ŋwi:ni* id. ||| ?Chad. (E) Mubi *'ínyéwī* (JI_2, 317), Minjile *iɲɛwî*, Kajakse *iɲawi* id. (Doornbos & Bender).

85.4. Om.: (N) Chara *s'éra*; Gonga **čˈe(:)r-* id., including Bworo *ts'iira* "hair"; Dizi *čˈɪr-u* id. The areal word, widespread also in various Ethio-Semitic and Cushitic languages: Cush. (E) Oromo *c'ira* "tail; bunch of hair, usually from a horse's tail" (Gamta), Qabenna, Sidamo *č'ira*; (C) Khamta *č'era*, Awngi *č'äri* (CDA 131–32); Ethio-Semitic: Tigrinya, Amhara, Argobba, Gafat, Harari, Gurage *ç̣əra* "tail, hair of tail" (Leslau 1979, 187: Ethio-Semitic < Cush.).

86. **that**

86.1. Om.: (N) **se-* > Cancha *sekkisi*, Zayse *sɔya*, Ganjule *sɛsa*, Koyra *sɛ'ɛssɪ*, Chara *sék*(*ä*)*n*; Yemsa (*h*)*aas*; Shako *yis* || (S) Dime *'ɪsɛno* & *sah* id. Perhaps related with NOm. **'isi* m./**'isa* f. "he/she"; ECush. **'usuu*/**'išii* id.; Beja *-uus* in *bar-uus/*

bat-uus id.; Semitic **s̲uwa/*s̲iya* id.; Egyptian *sw/sj* id. ('dependent'); Chad. **si* "he" (Blažek 1995, 51, 49, 48, 46, 45, 43).

86.2. Om.: (N) **ya-/*yi-* > Basketo *yita*, Male *jeja*, Malo *jéji*, Zergulla *ja*; Bench *yinkuš*; Shako *yis* id.

86.3. Om.: (N) Gonga: Anfillo *bɛni*, Kafa-Mocha **ebi* id. Cf. Bworo *bí* "he" : *bì* "she", Kafa & Mocha *bi* "he" & "she"; Yemsa *bár* "he" : *bàr* "she"; Wolaita *ba* "self" (Lamberti 1993b, 331–32); Benchnon *bàr* "self" (Rapold) ||| Cush.: (N) Beja *ba-* "der, die", m. *been*, f. *beet* "jener" (Reinisch).

87. **this**

87.1. Om.: (N) Dizi *ɛkɛ/yɛk* m./f. id., *ɛŋkɛ/yɛŋk* m./f. "that", cf. also Mocha *òkkabi* "that" || (S) Hamer *ka*, Ari *kona* id. ||| Cush.: (E) abs. m. **ka* : subj. m. **ku* "this" (Sasse 1982, 111); especially HECush. **hikka* m./**hitta* f. "this" (Hudson 1989, 150) || (S) **kaa* "this" m. ||| Eg. Pyr. *kjj* m., *kj.t*, Copt. *ke-* m., *ket-* f. "other, another" (Wb. V, 110; E 194, #309: Aroid & Dizoid + SCush. + Eg.).

87.2. Om.: (N) Ometo **ha-*; Bench *haši*; Yemsa (*h*)*an*; Gonga **hani*; ?Shako *à:z* id. ||| ?Cush.: (C) **'ən-* "this" : **'ən-dV* "that" (CDA 136, 134) ||| Sem. **hā-* in various demonstratives (Brockelmann 1908, 316–23).

88. **thou**

88.1. Om.: (N) abs. **ne*, subj. **ne-ni*, obj. **ne-na* (Blažek 1995, 51). The closest parallels occur in the Nilo-Saharan languages: Saharan: Teda-Daza *na* "thou": *ta* "I"; Kunama *ena*; Kwegu *ìnyá*, Majang *ini* "thou" : *eet* "I", etc. (Zaborski 2004, 181–82).

88.2. Om.: (N) Dizoid **yetV* id. ||| Cush.: (C) **'ənt* || (E) **'ati* || Dahalo *'àaṯa* ||| Chad.: (C) Tera *to*, Pidlimdi *toteŋ* ||| Sem. **'an-ta/-ti* m./f. id.; in the branches, where the independent 2sg. pronoun in *t* was not preserved, at least the verbal affixes reflect their old existence: Beja *ti-* .. *-a/-i* m./f.; Iraqwoid **-it*; Berb. **ta- … -ad/ḍ* and Egyptian Old Perfect *-t*(*j*) (Blažek 1995).

88.3. Om.: (S) **ya*(*-na*) id. : PL. **yV*(*-tV*) "you" (Blažek 1995, 51), plus Maoid: Mao *hi-ya* "thou" : *ti-ya* "I", Ganza *ye* "thou" : *ti* "I". Zaborski (2004, 181–83) finds the Nilo-Saharan parallels: Teso *ijo*, Nyangatom *iyo*; Anywa *yii-ni* "thou" : *aa-ni* "I"; Kunama obj. *-e/-ee* etc.

89. **tongue**

89.1. Om.: (N) **'ants'ir-/*'intsar-*(E 387, #797 reconstructs Gimirra-Gonga **hays'-*, but this isogloss cannot be separated from other NOm. forms).

89.2. Om.: (N) Yemsa *terma* id. There are interesting NS parallels: Saharan: Teda-Daza *termeso*; Kanuri *tə́làm*, Kanembu *dələm*; Maba *delmik* id. (Ehret 2000, 444, #850).

89.3. Om.: (N) Bworo *albeera*; Dizi *yabıl*, Shako *érb*, Nayi *yalb* id. (E 397, #808: NOm. **alib-*) ||| or borrowed from the same source as Cush.: (E) **'/ ᶜanrab-* id. (Sasse

1982, 28) ||| Chad.: (C) Gulfei *ariangu*; Mofu *'ərné*, Gisiga *'irne* (both Seignobos & Tourneaux), Mada *ạrrạ* id. (Mouchet).

89.4. Om.: (S) **'adinb-* id. ||| ?Cush.: (S) Qwadza *ondalimo*; Ma'a *lu-'ánda* id. Related with #89.3.? Cf. Sem. Arab. *ᶜaḏab* "tip of the tongue" (Steingass 1988, 679). There are suggestive parallels in NS: Kuliak: So & Nyangatom *ɛdɛb* "tongue" (Lamberti 1988, 123).

90. **tooth**

90.1. Om.: (N) **'ač-* id. || (S) Galila *ači*, Hamer-Banna *atsi*, Dime *ɪts* id., cf. also Galila *ič-*, Hamer-Banna *its-* etc., "to eat" (E 346, #676) ||| Cush.: (N) Beja *d'a'* "molar" (Reinisch) || (E) **ᶜac'-* > Saho *ad'a* "back teeth"; Arbore *'ač'éč'* "lower jaw" (Hayward) , Dasanech (Bender) *iiso* id. (if not from Om.); Dirayta *add'á* "cheek"; Gollango *ᶜad'o* id., etc. (Lamberti 1987, 533) ||| Chad. (W) **ḥa:cwV* "teeth" (Stolbova 1987, 226).

90.2. Om.: (N) Gimirra-Gonga **gaš-* id. ||| Cush.: (N) Beja *koos* "tooth/horn" (Reinisch) || (E) **gaws-* "molar, chin" (Sasse 1979, 45; Lamberti 1987, 533) || (S) Iraqw *goso* "incisor tooth", ?Qwadza *ko'osiko* "molar tooth" (Ehret 1980, 264) ||| ?Chad.: (W) Dira *gùskùɓa* "chin" : *guskùmɛ̀* "cheeks", Geji *gɨ̀sgɨ̀ɓà* "chin" : *gɨ̀skɨŋ* "cheeks" (Kraft) || (E) Sokoro *gússụmdụm* "Kinn" (Lukas) ||| Berb.: (N) Semlal *ákʊs*, Shilha *axʷs* "tooth"; (W) Zenaga *ūkšən* "teeth" ||| ?Sem.: Arab. *ḏ̣asaᶜa* "to chew the cud" (Steingass 1988, 233–34); cf. Soqotri *g-š-'* "manger" (Leslau 1938, 116–17; D 70–71: Om. + N+E+SCush; E 186, #287: Gimira-Gonga **gaš*$_1$- + Chad. + Arab.).

90.3. Om.: (S) Dime *kəsɪl*, Ari *kasel* id. < or > Cush.: (E) Tsamay *káasala* "molar tooth"; cf. also Nilo-Saharan: Ongota *káasala* id.

90.4. Om.: (N) Bench *san* id.; maybe NOm. **sin-* "face, front side" (Lamberti 1993a, 377) ||| Chad. **sina*(*n*) "tooth" (Stolbova 1996, 57) ||| Berb.: (E) Siwa *asᵃin*, PL. *isînən* (Laoust 1931, 224), Sokna *isîn*, Ghadames *asīn* | (S) Ahaggar *esîn*, PL. *isînän* (Foucauld; Prasse 1974, 125–26: **ē-siHin*, PL. **ī-siHin-an*), Ayr-Awlemidden *ešen*, PL. *išenăn* "dent (incisive; de peigne)" (Alojaly) | (N) Nefusa *sīn* id. ||| Sem. **šinn-* "tooth" (SED I, 219) ||| ?Eg. *sn* "sign depicting the "harpoon"" (Wb. IV, 148) ||| (E 161, #223: Bench & NOm. + Chad. + Berb. + Sem. + Eg. + ECush. **san* "nose"!).

90.5. Om. (N) Ometo **gagg-o* id., cf. Malo *gágginá* "molar" (Fleming) || (S) Ari *geegi* "tooth"; cf. SOm. **ga'-* "to bite".

91. **tree**

91.1. Om. **mits*$^{(,)}$–(H 284; D 250); Om. > Konso *miissa* "bush"; cf. Bworo *míttà* "Baum, Holz, Brennholz", Dache *mič'o* "Holzkohle" (Lamberti 1993a, 360–61). Cf. #12.4. "burn",

91.2. Om.: (N) Koyra *akka* id. (Cerulli) || (S) **'ahak'-a* ||| Cush.: (E) Afar *hak*, Saho Irob *ḥak* "branch"; HECush. *hakk'a* "tree, wood" (Sasse 1982, 90; Hudson 1989, 158).

92. **two**

92.1. Om. **lam-* id. ||| Cush.: (N) Beja *-raama* with "augmentative" *r* in *asaraama* "7" (Reinisch) || (C) **läŋa/*laŋa* id. (CDA 141) || (E) **lam*(*m*)- id. (Sasse 1979, 21, 25) || Dahalo *lííma* id. || (S) Asa *laam & xlam* (Fleming) id. (D 165–67: Om. + N+E+SCush.; E 424, #877: Om. + Cush. **ła(:)m-* + Sem. **ŝima'al* "left hand"!). Dolgopolsky (D 167) thinks about a relation to the Cushitic designation of one of "finger": (N) Beja *lumi* "finger" (Reinisch) || (C) Damot *laŋetii*, Awngi *laŋatii* "finger" (Conti Rossini) || (E) Oromo *elema* (Tutschek), *elemtu* (Vitterbo) "fore-finger". There are interesting parallels in NS: Kuliak: So *nebe-c*, Ik *le'be-ts*^{a} "2" (Lamberti 1988, 124).

92.2. Om.: (N) Dizoid **t'agn* id. There are problematic parallels in Beja *tág*^{w} *& dág*^{w}, PL. *tagúug* "20" and Oromo (Tutschek) *digetam* = (Gragg) *digdama* "20", cf. also Nilo-Saharan: Nara *dokuta* "2" : *doko* "1" (Blažek 1999, 49).

92.3. Om.: (S) Ari **k'asten*, Dime *k'Λt'ám* id. || ?(N) Gonga: Mocha *qáčč̣o* "half", Bworo *kač*, Kafa *qat*$^{(,\ )}$*o* id., cf. also Gurage Wolano *kitto* id. (Leslau 1959, 45).

92.4. Om.: (N) Gonga **gutt-/*gitt-* id. Perhaps of Ethio-Semitic origin: Gyeto, Gogot *k*^{w}*et*, Maskat *x*^{w}*et*, Aymellel *kit*, Harari *ko'ot, kōt* "2"cf. Tigre *kəl'ot* id. (Leslau 1963, 90; Lamberti 1993a, 316).

93. **walk/go**

93.1. Om.: (N) Ometo **b-* || (S) Dime *b-*; ?Ari *bad-* id. ||| Cush.: (N) Beja *ba*(*y*) "to walk" (Roper) ||| Sem. **b-w-'* > Akk. *bâ'u* "to go alongside, come in"; Ugaritic *bỏ*, Phoenician *b'*, Hebrew *bā'* "to come", Arabic *bā'a* "to return" (> Beja *bi'* "to return home"), Sabaic *bh'* "to enter", Geez *bo'a* id. (Cohen 1970f, 50; Leslau 1987, 114–15; E 90, #35: Ometo + Cush. + Sem.).

93.2. Om.: (N) **ham-* > Ometo **ham-*; Bench *ham*; Yemsa *ama*; Mocha *hàmmi-*, Bworo *am-* "to go" ||| ?Sem. **h-m-m* > Hebrew *hāmam* "se mettre en mouvement, troubler", Aramaic Targumic *h*^{a}*mam* "déranger, chasser" (Cohen 1970f, 424; E 382, #779: NOm. **h*$_{1}$*am-* + Sem.).

93.3. Om.: (N) Basketo-Dokka **luk*(*k*)- id.; cf. Arbore *lukk 'eh-ad'* "to walk" = "to take leg" ||| Cush.: (N) Beja *lɛkway* in *tə lɛkway aťagiig* "I am squatting on my heels" (Roper) || (C) **ləkʷ-* "foot/leg" (CDA 71) || (E) **lukk*$^{(,)}$- "foot/leg" (Sasse 1979, 5, 12, 41; Id. 1982, 136; Hudson 1989, 66) || Dahalo *lúk'a* "leg (from thigh to knee)" (Tosco) ||| Berb.: (S) Ayr-Awlemidden *eləɣ*, PL. *elyan* (Alojaly), Ahaggar *êleɣ*, PL. *êlyân* "jambe" (Prasse 1974, 124: **Hiliɣ*, PL. **Hilyān*), Taitoq *ileq*, PL. *ilyan* "jambe de devant (d'un animal)" (Masqueray) | (N) Zayan *ileɣ* "mollet" (Laoust 1920, 119). Appleyard (CDA 71) adds Semitic **h-l-k* "to walk, go" (Cohen 1970f, 413).

93.4. Om.: (N) Dizoid **teg-* || (S) Dime *tɪŋgɛ* ||| (E) Somali *tag* "to go" (Luling).
93.5. Om.: (S) **kay-* ||| Cush.: (E) Somali (Upper Juba) *kooy-*, Tunni *kǫy*, Jiddu *hǫy*, Bayso *koi* imper. "come!", Oromo *kooti* "come", Konso *khoyi* id. (Haberland & Lamberti 1986, 110) || (S) **kaw-* "to go, advance", NOM. act. **kaymu* "going" (Kiessling & Mous 2003, 174).
93.6. Om.: (S) Banna *rond-* id. ||| Cush.: (E) ?Somali *roor-* "to run", Bayso *roor-* "to pass" (Haberland & Lamberti 1988, 134); Yaaku *rɛ'ɛ-* "to run away" (Heine) || Dahalo *ro'-* "to walk" (Ehret) || (S) Ma'a *-ro* "to leave" ||| Chad. (W) Mupun *rú* "to disappear suddenly" ||| Eg. (OK) *rwj* "fortgehen, verlassen" (Wb. II, 406–07; Takács, *Afrikanistische Arbeitspapiere* 61, 2000, 101: Yaaku+Dahalo+Ma'a + Mupun + Eg.) or (E) Arbore *róot-* "walking" || Dahalo *rat-* "to walk" (Tosco).

94. **warm/hot**
94.1. Om.: (N) **k'e(:)ts(ts)'-(*H 280; Lamberti 1993a, 345) ||| ?Chad.: (E) Sokoro *góssuŋo* "Hitze" or *kújā* "warm" (Lukas) ||| Sem. **qayiṯ* - > Ugaritic *qẓ* "summer", Hebrew *qayiṣ*, Biblic Aramaic *qayṭå̊*, Arabic *qayẓ*, Sabaic *qyẓ* id.
94.2. Om.: (N) Dokka *so·lle* || (S) Dime *su:l-*, Banna *šoli* id. ||| ?Chad.: (C) Masa *sala* "hotness" (Kraft).
94.3. Om.: (N) Male *'oidi* || (S) **oyd'-/*oydd-* id. Perhaps borrowed from HECush.: (E) Burji *oyd'-* "to be hot". Sasse (1982, 157) finds a cognate in Konso *awd'-* "to be hot", separating here the root **'aw-*, attested in Oromo *ow-* "to be hot", plus the middle voice suffix *-d'-*. The Male word, isolated within North Omotic, represents probably a loan from South Omotic.
94.4. Om.: (N) Ometo **hop'-* id. ||| or > Cush.: (E) Bayso *oofe* "warm" (Haberland & Lamberti 1988, 64 compares Bayso *oofe*+Mossiya *hobaay-*, Konso *obay-* "to burn"; E 388, #801: Ometo **ho'-* + Cush. **hoo'-* or **'ooh-* "to burn").
94.5. Om.: (N) Ometo **miç'-a* id. (cf. H 280, 296). Cf. #12.4.
94.6. Om.: (N) Ometo **binn-a*; She *bins* id. ||| Chad.: (W) Chip & Ankwe *ɓàn* "hotness"; Dwot *rùp ɓannì* id. (Kraft) || (C) Pidlimdi *bìŋa* id. (Kraft); Wandala *mbrà* "heiss" (Lukas 1937, 121); *-r-* < **-n-* regularly.
94.7. Om.: (N) Malo *waré* id. ||| Cush.: (E) Afar *ure* "to be ignited" (Parker & Hayward) ||| Eg. (Med.) *w3* "dörren (des Getreides)", BD *w3w3w* "Glanz, Leuchten der Sonne", D18 *w3w3.t* "Feuer, Glut" (Wb. I, 244, 250) ||| Sem. **'-w-r* "briller" & Arab. *warā*, *warya* "to yield fire" (Cohen 1970f, 13; 627; Steingass 1988, 1209–10).
94.8. Om.: (N) Yemsa *gumu*; Kafa *game* id. ||| or > or < Cush.: (E) Kambatta *gummuuta*, Sidamo *gammuuta* "hot" (Hudson 1989, 81).
94.9. Om.: (N) Shako *sɛrəb* id. ||| Eg. (Pyr) *srf* id. (Wb. IV, 195; E 267, #486).

95. **water**
95.1. Om.: (N) **h_2a:ts-(*H 279, 284; E 524, #1014: **h_2a:tts-*). In spite of Ehret (E 386, #794), it is tempting to add Aroid: Ari *hats'-* "to wash", if the only difference

consists in glottalization, cf. also Zayse, Gidicho, Kachama *wáats'i*, Chara *á:s'a*, Kafa *ač'o*, etc.

95.2. Om.: (N) **ʾak'-* > Yemsa *aka* "water", cf. Mocha *a:k'ɔ* "wet" (Bender), Kafa *ok'-* "to be wet" (Lamberti 1993a, 263) and maybe She *aki* "to swim" ||| Cush. (C) **ʾaqʷ-* "water" (CDA 144) || (E) Konso *haqa*, Dirayta *hak'a* id. (E 243, #438: NOm. + CCush.) ||| ?Berb.: (S) Teneslemt PL. *qawən* "waters" (Militarev 1991, 247) ||| Sem.: Arab. *ᶜaqā* (*ᶜ-q-y*) "to give to drink, water" (Steingass 1988, 714).

95.3. Om.: (S) **luk'-(> *nuk'-* ?) ||| Cush.: (C) Awngi *lúgha* "rain" (Beke), Awngi of Kwakera *lu¦a-* "to rain" (Fleming) ||| Chad.: (W) Sura *lok* "wet", Mupun *lók* "moisten", *lòk* "wet" || (C) Kotoko: Affade *lókwíh* "moisten" (Stolbova 2005, 68: Chad. + Arab. *l-w-q* "adoucir, amollir") ||| Eg. (Med) *nqw.t* "moisture" (Wb. II, 343; E 320, #619: Aroid **no(:)q-* + Eg. + Cush. **-nok'-* "to flow").

96. **we**

96.1. Om.: (N) Ometo absolutive **nu*, subject **nuni*, object **nuna*; Gimira **nu-na*; Yemsa *inno*; Gonga **noo-/ *nuu-*; Dizi *inu*, Shako-Nayi **nat-*; Hozo *nu-nga* ||| Cush.: (N) Beja *hanán*, *hanín*, *hinín* (Reinisch) || (C) **ʾənn-(*CDA 145) || (E) obj. **na/*ni/*nu* || Dahalo *nyányi* || Ma'a *níne* id.; Iraqwoid **-Vn* "verbal ending of the 1PL." ||| Chad.: (W) Ron excl. **nin* || (C) Bura-Margi **[ʾi]yanu*; Bata *[ʾ]yin*; Lamang **-yiN* || (E) Lai **ni*; Sumrai **nVn*; Dangla *níí* ||| Berb. **na-* "verbal prefix of the 1PL."; the independent pronoun **ənakkʷanī* reflects the plural of **ənakkʷ* "I" ||| Eg. (late) *'inn* = **ʾanānu* (Satzinger) or **ʾanāna* (Vycichl 1983, 13) ||| Sem. gen.-acc. **niʾā-ti*, dat. **niʾā-si*, POSS. **na/*ni/*nu* (Blažek 1995, 46–47; E 363, #725: NOm. + Cush. + Eg. + Sem.).

96.2. Om.: (S) **wVtV* (Blažek 1995, 51–52). Zaborski (2004, 181–82) connects it with some of the Nilo-Saharan counterparts: Teso *isyo*, Nyangatom *suwa*; Kwegu *úúwà* "we".

97. **what**

97.1. Om.: (N) Ometo **ʾay-/*ʾaw-*; Bench *awu* & She *hai-*; Yemsa *aanba* < **ʾayn-*; Gonga **ʾay-b/g/n-* || ?(S) Dime *wɔyɔ:* "what", plus SOm. **ʾay-i* "who" ||| Cush.: (N) Beja *aw*, acc. *aab*, gen. *áay* "who?" (Reinisch) || (C) **ʾäw-* "who" (CDA 147: Awngi *ay* < gen. **ʾaw-i*) || (E) **ʾay(y)-* "who, which" (Sasse 1982, 30) || Dahalo *jíko* "who?", *joka* "which" (Tosco) ||| Chad. **wa* "who" (Newman 1977, 34); cf. also (W) Sura *-yə* "interrogative particle"; Bolewa *-ye* "relative particle", Gera *ya* "what" (Kraft) || (C) Margi *yà* "interrogative particle", Ngala *yagu, yaku* "who" || (E) Sokoro *ye* "wer" (Lukas) ||| Berb.: (N) Mzab, Ntifa *ai* "relative pronoun" ||| Sem. **ʾayy-* "which?" (Cohen 1970f, 16–17; Illič-Svityč 1971, 277: NOm. + N+C+ECush. + W+CChad. + Berb. + Sem.).

97.2. Om. **ʾam-* > (N) Koyra *'am*(*u-na*), Ganjule *aanma*, Chara OBJ. *ama:*, Gonga **ʾam-* id., cf. Chara *ama:bi* m. : *amena* f. "which?", *amitá* "when?", *ámne* "where?", etc.

|| (S) Hamer *am-* "where?", Ari *a̲mná* "which one?", Dime *am oyd* "when?", *amo:* "where?" (Bender 2000, 56, 166) ||| Cush.: (E) **ma*' "what?" (Sasse 1982, 146) || Dahalo *mááka* "what?" : *mááte* "what" (Tosco) ||| Chad. **mi/*mə* id. (Newman 1977, 34) ||| Berb. **mā* "what?", **mī* "who?" (Prasse 1972, 216f) ||| Eg. (OK) *m* "what?, who?" (Wb. II, 4) ||| Sem. **mi*'- & **man-* "who", *ma*' "what?" (E 301, #571: Cush. + Chad. + Eg. + Sem.).

97.3. Om.: (N) Bench *hare*; Dizoid **(y)ir-* || (S) **(h)ar-*.

97.4. Om.: (N) Mao *kɔ̀mísìjà*, Seze-Hozo **kin-*. Cf. #99.1. "who".

98. **white**

98.1. Om.: (N) Ometo **bo:ts*$^{(\cdot\,)}$- id. ||| ?Cush: (S) Qwadza *besiko* "sunlight"; Ma'a *bizári* "yellow" (Ehret 1980, 137: **beets*$^{(\cdot\,)}$-; E 88, #28: Ometo + SCush.) ||| Sem. **b-y-ṣ́* "to be white/yellow" (Cohen 1970f, 63; Leslau 1987, 41; SED I, 41 on the discussion about the relation to **bayṣ́-at-* "egg" & WChad.: South Bauchi **m-buy/'Vŝ* "egg"; cf. also Stolbova 1996, 122). There are interesting parallels in NS: Kuliak: So & Nyangatom *bej*, Ik *'bets'*^{a} "white" (Lamberti 1988, 125).

98.2. Om.: (N) Chara *dälá*; Bench *dal* id.; cf. Wolayta *adilliy*^{a} "yellow" (Lamberti) = *'adile-* id. (Alemayehu) > or < Sidamo *adale* "yellow" (Hudson) ||| Cush.: ?(N) Beja *adali* "pink, red, light brown" and/or *dili* "blackish" (Roper) || (E) Somali *dalas* "gelb" (Reinisch); Oromo of Wellega *daal-a-ca* m., *-i-ttii* f. "(blue-)grey" (Gregg), Borana *daal-a-ca* m., *-tti* f. "yellow, reddish" (Stroomer) > Burji *daal-ácci* "brown" and Amh *daläčča* id. (Sasse 1982, 51); cf. also Oromo of Wellega *daal-ee* "grey" (Gregg) || (S) Qwadza *deles-* "yellow" (Ehret) ||| Chad.: (C) Banana *doliyà* "yellow" (Kraft); (E) ?Sokoro *dálda* "grün" (Lukas) ||| ?Berb.: (S) Ahaggar *dal-ət* "être vert" (Foucauld), Awlemidden *dal-ăt* id. (Alojaly), *i-dāli-n* "vert, bleu clair" (Nicolas); (N) Sened *a-/i-dəl* "vert" (Provotelle), Mzab *a-dal-i* id. (Basset); but also "black": (W) Zenaga *ə̲də̲ž* "noir" (R. Basset) | (N) Wargla *dal* "couleur noir" (Biarnay) ||| Sem.: Syriac *'dlṯ*' "rouge des scribes" (Cohen 1970f, 9).

98.3. Om.: (N) Yemsa *poro* (Lamberti) = *foroo* (Cerulli) ||| Chad.: (W) Hausa *fárìí* id. || (C) Margi *pə̀rtù* id. || (E) Dangla *pórtà*, Migama *púrtà*; Jegu *pórórân*, Birgit *fóróórà*, Mubi *fèrét* id. (JI$_2$, 344–45) ||| Berb.: (S) Ahaggar *ifraw* "être serein (clair, pur et calme)" (NZ III, 646) ||| (Vycichl 1934, 84: Hausa + Ahaggar + Copt. *prre* "glänzen", but primarily "hervorkommen; aufblühen" - see EDE II, 467; D 250–51: Yemsa + Hausa + Berb. **f-r-w*).

98.4. Om.: (N) Gonga **natts*'- id. < Amhara *näčč* "white" (Lamberti 1993a, 370).

99. **who**

99.1. Om.: (N) Ometo **'oo-ni*; Gimirra **'on-*; Yemsa *oon*; Gonga **koo-n(n)i*, cf. Kafa dat. *koo-c*; ?Maoid **ki(n)-* ||| Chad.: (C) Pidlimdi *ki* "who?" (Kraft); Daba *kin*

"what?" (Kraft); Logone *yâm, ywani* "was?" (Lukas); Masa *ge* "who?" (Kraft) || (E) Sumray *kāna* "was?"; Sokoro *kaŋkema* id.; Mubi *gin* "wer?" (Lukas).

99.2. Om.: (N) Hozo *haya* || (S) **'ay-i*, *-nV*. Cf. #97.1.

100. **woman**

100.1. *Om. (N) *ma:čč-(H 283).*

100.2. *Om.: (N)* Dache *indo,* Zergulla *'ında,* Koyra *ındo id.,* further Dokko-Kullo *indee* (Conti Rossini), Basketo *indee,* Koyra *indoo;* Kafa *indo,* Yemsa *intoo* "mother" (Cerulli) ||| *Cush.: (N) Beja enda & énde* "mother" (Reinisch) || (E) Somali Hawiya *hindo* "mother" (D 137: NOm. + N+ECush.).

5.1 Additional words

101. **five**

101.1. Om.: (N) Ometo **'ičč-;* Chara *učá;* Bench *uč;* Gonga **'u(:)č(č);* Dizoid **'uč(č) u(m); ?* Maoid **k'utsi id.* ||| ?Eg. (NK) *qḏt* "hand" = **qiḏw-* > **qiwḏ-* > Coptic "P" *kiǧ,* Bohairic *ǧiǧ,* Sahidic *ciǧ id.* (Vycichl 1983, 350).

101.2. Om.: (N) Male *dɔŋgo* || *or* < *(S) *doŋk' id.* ||| Chad.: (W) South Bauchi: Boghom *dàwnyi,* but Kir *to:n id.* indicates a probable Bantu origin.

101.3. Om.: (S) Dime *šənne id.* < ECush.: Oromo *šan(i)* "5" (cf. Sasse 1975, 249–50).

101.4. Om.: (N) Anfillo *amitto id.* < HECush.: Burji *umútta* "5", Sidamo *onte,* Gedeo *onde,* Hadiyya & Kambatta *onto id.* (Sasse 1982, 184).

102. **four**

102.1. Om.: (N) Ometo **'oydd-,* Chara *obdá;* Bench *od;* Gonga **'awudd- id.* || (S) **'oydd-i id.* The form Chara *ourtōntsā* "40" (Cerulli) allows to reconstruct pOm. **'aβurd-* "4", compatible with Beja *fád'ig* (Roper) ~ *farig* (Reinisch), in old records *fardik* (Krockow), *ferdik* (Lucas) < **fard/ḍig* || (E) **'af(f)ar-* ~ **'af(f)ur-* ||| Chad. **fari-du* or **faru-di* ||| Eg. *fdw m., fdt f.* "4", *ỉfd.t* "Vierheit" < **fdo3wět* < **fida3wat* < **fidarwat* (Blažek 1999, 32–36: from AA **far-* "finger").

102.2. Om.: (N) Yemsa *ačeeč* id.

102.3. Om.: (N) Dizoid **kubm* id.

102.4. Om.: (N) Maoid **(m)bets'-* id.

103. **three**

103.1. Om.: (N) Ometo **hayz(z)-;* Bench *kaz;* Yemsa *keez;* Gonga **keež(ž)-;* Dizoid **ka(:)du id.* (E 228, #396: NOm. **$x_2$ayz-* + Chad. **k-n-d'*).

103.2. ?Maoid **t/siyaz- id.* (or related to #103.1?).

103.3. Om.: (S) **makkan* ~ **makVm id.* ||| Chad. (W) South Bauchi: Jimi *mwaikan,* Geji *mekan id.* || (C) Bata *mwakən,* Bachama *mùwà:kún,* Gudu *makʌn;* Sukur *má:k^{h}ən;* Mafa *màkâr,* Mofu *máákàr,* Gisiga *maakar* with *-r* < **-n-* , although most of the West & Central Chadic cognates are without the prefix **ma-*(JI$_2$, 326–27).

104. **ye**

104.1. *Om.: (N) Ometo *'intuna/i;* Bench *yint(i);* Yemsa *nitto;* Gonga **'int-;* Dizoid **'iti id.* ||| Cush.: (C) **'əntən,* obl. **'ənta-* "you" (CDA 150) || (E) **'atuni/*'atinu* "you" (Sasse 1982, 29) || Dahalo *'atta* "you" (Tosco) ||| ?Chad.: (C) Tera *tun(u)* "you" (Newman) ||| Eg. *-tjwn* "2pl Old Perfect = Pseudoparticiple" ||| Sem. **'an-tumu m. / *'an-tin(n)a f.* "you" < AA **'an-tunwa m. : *'an-tinya f.* (Blažek 1995, 37, 42, 45, 46, 48–51).

104.2. Om.: (S) **yeta/i* id. Cf. the Nilo-Saharan parallels: Teso *yɛsi,* Nyangatom *ezi;* Kwegu *yɛ̀* "you" (Zaborski 2004, 181).

Conclusions

From the present material the following conclusions may be formulated:

1. The average value of the common cognates between the individual North Omotic and South Omotic idioms vacillates around 16%. Applying the glottochronological method modified by Sergei Starostin, it is possible to date the North/South Omotic disintegration to the beginning of the 5th mill. BC. For comparison, applying the same method, the disintegration of the East Cushitic languages is dated to the middle of the 6th mill. BC (13.6%), including the most deviant language Yaaku, or to the middle of the 5th mill. BC (17.7%), without Yaaku.

2. Comparing 33 North Omotic idioms versus 7 South Omotic idioms, the number of the common cognates is naturally higher: totally 91 lexical units, without apparent loanwords 87 lexical units, representing 60 semantic units. As evidence of genetic unity this is enough, although the South Omotic personal pronouns are totally different from the Afroasiatic pattern. But the same can be said about most of the North Omotic languages, at least for the singular series, with exception of the Dizoid group, which preserves best the primary Afroasiatic system of the personal pronouns.

3. The region of Great Ethiopia was and is an area of intensive mutual language contacts. Phonetic criteria do not always allow us to differentiate borrowings from inherited lexicon. As to the direction of borrowing, it is frequently possible only to speculate, if the word is widespread or isolated (so especially 3c):

3a. Ethiosemitic > Omotic: 1.6. "all"; 9.4. "blood"; 14.5. "cloud" & 14.6. "cloud"; 26.6. "fat", maybe also 26.5. "fat"; 29.3. "fish"; 31.8. "foot" & 31.9. "foot"; 36.7. "hair"; 40.5. "heart"; 44.3. "knee"; 48.7. liver"; 74.1. "seed"; 85.6. "tail"; 92.4. "two".

3b. East Cushitic > North Omotic: 22.6. "earth"; 27.5. "feather"; 40.4. "heart"; 84.6. "swim"; maybe also 10.2. "bone"; 48.4. "liver"; 65.3. "person".

3c. East Cushitic > South (& North) Omotic: 10.4. "bone"; 17.3. "die"; 40.6. "heart" (& Male); 56.2. "mouth"; maybe also 29.4. "fish"; 38.9. "head" (& Dizoid); 68.2. "road" (& Zayse); 81.3. "star"; 85.2. "tail" (& Koyra?); ?92.1. "two" (& Ometo-Gimirra).

4. One of the most important results of the present study is the confrontation of the numbers of cognates common to Omotic and other Cushitic branches on the one hand, and to Omotic and the other Afroasiatic families on the other [figures in these brackets indicate questionable cases]:

4a. East Cushitic: 130 [145]; Central Cushitic: 56 [59]; Beja 42 [44]; South Cushitic 33 [40]; Dahalo 25.

4b. Common Cushitic: 155 [171]; Chadic: 113 [115]; Semitic: 71 [77]; Egyptian: 59 [64]; Berber: 47 [50].

These figures cannot be interpreted mechanically. The highest numbers of cognates occur in the units (families) consisting of maximum numbers of languages (Cushitic; Chadic), and in the units occupying the same or neighboring territory (East Cushitic; especially Highland East Cushitic, Dullay, Konsoid, Oromo). There are similar results, *c*. 50±10, for Central Cushitic, Beja and South Cushitic on the one hand, and Egyptian and Berber (i.e., the individual languages or closely related groups) on the other hand. The Semitic family represents a specific family consisting of languages with both old and rich lexicological traditions. One of the Semitic branches, Ethio-Semitic, was and is in immediate contact with various Omotic languages. If we accept the explanation that the high number of cognates in the case of Omotic – East Cushitic was at least partially caused by mutual areal influence, it seems most natural to see in Omotic an independent member of the Afroasiatic macro-phylum.

(5) Occasionally possible Nilo-Saharan parallels were taken in account. This direction of research is very fruitful and needs much more detailed study.

Abbreviations of languages

AA Afroasiatic; Akk. Akkadian; Arab. Arabic; BD Book of Dead; C Central; Chad. Chadic; Copt. Coptic; CT Coffin Texts; Cush. Cushitic; Eg. Egyptian; Gr. Greek Period; Med. Medicine Texts; MK Middle Kingdom; N North; NK New Kingdom; NS Nilo-Saharan; OK Old Kingdom; Om. Omotic; Pyr. Pyramid Texts; S South; Sem. Semitic; W West.

Sources

Anfillo - main Bender 1971; additional Grottanelli 1940.
Ari - main: Fleming 1965.
Basketo - main: Fleming 1971 (ms.); additional: A = Alemayehu Abebe 1993, B = Bender 1971.
Banna - main: Fleming 1965.

Bench - main: Wedekind 1990; additional: Ra = Rapold 2005 (ms.).
Bworo - main: Lamberti 1993a.
Cancha - main: Alemayehu 1993.
Chara - main: Aklilu Yilma 1995.
Dache - main: Bender 1971.
Dime - main: Siebert 1995b; additional F = Fleming 1990 and Fleming apud Bender 1994.
Dizi = main Fleming 1965; additional: Y = Aklilu Yilma 1994 (ms.), Al = Allan 1976.
Dokka - main: Fleming 1971 (ms.).
Dorze - main: Alemayehu 1993.
Hamer - main Lydall 1976 and Lydall apud Bender 1994; additional F = Fleming apud Bender 1994, B = Bender 1975.
Galila - main: Fleming 1971 (ms.)
Gamo - main: Alemayehu 1993.
Ganjule - Brenzinger 1995; additional: S = Siebert (& Hoeft) 1994.
Gidicho - Brenzinger 1995.
Gofa - main: Alemayehu 1993.
Hozo - main: Bender 1994; additional: S = Siebert & Wedekind 1994.
Kachama - main: Brenzinger 1995; additional: S = Siebert (& Hoeft) 1994.
Kafa - main: Habte Mikael 1989.
Karo - main: Bender 1971; additional: M = Moges Yigezu 2004.
Koyra - main: Brenzinger 1995; additional: H = (Siebert &) Hoeft 1994.
Kullo - main: Alemayehu 1993.
Male - main: Siebert 1995.
Malo - main: Siebert & Caudwell 1995.
Mao - main: Siebert, R. - K. & Ch. Wedekind 1993; additional: Grottanelli 1940.
Moča - main: Leslau 1959.
Nayi - main: Aklilu Yilma 1994 (ms.); additional: B = Bender 1971, CR = Conti Rossini 1927..
Oyda - main: Fleming 1971 (ms.)
Seze - main: Siebert & Wedekind 1994.
Shako - main: Aklilu Yilma 1988; additional: B = Bender 1971.
She - main: Bender 1971; additional: CR = Conti Rossini 1925, M = Montadon apud Conti Rossini 1925.
Ubamer - main: Fleming 1971.
Wolaita - main: L = Lamberti & Sottile 1997; additional: A = Alemayehu 1993, B Bender 1971, Bk = Beke 1845, Ce = Cerulli 1929, Ch = Chiomio 1938.

Yemsa - main: Wedekind 1990; additional L = Lamberti 1993b.
Zala - main: Cerulli 1929.
Zayse - main: Brenzinger 1995; additional: S = Siebert (& Hoeft) 1994.
Zergulla - main: Siebert & Hoeft 1994.

Abbreviations used in References

AMS = Amborn, et al. 1980.
BSOAS = Bulletin of the School of Oriental and African Studies.
CDA = Appleyard 2006.
CO = Cushitic-Omotic. Bechhaus-Gerst, et al. 1988.
D = Dolgopolsky 1973.
E = Ehret 1995.
EDE = Etymological Dictionary of Egyptian. Takács 1999–2001.
H = Hayward 1988.
JI = Jungraithmayr & Ibriszimow $1994_{1-2.}$
NZ = Naït-Zerrad 1998–2002.
OLS = Omotic Language Studies. Hayward 1990b.
SAW = Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften.
SED = Semitic Etymological Dictionary. Militarev & Kogan 2000–2005.
SLLE = Survey of Little-known Languages of Ethiopia Linguistic Report.
Wb. = Wörterbuch der ägyptischen Sprache, I–IV. Erman & Grapow 1926–1953.

References

Aklilu, Yilma. 1995. Some Notes on the Chara Language. Sound System and Noun Morphology. *SLLE* 32. 1–12.
Alemayehu, Abebe. 1993. A pilot Survey on Ometo languages. *SLLE* 4. 2–10.
Alemayehu, Abebe. 1993. A Sketch of the Mesketo Grammar. *SLLE* 8. 1–9.
Allan, E.J. 1976. Dizi. Bender 1976: 377–392.
Amborn, H., Minker, G. & Sasse, H.-J. 1980. *Das Dullay: Materialen zu einer ostkuschitischen Sprache*. Berlin: Reimer.
Appleyard, D.L. 1984. The internal classification of the Agaw languages: a comparative and historical phonology. *Current Progress in Afro-Asiatic Linguistics: Papers of the Third International hamito-Semitic Congress*, Ed. by J. Bynon, 33–67. Amsterdam/Philadelphia: Benjamins.
Appleyard, David L. 1986. Agaw, Cushitic and Afroasiatic: the personal pronoun revisited. *Journal of Semitic Studies* 31. 195–236.
Appleyard, David L. 1989. Agaw and Omotic Links: The Evidence of the Lexicon. Paper presented at *2nd International Symposium on Cushitic and Omotic Languages*, held in Turin, Nov. 1989.

Appleyard, David L. 2006. *A Comparative Dictionary of the Agaw Languages.* Köln: Köppe.
Bechhaus-Gerst, M. & Serzisko, F. (Eds) 1988. *Cushitic-Omotic. Papers from the International Symposium on Cushitic and Omotic Languages* (Cologne, Jan. 1986). Hamburg: Buske.
Beke, C.T. 1845. On the Languages and Dialects of Abyssinia and the Countries to the South. *Proceedings of the Philological Society* 2: 33. 89–107.
Bender, M.L. 1971. The Languages of Ethiopia: A New Lexicostatistical Classification and some Problems of Diffusion. *Anthropological Linguistics* 13: 5. 165–288.
Bender, M.L. 1975. *Omotic: A new Afroasiatic language family*. Carbondale (Ill.): University Museum Studies 3.
Bender, M.L. (Ed.) 1976. *The Non-Semitic Languges of Ethiopia*. East Lansing, Mich.: African Studies Center, Michigan State University.
Bender, M.L. 1988. Proto-Omotic Phonology and Lexicon. *CO*. 121–159.
Bender, M.L. 1990. The limits of Omotic. *OLS*. 584–616.
Bender, M.L. 1994a. Aroid (South Omotic) Lexicon. *AAP* 38. 133–162.
Bender, M.L. 1994b. The Mystery Languages of Ethiopia. *New Trends in Ethiopian Studies. Papers of the 12th International Conference of Ethiopian Studies* (Sept 1994), Eds. H.G. Marcus & G. Hudson, 1153–1174. Lawrenceville: Red Sea Press.
Bender, M.L. 2000. *Comparative Morphology of Omotic Languages*. München: Lincom Europa.
Bender, Lionel M. 2003. *Omotic Lexicon and Phonology*. Carbondale: Southern Illinois University.
Benton, P.A. 1912. *The Languages and Peoples of Bornu*, Vol. I. London: Cass.
Biella, Joan C. 1982. *Dictionary of Old South Arabic. Sabean Dialect*. Chico, Calif.: Scholars Press.
Black, Paul D. 1974. *Lowland East Cushitic: Subgrouping and Reconstruction*. Yale University: Ph.D. Dissertation.
Blažek, Václav. 1989. Omotic Lexicon in Afroasiatic Perspective: Body Parts Cognates. Paper presented at *2nd International Symposium on Cushitic and Omotic Languages*, held in Turin, Nov. 1989.
Blažek, Václav. 1995. The microsystems of personal pronouns in Chadic, compared with Afroasiatic. *Studia Chadica et Hamitosemitica: Akten des Internationalen Symposions zur Tschadsprachenforschung* (Frankfurt am main, May 1991), Ed. by Dymitr Ibriszimow & Rudolf Leger, 36–57. Köln: R. Köppe.
Blažek, Václav. 1999. Numerals. Comparative analyses and their implications. Brno: Masarykova univerzita.
Blažek, Václav. 2005. Natural Phenomena, Time and Geographical Terminology in Beja Lexicon. Fragment of a Comparative and Etymological Dictionary of Beja . *Memoriae Igor M. Diakonoff*, Ed. L. Kogan et al., 365–407. Winona Lake (Indiana): Eisenbrauns (Babel und Bibel 2).
Blažek, Václav & Boisson, Claude. 1992. The Diffusion of Agricultural Terms from Mesopotamia. *Archív orientální* 60: 1. 16–37.
Brenzinger, M. 1995. The "Islanders" of Lake Abaya and Lake Ch'amo: *Harro, Ganjule, Gats'ame* and *Bayso*. *SLLE* 26. 2–32.
Brockelmann, Carl. 1908. *Grundriss der vergleichenden Grammatik der semitischen Sprachen*, I. Berlin: Reuther & Reichard.
Cerulli, Enrico. 1929. Note su alcuni popolazioni sidāmā dell' Abyssinia. *Rivista degli Studi Orientali* 12. 1–70.

Cerulli, Enrico. 1938a. *Studi etiopici* II: *La lingua e la storia dei Sidamo*. Roma: Istituto per l'Oriente.

Cerulli, Enrico. 1938b. *Studi etiopici III: Il linguaggio dei Giangerò ed alcune lingue Sidama dell'Omo (Basketo, Ciara, Zaissè)*. Roma: Istituto per l'Oriente.

Cerulli, Enrico. 1951. *Studi etiopici* IV: *La lingua caffina*. Roma: Istituto per l'Oriente.

Chiomio, G. 1938. *Brevi appunti di lingua Wollamo: grammatica e dizionario.* Torino: Istituto Missioni Consolata.

Cohen, David (et alii). 1970f. *Dictionnaire des racines sémitiques*. Paris (-La Haye): Mouton.

Cohen, Marcel. 1947. *Essai comparatif sur le vocabulaire et la phonétique du chamito-sémitique.* Paris: Klincksieck.

Conti Rossini, Carlo. 1904. Appunti sulla lingua khamta dell'Averghelle. *Giornale della Società Asiatica Italiana* 17. 183–242.

Conti Rossini, Carlo. 1905. Appunti sulla lingua awiya del Danghelà. *Giornale della Società Asiatica Italiana* 18. 103–194.

Conti Rossini, Carlo. 1912. *La langue des Kemant en Abyssinie*. Wien: Hölder (Kaiserliche Akademie der Wissenschaften, Schriften der Sprachkommission).

Conti Rossini, Carlo. 1925. Sui linguaggi dei Naa e dei Ghimirra (Sce) nell'Etiopia meridionale. *Rendiconti della Reale Accademia dei Lincei*, classe di scienze morali, storiche e filologiche, Vol. 1, Ser. 6, 612–636.

Conti Rossini, Carlo. 1927. Sui linguaggi parlati a nord dei laghi Rodolfo e Stefania. *Festschrift Meinhof*, Ed. F. Boas et al., 247–255. Glückstadt - Hamburg: Augustin.

Conti Rossini, Carlo. 1936. Contributi per la conoszenza della lingua Haruro (Isole del Lago Margherita). *Rendiconti della Reale Accademia dei Lincei*, classe di scienze morali, storiche e filologiche, ser. VI, vol. XII, fasc. 7–10, 621–679.

Cowley, R. et al. 1971. The Kunfäl people and their language. *Journal of Ethiopian Studies* 9. 99–106.

Dolgopolsky [Dolgopoľskij], A.B. 1973. *Sravniteľno-istoričeskaja fonetika kušitskix jazykov*. Moskva: Nauka.

Ehret, Christopher. 1980. *The Historical Reconstruction of Southern Cushitic Phonology and Vocabulary.* Berlin: Reimer.

Ehret, Christopher. 1981. Revising Proto-Kuliak. *Afrika und Übersee* 64.81–100.

Ehret, Christopher. 1987. Proto-Cushitic Reconstruction. *Sprache und Geschichte in Afrika* 8. 7–180.

Ehret, Christopher. 1991. The Consonant Inventory of Proto-Eastern Cushitic. *Studies in African Linguistics* 22: 3. 211–275.

Ehret, Christopher. 1995. *Reconstructing Proto-Afroasiatic* (*Proto-Afrasian*). *Vowels, Tone, Consonants and Vocabulary*. Berkeley – Los Angeles – London: University of California Press.

Ehret, Christopher. 2000. *A Historical-Comparative Reconstruction of Nilo-Saharan.* Köln: Köppe.

Ehret, Christopher & Nuuh, Ali M. 1984. Soomaali Classification. In: *Proceedings of the Second International Congress of Somali Studies*, Vol. I., Ed. T. Labahn, 201–269. Hamburg: Buske.

Erman, J.P.A. & Grapow, H. Eds. 1926–1953. *Wörterbuch der aegyptischen Sprache im Auftrage der deutschen Akademien*. 6 vols. Leipzig: J.C. Hinrichs'schen Buchhandlungen. (Reprinted Berlin: Akademie-Verlag GmbH, 1971).

Fleming, Harold C. 1964. Baiso and Rendille: Somali Outliers. *Rivista degli Studi Etiopici* 20. 35–96.

Fleming, Harold C. 1965. *The Age-Grading Cultures of East Africa: A Historical Inquiry.* Ph.D. Diss.

Fleming, Harold C. 1969a. Asa and Aramanik: Cushitic Hunters in Masai-Land. *Ethnology* 8: 1. 1–36.

Fleming, Harold C. 1969b. The Classification of West Cushitic within Hamito-Semitic. *Eastern African History*, Ed. D.F. McCall et al., 3–27. New York: Praeger.

Fleming, Harold C. 1971. *Wordlists of Dasenech, Doko, Galila, Ganjule, Gofa, Malo, Oyda, Ubamer* (unpublished field notes kindly offered to the author).

Fleming, Harold C. 1974. Omotic as an Afroasiatic Family. *Studies in African Linguistics*, Supplement 5.81–94.

Fleming, Harold C. 1975. Recent Research in Omotic-Speaking Areas. *Proceedings of the First United States Conference on Ethiopian Studies*, Ed. H.G. Marcus, 261–278. East Lansing: Michigan State University.

Fleming, Harold C. 1976a. Omotic Overview. Bender 1976. 298–323.

Fleming, Harold C. 1976b. Kefa (Gonga) Languages. Bender 1976. 351–376.

Fleming, Harold C. 1983. Kuliak External Relations: Step One. *Nilotic Studies. Proceedings of the International Symposium on Languages and History of the Nilotic People* (Cologne, Jan 1982), Eds. R. Vossen & M. Bechhaus-Gerst, 423–477. Berlin: Reimer.

Fleming, Harold C. 1984. The Importance of Mao in Ethiopian History. In: *Proceedings of the Seventh International Conference of Ethiopian Studies*, Ed. S. Rubenson, 31–98. Uppsala: Scandinavian Institute of African Studies,.

Fleming, Harold C. 1987. Proto-Gongan Consonant Phonemes: Stage One. *Leo Reinisch – Werk und Erbe*, Ed. H.G. Mukarovsky, 141–159. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Fleming, Harold C. 1988a. Proto-South-Omotic or Proto-Somotic Consonant Phonemes: Stage One. *CO.*163–175.

Fleming, Harold C. 1988b. Mao's Ancestor. Consonant Phonemes of Proto-Mao. Stage One. In: *Proceedings of the Ninth International Congress of Ethiopian Studies* (Moscow, Aug. 1986), Vol. 5, Ed. A.A. Gromyko, 35–44. Moscow: Nauka.

Fleming, Harold C. 1990. A Grammatical Sketch of Dime (Dim-Af) of the Lower Omo. *OLS.*494–583.

Greenberg, Josph H. 1963. *The Languages of Africa.* Bloomington: *International Journal of American Linguistics* 29.

Greenberg, Joseph H. 1963a. The Mogogodo, a forgotten Cushitic people. *Journal of African Languages* 2.29–43.

Gregg, Gene B. 1982. *Oromo Dictionary*. East Lansing: African Studies Center.

Grottanelli, V.L. 1940. *Missione etnografica nel Uollega occidentale*, I: *I Mao*. Roma: Reale Accademia D'Italia.

H = Hayward, Dick. 1988. Remarks on the Omotic Sibilants. In: *CO*, 261–299.

Haberland, Eike. 1966. Zur Sprache der Bodi, Mursi und Yidenič Südwest-Äthiopien. In: *Neue Afrikanistische Studien*, Ed. J. Lukas. Hamburg: Deutsches Institut für Afrika-Forschung, 88–99.

Haberland, Eike & Lamberti, Marcello. 1988. *Ibaaddo ka-Ba'iso: culture and language of the Ba'iso*. Heidelberg: Winter.

Habte Mikael, Habte Wold. 1989. *English Kaffinya Dictionary*. Addis Abeba: Lazarist School.

Hayward, R.J. 1979. Bayso Revisited: Some Preliminary Linguistic Observations, II. *BSOAS* 62.101–132.

Hayward, R.J. 1982. Notes on the Koyra Language. *Afrika und Übersee* 65. 211–268.
Hayward, R.J. 1984. *The Arbore Language*. Hamburg: Buske.
Hayward, R.J. 1989. Comparative notes on the language of S'aamakko. *Journal of African Languages and Linguistics* 1. 1–53.
Hayward, R.J. 1990a. Notes on the Aari language. *OLS*. 425–493.
Hayward, R.J., Ed. 1990b. *Omotic Language Studies*. London: SOAS.
Heine, Bernd. 1975. Notes on the Yaaku Language (Kenya). *Afrika und Übersee* 58. 119–138.
Heine, Bernd. 1976. Notes on the Rendille Language. *Afrika und Übersee* 59. 176–223.
Heine, Bernd. 1978. The Sam Languages. A History of Rendille, Boni and Somali. *Afroasiatic Linguistics* 6: 2. 23–115.
Heine, Bernd. 1980/82. *Language and Dialect Atlas of Kenya,* Vol. II: *The Non-Bantu Languages of Kenya* (pp. 173–18 *Elmolo*); Vol. X: *Boni Dialects.*
Hetzron, Robert. 1969. *The Verbal System of Southern Agaw*. Berkeley – Los Angeles: University of California Press.
Hetzron, Robert. 1978. The nominal system of Awngi (Southern Agaw). *BSOAS* 41. 121–141.
Hetzron, Robert. 1988. Position of Omotic. *CO.*107–117.
Hudson, Grover. 1989. *Highland East Cushitic Dictionary*. Hamburg: Buske.
Hudson, Richard. 1996. *A Dictionary of Beja*. Draft printout prepared by Roger Blench, Ms.
Illič-Svityč, V.M. 1971/76/84. *Opyt sravnenija nostratičeskix jazykov*, I-III. Moskva: Nauka.
Jungraithmayr, H. & Ibriszimow, D. 1994_{1-2}. *Chadic Lexical Roots*, vol. I-II. Berlin: Reimer.
Jungraithmayr, Herrmann. 1970. *Die Ron-Sprachen. Tschadohamitische Studien in Nordnigerien.* Glückstadt: Augustin.
Jungraithmayr, Herrmann. 1990. *Lexique mokilko*. Berlin: Reimer.
Jungraithmayr, Herrmann & Adams, A. 1992. *Lexique migama*. Berlin: Reimer.
Jungraithmayr, Herrmann & Alio, Khalil. 1989. *Lexique bidiya*. Frankfurt am Main: Klostermann.
Kiessling, R. 1994. *Eine Grammatik des Burunge*. Hamburg: Research and Progress Verlag.
Kiessling, Roland & Mous, Maarten. 2003. *The Lexical Reconstruction of West-Rift Southern Cushitic.* Köln: Köppe.
Kraft, Charles. 1981. *Chadic Wordlists*. Berlin: Reimer.
Lamberti, Marcello. 1986. *Die Somali-Dialekte: Eine vergleichende Untersuchung.* Hamburg: Buske.
Lamberti, Marcello. 1987. Some Konsoid Etymologies. *Anthropos* 82. 529–541.
Lamberti, Marcello. 1988. *Kuliak and Cushitic. A comparative study.* Heidelberg: Winter.
Lamberti, Marcello. 1992. Some Phonetic Laws of the Gonga Languages, I. *Rassegna di Studi Etiopici* 36. 57–76.
Lamberti, Marcello. 1993a. *Die Shinassha-Sprache: Materialien zum Boro.* Heidelberg: Winter.
Lamberti, Marcello. 1993b. *Materialien zum Yemsa.* Heidelberg: Winter.
Lamberti, Marcello. 1993c. Some Phonetic Laws of the Gonga Languages, I. *Rassegna di Studi Etiopici* 37. 89–114.
Lamberti, Marcello. 1993d. The Ari-Banna Group and its Classification. *Studi Italiani di Linguistica Teorica e Applicata* 22: 1. 39–87.
Lamberti, Marcello & Sottile, Roberto. 1997. *The Wolaytta Language* (Studia Linguarum Africae Orientalis, Bd. 6). Köln: R. Köppe Verlag.
Laoust, E. 1931. *Siwa* I: *Son parler*. Paris: Leroux.
Leslau, Wolf. 1959. *A Dictionary of Moča* (*Southwestern Ethiopia*). Berkeley – Los Angeles: University of California Press, Publications in Linguistics, Vol. 18.

Leslau, Wolf. 1963. *Etymological Dictionary of Harari*. Berkeley-Los Angeles: University of California Press.

Leslau, Wolf. 1979. *Etymological Dictionary of Gurage* (*Ethiopic*), vol. III. Wiesbaden: Harrassowitz.

Leslau, Wolf. 1980. Proto-*Sidamo **z. Afrika und Übersee* 63. 119–129.

Leslau, Wolf. 1987. *Comparative Dictionary of Geʿez* (*Classical Ethiopic*). Wiesbaden: Harrassowitz.

Lukas, Johann. 1937. *Zentralsudanische Studien*. Hamburg: *Abhandlungen aus dem Gebiet der Auslandskunde*, Hansische Universität, Reihe B, Bd. 45/24.

Militarev, A.Y. 1991. Istoričeskaja fonetika i leksika livijsko-guančskix jazykov. In: *Jazyki Azii i Afriki* IV.2. Moskva: Nauka, 249–267.

Militarev, Alexander & Kogan, Leonid. 2000–2005. *Semitic Etymological Dictionary*, Vol. I: *Anatomy of Man and Animals*; Vol. II: *Animal Names*. Münster: Ugarit-Verlag.

Mouchet, J. 1950. Vocabulaire comparatifs de quinze parlers du Nord-Cameroun. *Études camerounaises* III:29–30. 5–74.

Mouchet, J. 1953. Vocabulaires comparatifs de sept parlers du Nord-Cameroun. *Études camerounaises* VI.41–42, 137–199.

Mukarovsky, Hans G. 1981. Wo steht das Saharanische? *Afrika und Übersee* 64. 187–226.

Mukarovsky, Hans G. 1989. On the relations of Cushitic, Omotic and Chadic languages. Paper presented at *2nd International Symposium on Cushitic and Omotic Languages*, held in Turin, Nov. 1989.

Müller, Walter W. 1975. Beiträge zur hamito-semitischen Wortvergleichung. In: *Hamito-Semitica, Proceedings of a Colloquium held by the Historical section* (London, 1970), Ed. J. & T. Bynon, 33–50. The Hague – Paris: Mouton.

Newman, Paul. 1964. A Word List of Tera. *Journal of West African Languages* 1: 2.33–50.

Newman, Paul. 1977. Chadic Classification and Reconstruction. *Afroasiatic Linguistics* 5: 1.1–42.

Newman, Paul & Ma, Roxana. 1966. Comparative Chadic: Phonology and Lexicon. *Journal of African Languages* 5: 3. 218–251.

Naït-Zerrad, Kamal. 1998–2002. *Dictionnaire des racines berbères*, I-III. Paris-Louvain: Peeters.

Palmer, F.R. 1958. The noun in Bilin. *BSOAS* 21. 377–391.

Parker, E. & Hayward, R.J. 1985. *An Afar-English-French Dictionary*. London: SOAS.

Pillinger, S. & Galboran, L. 1999. *A Rendille Dictionary*. Köln: Köppe.

Prasse, Karl-G. 1972–73–74. *Manuel de grammaire touaregue* (*tăhăggart*). Copenhague: Akademisk Forlag.

Rabin, Chaim. 1981. Ron-Semitic Etymologies. *The Chad Languages in the Hamitosemitic/ Nigritic border area*, Ed. H. Jungraithmayr, 24–31. Berlin: Reimer.

Rapold, Christian. 2005. *Towards a grammar of Benchnon*. Leiden: Ph.D. Diss.

Reinisch, Leo. 1884. Die Chamirsprache in Abessinien, II: Chamir-deutsches Wörterbuch. *SAW*, phil.-hist. Cl. 106. 330–450.

Reinisch, Leo. 1885. Die Quarasprache in Abessinien, II: Quarisch-deutsches Wörterbuch. *SAW*, phil.-hist. Cl. 109: 1. 3–152.

Reinisch, Leo. 1887. *Wörterbuch der Bilin Sprache*. Wien: Hölder.

Reinisch, Leo. 1888. Die Kafa-Sprache in Nordost-Afrika, II: Kafa-Deutsches Wörterbuch. *SAW*, phil.-hist. Cl. 116. 251–386.

Reinisch, Leo. 1890. *Wörterbuch der Saho-Sprache*. Wien: Holder.

Reinisch, Leo. 1895. *Wörterbuch der Bedawye-Sprache*. Wien: Hölder.
Reinisch, Leo. 1902. *Die Somali-Sprache*, II: *Wörterbuch*. Wien: Hölder.
Roper, E.-M. 1928. *Tu Beḍawye. An Elementary Handbook for the Use of Sudan Government Officials*. Hertford: Austin & Sons.
Sasse, Hans-Jürgen. 1973. Spuren von Pharyngalen im Galab. *Afrika und Übersee* 56. 266–275.
Sasse, Hans-Jürgen. 1975. Galla /š/, /s/ un /f/. *Afrika und Übersee* 58. 244–263.
Sasse, H.-J. 1976. Weiteres zu den ostkuschitischen Sibilanten. *Afrika und Übersee* 59. 125–142.
Sasse, H.-J. 1979. The consonant phonemes of Proto-Eastern-Cushitic (PEC): A first approximation. *Afroasiatic Linguistics* 7: 1. 1–67.
Sasse, H.-J. 1982. *An Etymological Dictionary of Burji.* Hamburg: Buske.
Sasse, H.-J. 1993. Ein weltweites Hundewort. *Sprachen und Schriften des antiken Mittelmeerraums. Festschrift für Jürgen Untermann*. Hrsg. F. Heidermanns, H. Rix und E. Seebold, 348–366. Innsbruck: IBS 78.
Shimizu, Kiyoshi. 1978. The Southern Bauchi Group of Chadic Languages. *Africana Marburgensia*, Special Issue 2.
Siebert, R. 1995a. A Survey of the Male Language. *SLLE* 24. 1–12.
Siebert, R. 1995b. A Survey of the *Dime* Language. *SLLE* 31. 1–14.
Siebert, R. & Caudwell, S. 1995. On Melo (Malo) and Mursi. *SLLE* 27.1–24.
Siebert, R. & Hoeft, L. 1994. Languages of the Abbaya/Chamo Area – Report Part I. *SLLE* 21. 1–27.
Siebert, R. & K. & Wedekind, Ch. 1993. Survey on Languages of the Asosa - Begi - Komosha Area. *SLLE* 11. 1–22.
Siebert, R. & Wedekind, K. 1994. Third S.L.L.E. Survey on Languages of the Begi/Asosa Area. *SLLE* 15. 1–18.
Skinner, Neil. 1977. North Bauchi Chadic Languages: Common Roots. *Afroasiatic Linguistics* 4: 1. 1–49.
Skinner, N. 1996. *Hausa Comparative Dictionary.* Köln: Köppe.
Steingass, F. 1988. *Arabic-English Dictionary*. Jalandhar City: Gaurav.
Stolbova, Olga V. 1987. Sravniteľno-istoričeskaja fonetika i slovaŕ zapadnočadskix jazykov. *Afrikanskoe istoričeskoe jazykoznanie,* 30–268. Moskva: Nauka.
Stolbova, Olga V. 1996. *Studies in Chadic Comparative Phonology*. Moscow: Diaphragma.
Stolbova, Olga V. 2005. *Chadic Lexical Database*, Issue I: **L**, **N**, **Ny**, **R**. Kaluga: Poligrafiya.
Takács, Gábor. 1996. Egyptian lexics in an Afrasian perspective: New etymologies. *Studia Etymologica Cracoviensia* 1. 125–180.
Takács, Gábor. 1999–2001. *Etymological Dictionary of Egyptian*, I-II. Leiden – Boston – Köln: Brill.
Tosco, Mauro. 1991. *A Grammatical Sketch of Dahalo*. Hamburg: Buske.
Tosco, Mauro. 2001. *The Dhaasanac Language*. Köln: Köppe.
Vycichl, Werner. 1934. Hausa und Ägyptisch. *Mitteilungen des Seminars für Orientalische Sprachen* 37, Abt. 3. 36–117.
Vycichl, Werner. 1955. Der Umlaut in der Berbersprachen Nordafrikas. *WZKM* 52.304–325.
Vycichl, Werner. 1978. L'état actuel des études chamito-sémitiques. *Atti del Secondo Congresso Internazionale di Linguistica Camito-Semitica* (Firenze 1974), Ed. by P. Fronzaroli, 63–76. Firenze 1978 (Quaderni di Semitistica 5).
Vycichl, Werner. 1983. *Dictionnaire étymologique de la langue copte.* Leuven: Peeters.
Wedekind, K. 1990. Gimo-Jan or Ben-Yem-Om: Benč, – Yemsa phonemes, tones, and words. *OLS*.68–184.

Welmers, W.E. 1952. Notes on the structure of Saho. *Word* 8.145–162, 236–251.
Woelfel, Dominik J. 1955. *Eurafrikanische Wortschatzten als Kulturschichten*. Salamanca: Acta Salmanticensia, Filosofía y Letras, T. IX, núm. 1.
Woelfel, Dominik J. 1965. *Monumenta Linguae Canariae*. Graz: Akademische Druck.
Yigezu, Moges. 2004. The vowel system of Kara with comparative-historical notes. Paper presented at the 7th *International Samito-Hamitic Congress* (Berlin: Sept. 13–15, 2004).
Zaborski, Andrzej. 1983. Basic Numerals in the Omotic Languages. In: *Ethiopian Studies dedicated to W. Leslau* (75th birthday), Eds. St. Segert & A.J.E. Bodrogligeti, 375–390. Wiesbaden: Harrassowitz.
Zaborski, Andrzej. 1984–86. A Note on Cushitic Demonstrative Pronouns. *Orientalia Suecana* 33–35, 505–511.
Zaborski, Andrzej. 1987. Basic Numerals in Cushitic. *Proceedings of the Fourth International Hamito-Semitic Congress* (Marburg, Sept. 1983), Ed. H. Jungraithmayr & W.W. Müller, 317–347. Amsterdam – Philadelphia: Benjamins.
Zaborski, Andrzej. 2004. West Cushitic – A genetic reality. *Lingua Posnaniensis* 46. 173–186.

The primary branches of Cushitic

Seriating the diagnostic sound change rules

Christopher Ehret
University of California, Los Angeles

This article situates Dahalo and Southern Cushitic within an overall seriation of the early phonological evolution of each of the Cushitic sub-branches. A sequence of nine diagnostic sound changes solidly establishes Dahalo's membership in the Southern Cushitic branch. Different sound change histories characterized the evolution of each of the proto-languages of Eastern Cushitic, Agaw, and Beja. Seriating the sound change histories in each branch reveals that the presence of the same or similar reflexes of certain consonants in different branches does not reflect closer branch relationships, but came about separately through the operation of differently ordered and structured sound change rules.

This article revisits the phonological evidence first clearly set forth more than twenty-five years ago (Ehret 1980), and presented with additional detail twenty years ago (Ehret 1987: 146–148), that the Dahalo language belongs to the Southern Cushitic sub-branch of the Afrasian (Afroasiatic) language family. Such an effort should, by this point in time, be a redundant exercise, but the continuing disregard of, or perhaps simply lack of acquaintance with, the relevant published sources has meant that certain scholars have continued to misconstrue Dahalo's relationships. To bring closure to this issue, the article situates Dahalo and Southern Cushitic within an overall seriation of the early phonological evolution of each of the Cushitic sub-branches.

A by-product of this overall seriation is a new perspective on the interrelationships of the four deep branchings of Cushitic: Northern Cushitic (represented today solely by the Beja language), Agaw, Eastern Cushitic, and Southern Cushitic. In particular, one of the features of Cushitic phonological history, already noted in this author's reconstruction of the Cushitic branch as a whole (Ehret 1987), was a trend in the Cushitic languages of the northern Horn of Africa toward the development of phonetically similar modern-day reflexes for a number of the consonant phonemes of proto-Cushitic. In contrast, a greater preservation of ancient features of articulation occurred in southern Horn languages and, above all, in the Southern Cushitic languages spoken still farther south, in East Africa. Clearly, this trend in northerly

Cushitic-speaking areas reflected areal influences of recent millennia, rather than ancient reconstructible Cushitic commonalities. The demonstration of the different histories of sound change that lay behind the similar outcomes in the different branches was left aside in previous publications (Ehret 1987), but is taken up here.

We first reprise the shared sound change histories that firmly locate Dahalo within Southern Cushitic. The set of diagnostic sound changes characteristic of the Southern Cushitic branch, which occur in Dahalo and resoundingly placing Dahalo in Southern Cushitic, have already been published, but they will be presented again here in restated and updated versions. We will then sketch out the systemic patterns of consonantal sound change in each of the major divisions of Cushitic. These patterns reveal that separate courses of sound change took place in each branch, strikingly different from those that characterize Southern Cushitic and Dahalo as a member of Southern Cushitic.

The proto-Cushitic consonant system (Figure 1) consisted of 38 or 39 consonants (Ehret 1995):

b	d	dz	dl		g	g^w	ʕ
p	t	ts			k	k^w	ʔ
p’	t’	ts’	tl’	c’	k’	k^w’	
f	s		ɬ	ʃ	x	x^w	ħ
	z						
m	n			ɲ	ŋ	$ŋ^w$ (ʔ)	
w	l, r			y			h

Figure 1. The proto-Cushitic consonant system.

The historically most notable sound changes in the different branches affected especially the glottalic (originally, ejective) and lateral obstruent consonants of proto-Cushitic, often yielding resemblant outcomes in the proto-languages of the branches. But the evolution of the changes took place in different systemic contexts and were the outcomes of differently ordered courses of sound change. On the whole this exposition leaves aside vowel sound changes and concentrates on the ordered sound changes affecting consonants. Consonant sound change seriations provide the more telling diagnostic evidence for the branches, although vowel sound changes in some cases must be evoked because they fit within these sequences.

1. Sound changes leading down to proto-Southern Cushitic

Having viewed the consonant system of proto-Cushitic, we are ready to move on to the defining sound changes of proto-Southern Cushitic, attested in the Dahalo evidence as well as the evidence of the rest of the languages of the Southern Cushitic branch.

All of these rules have been previously published (Ehret 1987), but it is clear that some scholars have remained unaware of these data.

Before proceeding further, we need to significantly revise the proto-Southern Cushitic (PSC) consonant system as presented in Ehret 1980. The bases of the revision cannot be adequately argued here, but a statement of what remains valid can be presented. The system in Ehret 1980 includes, along with a large number of quite valid reconstructions, several postulations that should not have appeared there. In particular, the supposed PSC retroflex series, with the exception of voiced *ɖ, should be dropped. The existence of voiced and unvoiced palatal stops in PSC (*d^{y} and *t^{y}') is also not supported. Additionally, the prenasal series of PSC must be reinterpreted as consonant clusters in the light of the wider Cushitic and Afrasan reconstructions now available (Ehret 1987, 1995). These items can now be seen to derive from a morphological operation, the addition to stems of an old prefix *in-. This prefix, traceable at least to proto-Cushitic and found in Agaw as well as Beja, was still productive down to recent times in the West Rift languages, in the alternant shapes *in- ~ *hin-.

One solidly reconstructed segment, represented as *ɖ in Ehret 1980, presents a different problem, in that it surely was not a retroflex. This segment coexisted in the PSC consonant system with an equally solidly reconstructed voiced stop previously rendered as PSC *d. Two ways of accounting for the existence of two alveolar/dental voiced stops in PSC are possible. One is that *ɖ was an implosive /ɗ/, as it is today in its Dahalo reflex, and that *d was indeed non-implosive /d/. The other is that the two consonants were, respectively alveolar *d, either implosive or not, and dental *d̪. Again Dahalo evidence is relevant here, in that the Dahalo reflex of PSC *d is indeed the non-implosive dental [d̪]. For the nonce, I treat the distinction as implosive versus non-implosive, representing the former *ɖ (Ehret 1980) as *ɗ and PSC *d as still *d. In addition, the simpler *c' replaces previous notation *t^{y}' for the PSC palatal ejective.

With these considerations taken into account, we are left with the following PSC pattern (Figure 2):

	ɗ						
b	d	dz	dl		g	g^{w}	ʕ
p	t	ts			k	k^{w}	ʔ
p'	t'	ts'	tl'	c'	k'	k^{w}'	
f	s		ɬ	ʃ	x	x^{w}	ħ
m	n			ɲ	ŋ	ŋ^{w} (ʔ)	
w	l, r			y			h

Figure 2. The consonants of proto-Southern Cushitic.

This system reshaped the proto-Cushitic consonant system in two notable respects, removing the unbalanced existence of a solitary voiced fricative *z and adding at least one implosive consonant. In Dahalo the inherited PC *b produces an implosive

outcome in word-initial contexts, so it is possible that PSC *b may have been implosive as well.

The most striking and compelling demonstration of the membership of Dahalo in Southern Cushitic is an interlocking set of four sound changes dating to the period of the separate evolution of the proto-Southern Cushitic (PSC) language out of its earlier Cushitic ancestry. This set of shifts occurred in uniquely delimited, intertwining environments. Listed after each rule are the specific reference numbers of the PC and PSC roots (in Ehret 1987) the Dahalo reflexes of which (in Ehret 1980) attest those particular sound changes:

1. PC *s > pre-PSC *z /V_V (but remained *s word-initial and verb-stem final) (Ehret 1987: roots 233, 234, 236).
2. PC *d > PSC *ɗ /_VC (i.e., everywhere except stem-final) (Ehret 1987: non-stem-final: roots 17, 18, 22, 23, 26, 32, 35; stem-final (> *d): roots 30, 31).

The new pre-PSC *z created by rule 1, along with original PC *z, were subsequently converted to stops by the third and fourth sound shifts:

3. Pre-PSC *z > PSC *d /#_V and noun-stem-final (Ehret 1987: word-initial: roots 196, 200, 201, 203, 270; noun-stem-final: 214, 216, 233, 234, 236).
4. Pre-PSC *z > PSC *ɗ elsewhere (/V_VC and verb-stem-final) (Ehret 1987: roots 207, 209, 210, 212, 217, 219, 222, 366).

The third rule created new non-implosive *d in environments in which rule #1 had removed PC *d by changing it to PSC *ɗ, such that a push-pull effect seems present here; and the fourth rule partially filled a remaining gap in the distribution of *ɗ by producing verb-stem-final instances of it. A fifth rule can be placed subsequent to this series because, in at least one case, it removed an environment in which rules 2 and 4 operated:

5. PC *b > Ø /#CVCV_# (Ehret 1987: roots 32, 61, 481).

The full participation of Dahalo in such an intricately interlocking set of sound change rules puts the membership of Dahalo in Southern Cushitic beyond doubt.

Several additional PSC sound change rules that operated in labiovelar environments are evident in Dahalo as well as the other Southern Cushitic languages:

6. PC *C^w > PSC *C /[+round/+vocalic]_-# (stem-final) (Ehret 1987: roots 100, 307, 310, 400, 418, 516).
7. PC *C^w > PSC *C /(a)i_-# (stem-final), C^w = obstruent (Ehret 1987: root 364).

8. PC *C^wi(i) > PSC *Cɨ(ɨ) /#(Ci)_ (Ehret 1987: roots 32, 61, 236, 281, 418).

Finally, Dahalo shared in the PSC merger of the PC voiced velar fricatives with their voiceless equivalents. The ordering of this sound change with respect to the other rules is not self-evident:

9. PC *ɣ$^{(w)}$ > PSC *x$^{(w)}$ (Ehret 1987: roots 257, 258 [typographical error in both Ehret 1987 and 1980: should be *ʕaaʔ- instead of *ɣaaʕ-], 262, 270, 272, 273, 277, 279, 281).

Once these sound changes are taken account of, one of the mistakes that has been made in the past – of inflating the cognation counts of Dahalo with Eastern Cushitic by counting loanwords as cognates – can quickly be corrected. At least five Eastern Cushitic loanwords have penetrated the most basic vocabulary of Dahalo, one of them borrowed from the Garree branch of the Soomaali group in the early second millennium CE (*liima* "two") and four of them from Oromo in the past four centuries (*ɗiiga* "blood"; *ʔado* "sun"; *sina* "nose"; and *k'oro* "tree"). With these loanwords deleted, the valid percentages of retained cognates of Dahalo in basic vocabulary with the two best-known Eastern Cushitic languages, Oromo and Soomaali, fall to around 9–10 per cent, in keeping with the cognation range of the other Southern Cushitic languages with Eastern Cushitic (for instance, Iraqw has 9 per cent cognation in the 100-word list with Soomaali). In contrast, the Dahalo cognate range in basic vocabulary with the other Southern Cushitic languages runs consistently and strikingly higher, at 15–22 per cent (Ehret 1980: 20, 385–388), in keeping with the testimony of the sound change histories, which fit Dahalo integrally into the Southern Cushitic branch.

2. Northern Cushitic: Beja (Bedawiye)

Beja, consisting today of a number of fairly divergent dialects spoken from the Red Sea hills of farthest southeastern Egypt to the lowlands of northern Eritrea, is the only extant language of the Northern Cushitic branch, although we may suspect that other members of the branch existed in much earlier times (Ehret, forthcoming). A sequence of three straightforward, simple sound changes account most parsimoniously for the Beja outcomes of the proto-Cushitic (PC) ejective consonants. The first of the shifts merged the PC ejective *ts' with the PC palatal ejective *c'. The PC dental stop ejective then spirantized, filling the slot vacated by PC *s'. Logically subsequent to both rules came the simple deletion of ejection:

1. PC *ts' > pre-Beja *c'.
2. PC *t' > pre-Beja *ts'.
3. *C' > C.

As a result of rule #3,

pre-Beja *ts' > (later) pre-Beja *ts,
PC *tl' > pre-Beja *tl,
PC *p' > pre-Beja *p,
PC *c' > pre-Beja *c,
PC *k' > Beja /k/,
PC k^{w}' > Beja /k^{w}/.

Following sound change #3, a further highly natural rule, #4, then explains the outcomes of the two PC lateral affricates:

4. [+lateral/+obstruent] > [+retroflex/+obstruent];

i.e., PC *dl > /ɖ/, pre-Beja *tl (< PC *tl' by rule #3) > /ʈ/. Two more sound changes, similarly postdating rule #3, would account for the removal of the remaining affricates from the language and for the collapsing of the alveolar/dental and palatal ejectives with the equivalent fricatives:

5. PC *dz > pre-Beja *ts.
6. pre-Beja *ts > Beja /s/, pre-Beja *c > Beja /ʃ/.

At least one further rule also can be placed tentatively after rule #3, although its ordering relative to #4 is not self-evident:

7. pre-Beja *p > Beja /b/.

A separate sequence of two sound change rules, of uncertain ordering with respect to rules #1–4, together removed the category of velar fricatives from Beja. The first collapsed the four PC velar fricatives to just two by devoicing PC *ɣ and *$ɣ^{w}$; the second merged the two resulting voiceless fricatives with the two existing voiceless velar stops inherited from PC:

8. PC *$ɣ^{(w)}$ > *$x^{(w)}$.
9. pre-Beja *$x^{(w)}$ > *$k^{(w)}$.

At least four other striking consonant sound changes shaped the Beja system. Two of these affected the PC lateral fricative *ɬ. The other two separately merged word-initial and intervocalic PC *z with two other consonants:

10. PC *ɬ > r /V_V.
11. PC *ɬ > l /#_.
12. PC *z > y /V_.
13. PC *z > d /#_.

The ordering of these shifts remains unknown, but the shift of PC *z to *y in non-initial environments must be very ancient indeed, because it was already attested in the borrowing of the term for "goat" from Northern Cushitic by Nilo-Saharan peoples of the sixth or later seventh millennium BCE (Ehret 2001: 602).

It appears, finally, that the rare PC nasal *ŋ (and presumably also *ɲ) fell together in Beja with *n (Ehret 1987: 110).

The overall result of these various sound changes has been a radical reduction in Beja of the original proto-Cushitic inventory from 39 to just 20 consonants (Figure 3):

b	d	ḍ		g	g^w	
	t	ṭ		k	k^w	ʔ
f	s		ʃ			h
m	n					
w	r, l		y			

Figure 3. Consonants of Beja.

The change of PC *z to /y/ medially in Beja is of wider interest for Cushitic historical linguistics. This shift took place independently at least three times separately in different subgroups of the Eastern Cushitic branch (Sasse 1971; Ehret & Ali 1985; Ehret 1987), suggesting that interesting issues remain to be resolved about just how PC *z may have been articulated in early Cushitic. One possibility is that this consonant may actually have been not *z, but a palatal voiced stop *[ʄ], since examples of the shift *ʄ > *y (and also of *y > *ʄ) are common in language histories around the world. In addition, palatal voiced stops less commonly have been found to merge with dental or alveolar *d (this is a recurrent pattern in Nilo-Saharan, for instance; see Ehret 2001). Such a solution would have the doubly satisfactory outcome of resolving an imbalance in the PC consonant system, by filling the palatal gap in the voiced stop row and removing the single consonant of the voiced fricative row (see PC consonant chart above). Against this solution is the fact that PC consonant currently reconstructed as *z yields variously /z/, /s/, /d/, or /y/, but not /ʄ/, in the modern-day Cushitic languages.

3. Sound changes of the Agaw branch

Similar outcomes arose for certain of the PC consonants in the proto-Agaw language (Appleyard 2006), but the systemic contexts of the sound changes involved indicate different routes to these outcomes. One of the characteristic features of Proto-Agaw was the entire deletion of ejection in three of the PC ejectives and the

development of what were initially free alternances in the articulation of other three. The end result was that these consonants frequently show alternant outcomes in the modern-day Agaw languages. The most parsimonious accounting of the sound changes surrounding the developments in the ejectives would be the following series of shifts:

1. PC *k$^{(w)}$’ > pre-Agaw *q$^{(w)}$.
2. PC *p’ > pre-Agaw *ɓ.
3. PC *t’ > pre-Agaw *ts’.
4. *[+glottal/-lateral] > pre-Agaw *[-glottal/-lateral]

Rule #4 deleted the feature [+glottal] from the non-lateral glottalic consonants. As a result,

pre-Agaw *ɓ > proto-Agaw (PA) *b,
pre-Agaw *ts’ > PA *ts,
PC *c’ > PA *c.

The PC ejective lateral *tl’ maintained its ejection, but in tandem with the voiced PC lateral affricate *dl became a stop:

5. PC [+lateral/+obstruent] > pre-Agaw [-continuant/+obstruent].

That is, PC *dl > PA *d, PC *tl’ > pre-Agaw *t’. Subsequently, before the final breakup of proto-Agaw, an alternance in the articulation of this consonant developed:

6. pre-Agaw *t’ > *t ~ *t’.

This development created a number of instances of proto-Agaw roots in which the ejective articulation of *t’ was retained, but others (such as PA *taba “t’ef”) in which *t’ became simply *t. A further set of proto-Agaw alternances emerged affecting the pre-Agaw velars *q and *q^{w}:

7. Pre-Agaw *q$^{(w)}$ > PA *q$^{(w)}$ ~ *ɣ$^{(w)}$ ~ *x$^{(w)}$.

The effects of this alternance spread to original pre-Agaw cases of *ɣ, *ɣ$^{(w)}$, *x, and *x$^{(w)}$, creating further alternances in proto-Agaw between the reflexes of these two pairs of consonants (Ehret 1987; Appleyard 2006).

Several other notable consonant sound changes, the first two forming a sequence, contributed to the shaping of the eventual proto-Agaw consonant inventory. The latter three changes switch PC *m to PA *ŋ in several disparate environments (Ehret 1987).

8. PC *ɬ > *l.

9. pre-Agaw *l > *r /V_C.
10. PC *m > PA *ŋ /#_o.
11. PC *m > PA *ŋ /#_i(n).
12. PC *m > PA *ŋ /CV_-# (stem-final), where C ≠ [+velar] (> PA *ŋ^{w} /Cə_-#).

Proto-Agaw maintained PC *ŋ as PA *ŋ, but apparently merged the PC palatal nasal *ɲ with *n (one probable example is known so far: PA *na·w- "calf" < PC *ɲaʔ- "small, young").

The resulting consonant inventory of proto-Agaw (Figure 4), like that of Beja, considerably restructured the original proto-Cushitic array, and also shrank the number of consonants, but to a much lesser extent, from 38 or 39 to 28:

b	d		g	g^{w}			
p	t		k	k^{w}	q	q^{w}	ʔ
f	s		x	x^{w}			
	z	ʒ	ɤ	ɤ^{w}			
	c	č					
m	n						
w	r, l	y	ŋ	ŋ^{w}			

Figure 4. The consonants of proto-Agaw.

4. The evolution of proto-Eastern Cushitic

A very different sequence of sound changes led from proto-Cushitic down to the proto-Eastern Cushitic (PEC) language (Sasse 1979; Arvanites 1991; Ehret 1991). The most striking development distinguishing PEC was its creation of a new full series of consonants, voiced implosives, covering each of the major points of articulation. Four rules account for this series, with rules 3 and 4 left separate for now, although possibly combinable into a single rule. A linking feature of three of the four shifts is that each changed a continuant or partial continuant into an implosive, so possibly a single rule can eventually be formulated to capture the essence of those three:

1. PC *pʼ > PEC *ɓ.
2. PC *ɤ$^{(w)}$ > pre-PEC *ɠ$^{(w)}$.
3. PC *dl > PEC *ɗ.
4. PC *tlʼ > PEC *ƒ.

In addition, the PC voiceless labial fricative everywhere became PEC *l:

5. PEC *ɬ > PC *l.

A series of three other sound changes brought about another feature unique to PEC, the entire loss of the labiovelar consonants of PC. The two vowel sound changes of the sequence (rules 6 & 7), because they took place in the environment of an adjacent labiovelar consonant, necessarily preceded the disappearance of that environment through the merger of labiovelars with their equivalent regular velars. Rule #7 came after #5 because its following environment was clearly a PEC liquid, requiring that the PC obstruent *ɬ had previously merged with PEC *l:

6. PC *a(a), *e(e) > PEC *o(o) /C^w_C_2, C_2 = [+labial];
   PC *i(i) > PEC *u(u) ~ *i(i) /C^w_C_2, C_2 = [+labial].
7. PC *a(a), *e(e) > PEC *o(o) /C^w_C_2, C_2 = PEC liquid, i.e., *l or *r.
8. PC *$C^{(w)}$ > PEC *C.

Uniquely in Cushitic, the affricates of proto-Eastern Cushitic merged into a single consonant, plausibly explainable by the following sequence of sound changes:

9. PC *ts’ > pre-PEC *ts.
10. PC *dz > pre-PEC *ts.
11. pre-PEC *ts > PC *c.

The upshot of these sound changes was an overall proto-Eastern Cushitic consonant system (Figure 5) significantly smaller than the proto-Cushitic system, 30 consonants in place of 38 or 39. The entire loss, unique to PEC, of the velar/labiovelar distinction took place, along with the merging of the dental/alveolar affricates and, as noted previously, a considerable restructuring of the system overall:

b	d		g	ʕ
ɓ	ɗ	ʃ	ɠ	
	t	c	k	ʔ
	t’	c’	k’	
f	s	ç	x	ħ
	z			
m	n	ɲ	ŋ	
w	r, l	y		h

Figure 5. The consonants of proto-Eastern Cushitic.

5. Cushitic versus Cushitic

There is a last question we might consider here. Do the phonological seriations provide evidence for discerning any closer relationships among particular Cushitic branches, or do the various branches seem equally distant from each other? On the whole, the

answer seems to be that the course of sound change in the various branches followed separate trajectories, even if sometimes producing partially parallel outcomes. Two sound changes, however, may suggest possible closer relationships among particular pairs of branches.

The most salient of these is a change not yet mentioned, a simplification of the PC consonant cluster *nt to just *t in one particular, restricted environment in both Eastern Cushitic and Southern Cushitic:

A. PC *nt > *t /#ʔV_V#.

The known examples so far are just two root words, for "louse" and "you (sing.)" (*ʔint- and *ʔant-/*ʔint- respectively: Ehret 1987, roots 81, 82). But the environment is so precisely delimited that the sound change may nevertheless indeed be a valid shared innovation of Eastern and Southern Cushitic and thus a marker of an intermediate proto-Eastern-Southern Cushitic node in the Cushitic family tree, lying between proto-Cushitic and the proto-Eastern and proto-Southern Cushitic periods. More examples need to be sought, however, to fully validate this proposal.

An intriguing additional rule shared uniquely by two branches is a collapsing of the PC voiced velar fricatives with their voiceless equivalents. We have previously encountered this rule in both Beja and proto-Southern Cushitic:

B. PC *ɣ$^{(w)}$ > *x$^{(w)}$.

(In Beja, of course, *x and *x^{w} subsequently merged with, respectively *k and *k^{w}. In Southern Cushitic this further change did happen, but only in Dahalo.) The shift lacks the degree of uniqueness of the shared Eastern and Southern Cushitic shift, in that loss of voicing is one of the natural directions of change to affect voiced velar fricatives; it is more likely to be a parallel independent sound change than is shared rule A. Nevertheless, within Cushitic it is a unique sharing between two geographically distant branches, Southern Cushitic and Beja, not attributable to areal influences and therefore worthy of further consideration as a possible marker of an otherwise improbable Beja-Southern Cushitic node in the Cushitic classification.

References

Appleyard, David L. 2006. *A Comparative Dictionary of the Agaw Languages*. Köln: Rüdiger Köppe.

Arvanites, Linda 1991. The Glottalic Phonemes of Proto-Eastern Cushitic, Ph.D. dissertation, University of California at Los Angeles.

Ehret, C. 1980. *The Historical Reconstruction of Southern Cushitic Phonology and Vocabulary*. Berlin: Reimer.

Ehret, C. 1987. "Proto-Cushitic Reconstruction." *Sprache und Geschichte in Afrika* 8. 7–180.

Ehret, C. 1991. "Revising the Consonant Inventory of Proto-Eastern Cushitic." *Studies in African Linguistics* 22: 3. 211–275.

Ehret, C. 1995. *Reconstructing Proto-Afroasiatic (Proto-Afrasian): Vowels, Tone, Consonants, and Vocabulary*. Berkeley, Los Angeles: University of California Press.

Ehret, C. 2001. *A Comparative Historical Reconstruction of Proto-Nilo-Saharan*. Köln: Rüdiger Köppe.

Ehret, C., & Ali, M.N. 1985. "Soomaali Classification." *Proceedings of the Second International Congress of Somali Studies (Hamburg, August, 1983)*, Ed. by T. Labahn. Hamburg: Buske.

Sasse, Hans-Jürgen. "The Consonant Phonemes of Proto-East-Cushitic (PEC): A First Approximation." *Afroasiatic Linguistics* 7. 1–67.

Erosion in Chadic

Herrmann Jungraithmayr
Johann Wolfgang Goethe-Universität

The Chadic language family in the Central Sudan is comparable to a landscape the surface of which consists of plains, valleys and mountain ridges. Some areas were more resistant than others against the forces of erosion. Different degrees of hardness and resistibility caused languages to preserve or reduce the original substance with regard to lexicon or grammar of an individual language or the entire language group. Chadic with its ca. 130–150 languages is probably the one branch of Hamitosemitic (Afroaasiatic) which has been subject to the strongest transformational processes in which erosion, i.e., attrition of language material in space and time, played a dominant role. The paper presents numerous examples illustrating different stages of progressive erosion which may characterize the nature of Chadic linguistic history.

1. Erosion in geomorphology and linguistics

"Erosion" is a technical term taken from the realm of physical geography and geology. There it signifies the gradual reduction of parts of the substance of a given material, either by the water (of a river or of rain) or by the wind. High terraces, tablelands, or "Inselberge" standing out against the surrounding plains are typical resistant parts which, with their more solid substance, were able to withstand either the fluvial, pluvial or aerial forces; cf. the profile of the Tangale country in Brunck 1991: 103.

By analogy, "linguistic erosion" denotes the gradual attrition of language material in space and time. As far as language is a product of physical activity, it is subject to constant change, normally – in its sound and morphological systems – towards reduction, apocopation and syncopation of a given lexical item or morpheme, or towards total replacement. The inducing forces behind such changes are either mere internal evolutionary factors – being inherent to any living entity including language – or external impulses received from contacts with other language communities.

If we conceive of the Chadic language family as a landscape consisting of plains, valleys and mountain ridges, we observe and have to study mainly two types of linguistic erosion, i.e., (a) the erosion that led to the given surface differences within the entire area and between its individual sectors, and (b) the internal erosion inside each

individual member of the whole, that is each language – or even words or morphemes within a language – of the given Chadic family.

As there are different degrees of hardness within the rocky ground of a plateau, languages and – within a given language – the different parts of the vocabulary and grammar may display extremely variable degrees of resistibility against the different forces of change. Thus, to give an example, in Hausa a strong tendency towards forming denominative verbal expressions (by the auxiliary verb *yi* "to do") has led – in the course of the language's history – to the replacement of an older verbal vocabulary of a rather basic quality still existing in other Chadic languages where that process has not taken place, e.g.,

	Hausa	**Proto-Chadic**	**Chadic languages**
"to laugh"	*yí daaríyáa*	*GMS	Mubi: *gàmàsé*
"to dream"	*yí máfárkíi*	*SWN	Ngizim: *sáunú*
"to sleep"	*yí bárcíi*	*(W)SN	Warji: *sən-*

Other "vulnerable" features of the Hausa lexicon can be observed in the following glosses:

"foot, leg"	*ƙáfaa*	*SKR	Ron-Daffo: *sakur*
"moon"	*wátaa*	*T-R	Angas: *taar*
"sun"	*ráanáa*	*PT	Karekare: *pàtí*
"to give birth"	*haifáa*	*WY	Birgid: *wáay-*
"to kill"	*káshee*	*DWK	Tangale: *tuge*

In certain instances, Hausa has actually preserved the lexical root, but at the same time shifted its semantics to a closely related meaning, e.g.,

"foot(print)"	*sáu*	*SKR
"smiling"	*murmushíi*	*GMS
"beat"	*dookáa*	*DWK

On the other hand, Hausa shares with the majority of Chadic languages such resistant lexical items as

	Hausa	**Proto-Chadic (PC)**	**Chadic languages**
"four"	*húdú*	*-Pɗ	Pa'a: *fáɗù*
"to drink"	*sháa*	*S_2^{w}/$_y$ʔh	Ron-Daffo: *shoh*

(*continued*)

Continued

	Hausa	Proto-Chadic (PC)	Chadic languages
"to eat"	*cí*	*TWY	Dera: *twi*
"wind, cold"	*sányíi*	*SMɗ	Diri: *súmbùɗù*
"ash(es)"	*hábɗíi*	*BT	Kirfi: *bùtó*
"to blow"	*húuraa*	*PLT	Bole: *fiit-*
"bone"	*ƙashíi*	*ƙS_3	Jimbi: *ƙəƙásì*
"body"	*jikíi*	*ZK	Boghom: *shòk*
"eye"	*ídoo*	*YDN	Karekare: *ìdó*
"fat"	*kítsee*	*KD^{Y}R	Pa'a: *hìɗur*
"heart"	*zúuciyáa*	*DKT	Hausa (PL.): *zúkàatáa*
"fish"	*kíifíi*	*KLP	Siri: *kə̂rfí*
"to give"	*báa*	*BR	Ngizim: *bàrú*
"guinea-fowl"	*zaabóo*	*ZBL	Karya: *zábùr*

From the comparison of some of the Hausa reflexes with the respective reconstructed roots, as well as reflexes of other more conservative Chadic languages, it becomes clear to what extent the Hausa lexeme has been subject to the forces of erosion. Compare, e.g.,

"to eat" *cíì* (v.n.)	Mubi: *túwáà* (v.n.) Ron-Daffo: *cwaah* (v.n.)
"foot(print)" *sáu*	Ron-Daffo: *sakur*

Undoubtedly, Hausa like any other language, has undergone deep changes in the course of its history, phonetically by (1) weakening certain syllable-final consonants, which processes have become known by the name of "Klingenheben's law" (1927/28), or (2) clipping word-final consonants, thus producing in Hausa the modern canonical vowel-final structure (Jungraithmayr, Ms.).
Examples:

		Standard Hausa	West-Hausa
(1)	"to sit down"	*záunaa*	*zámnaa*
	"to descend"	*saukáa*	*sabkáa*
		Standard Hausa	**Proto-Chadic**
(2)	"crocodile"	*kádoo/kádaa*	*KDM
	"eye"	*ídoo* (PL. *ídaanúu*)	*YDN

2. Linguistic history and stages of erosion

Each of the approximately 130–150 Chadic languages is a living historic document inasmuch it represents a certain stage of development within the general common history of the entire family. For the purpose of a working hypothesis, we may group these almost infinitely divergent reflexes of one and the same root (PC) into three "typological stages", i.e., "Old Chadic", "Middle Chadic" and "New Chadic". These stages are characterized by different degrees of phonological attrition and erosion (by apo- or syncopation), comparable within Germanic to Gothic: *fidwor* / Engl.: *four* / German: *vier*, or Latin: *mater* / Italian: *madre* / French: *mère*.

Typological stages

Gloss	PC	"Old Chadic"		"Middle Chadic"		"New Chadic"	
"tail"	*K^WTR	Warji:	*kwátárę́*	Cibak:	*$k^wuɗæ$*	Sukur:	*t^hur*
		Tangale:	*kọ̀dọ̀r*	Hausa:	*wútsíy-*	Zaar:	*kiir*
		Musgu:	*gider*	Sibine:	*wúdíny*	Kir:	*kə̀r*
		Kirfi:	*kìtírí*			Boghom:	*kay*
		Gidar:	*kútra*				
"neck"	*$G^WɗʸR$*	Diri:	*ŋgwáɗù*	Boghom:	*gway*	Ron-Sha:	*woḥ*
		Ndam:	*kwad*	Pa'a:	*rèy*	Hausa:	*wúy-*
		Guduf:	*kwindà*	Musgu:	*ulla*	Maha:	*wah*
		Mburku:	*(ʏ)wíírò*	Mokilko:	*ʔóṛé*	Ngweshe:	*íʔà*
		Mubi:	*wíírì*			Gisiga:	*ʔay*
		Karya:	*(ʏ)wir*			Mandara:	*íye*
		Tala:	*yaar*			Yedina:	*wui*
		Kilba:	*wulya*				
"die"	*MWT	Mubi:	*muwaat*	Tangale:	*mụdẹ*	Tumak:	*ma*
				Hausa:	*mutu*		
"bone"	*$ƙS^3$	Hausa:	*ƙas-*	Tangale:	*wush*	Kofyar:	*és*
		Ndam:	*gusé*	Migama:	*àssú*	Yedina:	*háy*
"root"	*LRW	Sibine:	*sárbə̄*	Hausa:	*saywaa*	Tangale:	*yara*
		Miya:	*tlerwa*	Masa:	*soʾol-*	Tumak:	*həraw*
		Ndam:	*sírwé*	Zime:	*tloor*	Muktile:	*tlíh*
		Siri:	*tluura*			Kera:	*kə́-sar*
"wind, cold"	*SMɗ	Zime:	*simbeɗe*	Hausa:	*sányíi*	Bokkos:	*se*
		Diri:	*sumbuɗu*	Gidar:	*semia*	Tumak:	*had*

(*continued*)

Typological stages (Continued)

Gloss	PC	"Old Chadic"		"Middle Chadic"		"New Chadic"	
		Kotoko:	*sə̀máaɗə̀*	Boghom:	*swam*	Mokilko:	*maye*
		Yedina:	*hambura*	Masa:	*sime*	Daba:	*miɗ*
		Pa'a:	*sənɗi*				
		Tangale:	*yibat*				
		Dera:	*yiwet*				
"dream"	*SWN	Kabalai:	*suwən*	Dangla:	*soone*	Tera:	*zhine*
		Ngizim:	*suwan*			Daffo:	*sun-an*
		Dera:	*jowan*			Tangale:	*wuna*
		Sura:	*sugun*				
		Kotoko:	*suwane*				
"leg, foot"	*SKR	Daffo:	*sakur*	Bole:	*sheke*	Hausa:	*sau*
		Ngizim:	*zəgər*	Mafa:	*sak*	Sura:	*shii*
		Bade:	*zgil*			Tangale:	*yoo*

3. The case of "neck"

One of the most divergent and deeply eroded roots in Chadic is that for the gloss "neck". We are not yet quite sure about its definitive reconstruction, yet the arguments for the explanatory analysis outlined below are strong. So far the following reconstructions have been proposed: Newman & Ma (1966): **wVrV*; Newman (1977): **wəra*; Jungraithmayr & Shimizu (1981): **g^{w}r*.

Compare the following data:

Jegu	*were*	Karya	*(ɣ)wîr*
Kulere	*wur*	Kotoko	*ʏwàyè*
Mburku	*(ɣ)wiiro*	Yiwom	*ɤʏa*
Miya	*wír*	Warji	*yiráy*
Nzangi	*wura*	Siri	*yere, yírí*
Migama	*úrè*	Bade	*ùlá*
Guruntum	*iri*	Ron/Daffo	*là*

On the basis of reflexes like

Mubi:	*wíirì*	Tala:	*yaar*

the root may rather have had the structure **g^{w}-r*.

However, if we include more evidence, we should probably arrive at the reconstruction *g^wd^yr. Compare the following data:

Tera:	*ɠura*	‹ **ɠuura* (‹ **guɗʸura?*)	
Ga'anda:	*ɓira-ta*	‹ **ɓiira* (‹ **ɓiɗyira?*)	
Karekare:	*'wulau*		
Bole:	*diɗo*	Bole-Tangale	**(ŋ)giɗo*
Muktele:	*kùɗá*		(Schuh 1984: 176)
Guduf:	*kwíndà*		
Ndam:	*kwad*		
Dangla:	*gáaɗyá*		
Bure:	*gòojìrè*	gwV- (= gwa- ?) ›	gVV- (= goo-)
Kwami:	*ngozor*		
Bura:	*wulya*	Bade:	*ùlá*
Karekare:	*'wulau*	Daba:	*wìlà*
Lamang:	*úlé-ké*	Gudu:	*kwíndà*
Gidar:	*ulokó*	Sukur:	*wo'i*
		Ngweshe:	*íà*
Bura:	*wulya*	‹ **wuɗʸra*	‹ **gwuɗʸar*
Hausa:	*wúyàa*	‹ **wuyay* ‹ **wuyar*	‹ **wuɗyar*
Mofu:	*ɗáy* (‹ **(g)waɗ- ?*)	cf. Muktele	*kùɗá*
Mafa:	*dáâày* (‹ **gaɗyar*)		
Giziga:	*'ay*		
Zime (Batna):	*dray/dərey* (‹ **gèɗʸer ?*)		
Masa:	*del* (‹ **geɗy- ?*)		
Kwang:	kàrí		
Kera:	kur		
Lele:	korga (=kor-ga ?).		

Conclusion

Linguistic erosion is a feature of linguistic history. In general, the older a language is, the more it has been subject to processes such as erosion (apocopation, syncopation, word-thinning, consonant-weaking, etc.). It is among the more important tasks of comparative linguistics to develop methods by which the phenomena described above may be exploited for the study of the ethnic history of the Chadic-speaking world.

References

Brunck, K. 1991. "Natural resources and landscape development in the Southern Gongola Basin, NE Nigeria." *West African Savannah. Culture, Language and Environment in an historical Perspective*, Prelim. Report 1989–91, ed. by H. Jungraithmayr & G. Nagel, 97–104. Frankfurt am Main.

Jungraithmayr, H. 1977. "Sprachhistorische Schichtstufen im Tschadraum." *Paideuma* 23. 95–100.
Jungraithmayr, H. 1978. "A tentative four stage model for the development of the Chadic languages." *Quaderni di Semitistica* 5. 381–388.
Jungraithmayr, H. Ms. "Final consonants in Old Hausa." Kano.
Jungraithmayr, H. & Shimizu, K. 1981. *Chadic Lexical Roots*, vol. II. Berlin.
Jungraithmayr, H. & Ibriszimow, D. 1994. *Chadic Lexical Roots*, 2 vols. Berlin.
Klingenheben, A. 1927/28. "Die Silbenauslautgesetze des Hausa." *Zeitschrift für Eingeborenen-Sprachen* XVIII: 4. 279–297.
Leger, R. 1994. *Eine Grammatik der Kwami-Sprache (Nordnigeria)*. Cologne.
Leger, R. Unpublished notes on Bure. Frankfurt am Main.
Newman, P. 1977. "Chadic classification and reconstructions." *Afroasiatic Linguistics* 5: 1. 65–84.
Newman, P. & Ma, R. 1966. "Comparative Chadic: phonology and lexicon." *Journal of African Languages* 5. 218–51.
Schuh, R.G. 1984. "West Chadic Vowel Correspondences." *Current Progress in Afro-Asiatic Linguistics*. Papers of the Third International Hamito-Semitic Congress, Ed. by J. Bynon, 167–223. Amsterdam.

On Kunama *ukunkula* "elbow" and its proposed cognates in Nilo-Saharan languages

Philippe Bürgisser
Lausanne, Switzerland

This article is dedicated to Professor Harold C. Fleming, with my congratulations and my deep gratitude

There have been two attempts at reconstructing proto-Nilo-Saharan (pNS), yet they diverge widely. To illustrate the lack of consensus even at the level of cognate recognition among NS languages, I selected the Kunama word *ukunkula* "elbow" and found it to be included by several scholars in different etymologies which poorly overlap. One major disagreement bears on whether *ukunkula* consists of a unique duplicated stem or is a compound word. *Ukunkula* is often considered cognate with Songay *haŋkoro* (*honkoro*) and Berta *k'oŋk'oloŋ* "elbow". I propose here that *ukunkula* is unrelated and that the real Kunama cognate of the Songay and Berta nouns is *ikoŋkoro-* "to bend". I could not find convincing cognates of *ukunkula* or *ikoŋkoro-* in Shabo and the Kadu family, likely affiliated to NS. Nevertheless, I present some other unpublished potential cognates between Shabo, Kadu and NS that may strengthen the case for a genetic relationship among them.

While looking at some proposed Nilo-Saharan (NS) etymologies, I fell quite by chance on the Kunama word *ukunkula* meaning "elbow" and, just out of curiosity, undertook to look more in depth how different scholars dealt with this word. At first, such an investigation may seem quite trivial. However, it soon became apparent that it brought more questions than answers. Indeed, roots such as *kVl* are numerous in many NS languages and there is little agreement between researchers as to which sets of potential cognates they should be assigned. In the first part of this paper, I shall compare and discuss some of the published etymologies involving *ukunkula* and offer an alternative hypothesis. In a second part, I will present potential new cognates between Shabo, Kadu and NS languages that came to light in the course of this study. Shabo is an isolate spoken in southwestern Ethiopia, and Kadu is a small family of

nine closely related languages spoken in southern Sudan. The taxonomic position of both Shabo and Kadu is still controversial, but several scholars presented evidence for a genetic relationship to NS (see ref. 1–4 for discussion).

In preparing this paper, I have used the two presently published attempts at reconstructing proto-Nilo-Saharan (pNS), and it is appropriate to briefly introduce them here. One was published by C. Ehret in 2001 (ref. 5), the other by M.L. Bender in 1996 (ref. 6). I have already discussed the qualities and weaknesses of Ehret's reconstruction in a previous article (ref. 3). In short, Ehret's work has three major characteristics: (1) it presents an extremely, if not excessively, sophisticated phonetic approach; (2) the author believes in the exceptionless nature of sound "laws" (3) the work is fraught with far-fetched semantics that invalidates up to one third of the proposed etymologies. In contrast, Bender proposes a pNS phonetic inventory that is very simple, probably even simplistic, and sound correspondences that are poorly worked-out. In addition, Bender, like Ehret, allows for semantic latitude that goes too far to my taste. Let me admit here my preference for Ehret's work, because it is very carefully done. I would nevertheless emphasize that while I believe that regular sound correspondences must be uncovered as far as possible, I do not share Ehret's faith in the existence of exceptionless sound laws and consider that exceptions must be permitted with proper judgment. In addition, those of his proposed cognates with poor semantic fit must be discarded.

As far as Shabo and Kadu are concerned, systems of sound correspondences are far from being established. The cognates that will be proposed have been selected on the principle of phonetic and semantic plausibility used by Greenberg, Fleming, and others.

Part 1. Kunama *ukunkula* and its proposed cognates

- **Ehret and Kunama *ukunkula***. Ehret connects Kunama *ukunkula* with just one pNS root meaning "to bend" (etymology ⟨1⟩).

⟨1⟩ E1059 ***k^{h}ul*** "to bend (intr.)":
Koman: Uduk *k^{h}ululak^{h}ulul* "bent, crooked" [ADJ. by reduplicated stem]
Kunama: *ukunkula* "elbow" [n. derivative pref. **o* > *u-* plus partially reduplicated stem]
Saharan: Kanuri *këla* "to roll into a ball" [NS suff. **-a* dispunctive]; Kanuri *këlë* "to wrap around" [NS suff. **'y* essive-active (?)]
Songay: Zerma *kulla-kulla* "to have knock-need walk" [NS suff. **-a* dispunctive]. Also present in Songay Kaado *kulla-kulla* "avoir une démarche cagneuse (Fr.)" (ref. 7).
W. Nilot.: Ocolo (Shilluk) *kul* "to bow the head".

Rub (= Kuliak): Ik *ikukul-* "to go the wrong way and come back" [**i-* v. class pref. plus partial reduplication. "Loan", expected **kul.-* with *l.* instead of *l*].

The connection between "to bend" and either "elbow" or "knee" (namely "bent part of a limb") is too universal to necessitate any comment. The proposed Kanuri reflex appears somewhat questionable from a semantic point of view. The Ik word may reflect an irregular sound correspondence rather than being a loan. Most importantly with respect to our subject, the phonetic analysis of Kunama *ukunkula* seems to me unclear. Ehret interprets this word as the partial duplication of a unique root. However, for a partial reduplication, one would expect **ukukula*, and if the reduplication was complete, one would expect **ukulkula*. Thus the internal *-n-* is not explained satisfactorily.

- **Bender and Kunama *ukunkula***. In sharp contrast to Ehret, Bender analyzes the Kunama word as a compound of two distinct roots with various meanings related to the limbs anatomy (etymologies ⟨2⟩ and ⟨3⟩):

⟨2⟩ B320 ****kump, *koN, *kon***- for the 1st part of Kunama *ukun-kula*
Koman: Kwama *-kump*" "claw, finger". Gumuz Sai *kumb-ërë* "knee"
Centr. Sud.: *kon* "elbow", *(k)on^{y}* "claw, finger", **kum- ~ gomo* "knee" (no further details given in ref. 6!) →
Kunama: *ukun*-(kula) "elbow"; Kunama, Ilit *kona* "hand, arm"
Songay: Gao *hon*-(koro) "elbow"
Maban: Masalit *-koŋ* "elbow". The Masalit full form is *muturkoŋ-yi ~ muturkon-ji ~ muturk(w)un^{y}-i* (ref. 8).
Berta: *k´on-, kwòn-* "elbow".
East Sud.**:** *kum, kòm-* "elbow", *kuun* "elbow, knee" (no further details given in ref. 6). See e.g., Taman Merarit *kum* "knee" (ref. 9), Surmic Mursi *kòmi* "knee" (ref. 10), and also pE.Nil. **-kvŋ,* e.g., Bari *kvŋ,* Turkana *a-kvŋ* "knee" (ref. 11).
→ For Centr. Sud. *(k)on^{y}* "claw, finger", Bender quotes, in ref. 12, Miza (Moru) *k$^{(w)}$ònyı,* Lugbara *òn^{y}ago-*, Lokai (Madi) *an^{y}u,* Mangbetu *-kònzò*, Aja *koni/kan^{y}an*, Fongoro *kunza*, Fer *ka(n)za* and Modo *kon^{y}o*. This clearly points at a root such as **on^{y}* preceded in some languages by the "movable" *k* prefix identified by Greenberg (ref. 13; < pNS **k^{h}* according to Ehret, ref. 5).

I could find such a set neither in Ehret's (ref. 5) nor in Greenberg's (ref. 13) work.

⟨3⟩ B11 ****kol-, kor***$_2$ for the 2nd part of *ukun-kula.*
Koman: Komo *kol-o* "hand, arm".
Kunama: (ukun)-*kul-a* "elbow".
Saharan: Kanuri, Tubu *-kol-* "foot". Tubu *kur-u* "elbow"
For: *ka(a)r-u* "claw".
Songay: Gao (hon)-*koro* "elbow".

Maban: Maba, Mimi *kar-* "finger" (in fact *kar-tuŋgulak* "hand-ancestor", ref. 8), Masalit *koro* "hand, arm", Aiki (Runga) *kòrò-* "elbow"
Berta: Gebeto *kol-ò ~ k'(w)òl-o* "foot"; other varieties *k'òlòò* "leg" (ref. 14)
East Sud.: *-kol-* "claw"; *-kol-, kòl-* "elbow" (no further details given!)
Rub: Ik *kòr-òk* "finger".

Bender (ref. 6) also puts in ⟨3⟩ Kadu Tolibi *-gòòr-ò* and Krongo *-koor-o* "elbow". When looking at the non-truncated forms, Tolibi (ref. 15) *ma-t'òŋgòòrò* (PL. *ki-t'òŋgòòrò*), Tolibi (ref. 16) *mutòŋgò@rò* (PL. *na-mutòŋgò@rò*); Krongo (ref. 15) *kòtigòòrò* (PL. *ni-kòtigòòrò*), Krongo (ref. 17) *kattıgooro* (pl. *nı-kattıgooro*), I wonder whether Bender's segmentation is correct. For the Krongo form, M. Reh does not propose any segmentation (ref. 17). Thus the Kadu forms may not belong here.

My first comment to the ⟨2⟩ plus ⟨3⟩ proposal is the great semantic latitude allowed by Bender, with which I do not feel quite comfortable. This may be a personal bias due to my education as a physician, making it difficult for me to accept meanings as different as "foot", "claw", "finger", "elbow" in the same etymology! Note however that Bender himself concedes that "The semantic spread [in etymology ⟨2⟩] … is a bit wide" (ref. 18). Another problem is the poorly defined state of the pNS roots postulated. This is particularly evident for etymology ⟨2⟩, which I suspect to be made up of several genetically distinct roots. A third observation is the absence of overlap between etymologies ⟨1⟩ and ⟨3⟩, in spite of phonetic similarity. This is partly the consequence of the strict observance by Ehret of the phonetic "laws" he has postulated. This scholar has reconstructed no less than three pNS unvoiced velar stops, i.e., **k,* **k^h* and **k'* with different outcomes, notably in Songay (*g, k* and *h,* respectively), in word-initial position (ref. 5). In sharp contrast, Bender reconstructs only **k* and **k^h,* with **k* > *k* and **k^h* > *k ~ h* in Songay (ref. 6)! Finally, the most perplexing point in Bender's proposal is that he regards *u-kun-kul-a* as the juxtaposition of two pNS words with exactly the same spectrum of meanings (semantic redundancy). For all those reasons, Bender's double etymology for *ukunkula* does not look very satisfactory.

Another point of disagreement between Bender and Ehret bears on Kunama *kona* and *ukun-*. Bender puts Kunama *kona* "hand, arm" together with *ukun-*(kula) in comparison ⟨2⟩. For Ehret, these two words are not cognates because their proto-form begins with two different velar stops, i.e., **k'-* in the case of *kona* (etymology ⟨4⟩, shown below) and **k^h* in the case of *ukun-*(etymology ⟨1⟩).

⟨4⟩ E1114 ****k'wɛy*** "fingers" (suppletive PL.)
Koman: Opo *k'wi* "hand"
Centr. Sud.: pCS **kwe* or **k'we* "finger"
Kunama: *ko-n-a* [NS **n* n. suff.] "hand" →
Temein: *kwi* "hands"
Rub: pRub **a-k'w*, Ik *ak'w* "palm, sole" [NS **a-* n. derivative pref.]. Ik *kwɛ-t* "hand, foreleg" [NS **t^h* n. suff.] is considered a "loan" (*k* instead of the expected *k'*).

→ The same NS **n* n. suff. is recurrent in Kunama according to Ehret, e.g., E1214 "grass", Kunama *se-n-a vs* pEast Centr. Sud. **sʸɛ* or **sɛ* or **θɛ* (Bender **ese*, ref. 12) and Saharan Kanuri *sheshe* (pNS **sʸɛ* or **sɛ:*).

I find etymology ⟨4⟩ quite acceptable. The Ik reflex may be a case of irregular correspondence. We will return to this etymology later.

- **Greenberg and Kunama *ukunkula*.** Greenberg proposes a quite different cognation, comparing *ukunkula* with several words meaning "armpit" (ref. 13):

⟨5⟩ G/NS7 "armpit"
Kunama: *ukunkula* "elbow".
Saharan: Daza *kilikili*, Zaghawa *kallikalli* "armpit".
Berta: Fazoglo *yoŋyòleŋ* "elbow".
Surmic: Didinga *kalkic* "armpit".
S. Nilot.: Nandi *kulkul* "armpit". pSNil. **kulʰkulʰ* according to Rottland (ref. 19).

Greenberg also gives the additional meaning "armpit" for *ukunkula,* although Kunama has a different word for it, which is *akula* (ref. 20). Thus, for him, "elbow" and "armpit" can be connected semantically. On the contrary, both Bender and Ehret propose different etymologies for "armpit" (B132 **kol* "armpit, breast or chest"; E1030 **kʰil, *kʰilkʰil* "to tickle, armpit") and for "elbow" (cf. ⟨1⟩ to ⟨3⟩). This is what I expect because armpit and elbow are in my view quite distinct body parts.

- **Kunama *ukunkula* and global etymologies.** In 1994, Bengtson and Ruhlen published an article in which they propose 27 global etymologies (ref. 21). These scholars connect *ukunkula* with two of them (etymologies ⟨6⟩ and ⟨7⟩ below) by segmenting this word in the same way as Bender does.

⟨6⟩ BR6 ***kano*** "arm" (as well as "hand", "shoulder")
Kunama: Kunama *kona*, Ilit *kon* "hand", also "arm" (ref. 20); *u-kun-*(kula) "elbow"
Berta: *k'oŋ-*(k'oloŋ) "elbow"
E. Nilot.: Teso *(a-)kani*, Maasai *(eŋ-)kaina* "hand" →
→ pE.Nil. **kain-*(note Ehret's discordant interpretation: **kʰ* n. prefix + root **a'y* + **n* n. suff., E1533!)
As far as NS is involved, proposal ⟨6⟩ and Bender's ⟨2⟩ overlap, so that the reader can refer to the latter for discussion. Global root ⟨6⟩ is reminiscent of two similar but apparently unrelated roots identified by Greenberg (ref. 22) in his recent work on Eurasiatic, i.e., **kon* "armpit" (also "hollow, cavity") (Uralic, Altaic) and **ken* "elbow" (i.e., "elbow, forearm/knee, shin") (Indo-European, Uralic, Altaic, Japanese, Eskimo-Aleut). Thus, in this instance, Greenberg makes a clear semantic distinction between "elbow" and "armpit".

Even though I believe that etymologies ⟨2⟩ and ⟨6⟩ are genetically heterogeneous, and in spite of Ehret's satisfactory analysis of Kunama *kona* (etymology ⟨4⟩), I cannot exclude that either Kunama *kona* or *ukun-* represent an old NS remnant of a global etymology that needs however more stringent delineation.

⟨7⟩ BR8 ***k'olo*** "hole"
Koman: Gumuz (Buldiit) *kul(ma)* "buttocks"
Kunama: *kura* "anus" (in fact "buttock, rump, bottom" according to Bender, ref. 20), (u-kun)-*kula* "armpit, elbow"
Saharan: Kanuri *kuli,* Teda *kulo* "anus"
Songay: Gao *nkoro* "buttocks"
Berta: (k'oŋ)-*k'oloŋ* "elbow"
W. Jebel: Gaam *kura-n* "hollow (in ground)"
Temein: *kukuruk(it)* "buttocks"
S. Nilot.: Nandi *kulkul* "armpit"
Rub: Soo *ukòlkòl* "armpit" (but *ukokol* in B132, ref. 6)

Referring to NS, etymology ⟨7⟩ seems to me as heterogeneous as etymology ⟨6⟩. I am reluctant to accept it as such, for phonetic as well as semantic reasons. Etymology ⟨7⟩ contains some words that I would relate in a more satisfying way to one of two reconstructions by Ehret, namely: E1075 **k^{h}wil* or **kwil* "to crack, break open, cavity, hole, mouth"; E1461 **k^{h}wil* or **k^h'wil* (< ***k^h-wil* or ***k^h-'wil*, with the "movable" **k^h-* prefix) "buttocks, anus, hip, tail".

With respect to other world's language families, some of the forms adduced by Bengtson and Ruhlen, such as Dravidian Kannada *koŋkur, gaŋkëlu* or Tulu *kaŋkula* "armpit" (approximate phonetic notation), certainly look at first quite similar to the Kunama word under discussion (ref. 23). This is also the case for Uralic Estonian *kangla ~ kangel* "armpit, shoulder" and Mari *koŋyëla ~ koŋla* "armpit" (ref. 24), as well as for Mongolian Written Mongolian *qonyur ~ qonyil* and Khalkha *xonxor ~ xongil* "hollow, cavity" (ref. 25). But reconstructed forms (ref. 23–25), if correct, do not support a genetic connection with *ukunkula* (pSouth Dravidian **kavuŋ-kat-*, pTelugu **kavŋ-il-*, pKolami-Gadba **kavŋ-*; pUralic **kon*(*V*) + **kaŋla* (*sic*); pAltaic **kobani*; all supposedly from **K'awingV*, Dolgopolsky, ref. 23–25, an etymology considered "possible" by Bomhard in his unpublished review of Dolgopolsky's Nostratic Dictionary). From a semantic point of view, it should be emphasized that again all of the items considered mean "armpit, shoulder or cavity" rather than "elbow" or "bend".

To sum up, *ukunkula* "elbow" is not well accounted for by the juxtaposition of etymologies ⟨6⟩ and ⟨7⟩, which may be better suited to designate the armpit (= "hole of the arm").

- **"Elbow" as a possible compound word in Kunama, Songay and Berta.** The reader has probably been struck by the apparent similarity of three words viewed as compounds by several scholars, notably Bender (but not Ehret):

Kunama	*ukunkula*	=	*u-kun-kul-a*
Songay of Gao	*honkoro*	=	*hon-kor-o*
Berta	*k'oŋk'oloŋ*	=	*k'on-k'olo-ŋ*

But are these forms really compounds? And are they genetically related?

Before attempting to answer, we must first introduce the Berta word that was only mentioned in part in etymologies ⟨2⟩ (two forms given), ⟨6⟩ and ⟨7⟩, and bring our data on Songay up to date. As for Berta, the basic form is *k'oŋk'oloŋ* (Undu, Mayu and Fadasi varieties) (ref. 14, 15, 26). A second form, *k^{w}òŋk^{w}òlòŋ*, was collected besides *k'oŋk'oloŋ* from one speaker originating from the Jebel Ura area (ref. 14). Only the major Berta representative, *k'oŋk'oloŋ*, will be considered here. In this word, the final *ŋ* probably represents a NS noun suffix (Ehret, ref. 5; see e.g., Berta *k'òrò-ŋ* "back of neck", *amv-ŋ* "nose").

In the Songay languages, the full expression to mean "elbow" generally begins with *ka(m)ba*, which stands for "hand, arm". *kaba ~ kamba ~ kambe* is the basis of numerous compound words, e.g., Gao *kamba hamo* "upper arm", *kamba tola* "hollow in the arm" > "anticubital fossa" (*vs ce: tola* "hollow in the leg"> "popliteal fossa"), *kaba banda* "dorsum of the hand" (*banda* = "external face, back"), *kaba jinde* "narrow part of the arm" > "wrist", and many others (ref. 7, 27–32). The designation of the "elbow" is astonishingly diverse among the closely related Songay languages and involves several unrelated roots:

Southern group			
Tondi Songway	Douentza [33]	*kambɛ tòtoŋgo*	
Humburi Senni	Hombori [33]	*kamb-u to:t-iy-a*	also used for "knee"
Western Songay	Timbuktu [28]	*kamba tokone*	
	Niafounké [28]	*kamba to:na*	
	Djenné [29]	*sokone*	
	Goundam [28]	*kamba to:ne*	
		kamba koro	
Eastern Songay	Gao [27, 30]	*ka(m)ba honkoro*	
		kaba haŋkoro	
		kaba haŋkore	
	Bamba [30]	*kamba haŋkoro*	
	Ansongo [30]	*kamba haŋkor*	
	Téra (Kaado) [7]	*kab koro*	

(*continued*)

(Continued)

Southern group			
	Zarmaciine [31]	*kambe gollo*	cf. Berta *gòli* "crooked, curved", *agoogole* "round".[19] Maban Masalit *ŋgullɛ* & Kibet *gulana* "crooked", Aiki *gul-i* "articulation".[17] Kunama *gugula* "round".[25]
	Zerma [33]	*kambe golo*	
	Zerma [34]	*gollo*	
Northern group			
	Tadaksahak [35]	*tëymurt*	borrowing from Tamasheq

The Djenné form is cognate with Songay of Gao *sokono* "to bend (legs)", Dendi *sokom* (v.) and *sokoniam* (n.) "crouch" (ref. 27, 30, 32). For its part Zerma *gol(l)o* is related to Songay of Timbuktu *golle* "hook, metal hook" and *golli* "to bend (intr., tr.), twist", Songay of Goundam *golo* "hook, metal hook" and Songay of Gao *golli* ~ *gulli* "to be crooked, bent, curved" (ref. 28, 30). Those two roots are underlined here because they share the meaning "to bend".

In the Songay of Gao the full expression for "elbow" is *ka(m)ba haŋkoro*. I interpret it as the "bent part of the arm", in which *haŋkoro* by itself appears to mean "bend" (cf. Fr. "coude" *sensu lato*), as do *sokone* and *gol(l)o* in Djenné and Zerma, respectively. This interpretation is also strongly supported by the notable parallelism in Songay Ansongo (ref. 30) between *kamba haŋkor* "elbow" and *ce: haŋkor* "knee" (*ce:* = "foot, leg"). Finally, Goundam and Kaado *koro* may be viewed as a shortened form of *haŋkoro*, even though alternative analyses are possible (ref. 7).

We can return now to our comparison *ukunkula*/*honkoro*/*k'oŋk'oloŋ*. First, we must take into account the more recent data on the Songay of Gao mentioned above that give *haŋkoro* (and even *haŋakora*, DN95, cited in ref. 30) as a more accurate form than *honkoro*. Second, we shall assume for a while that Ehret's sound "laws" (detailed in ref. 5) are correct. If so, and still neglecting vocalic correspondences, the postulated 1st element should derive from pNS ***k'un-*, ***kun-* or ***k^{h}un-* in the case of Kunama, pNS ***k'aŋ-* in the case of Songay, and pNS ***k'on-* or ***k'oŋ-* in the case of Berta (it is not clear whether Berta *-ŋ-* is primary or represents a velar realization of *-n-* before *k'*). This points to a possible genetic heterogeneity, as already discussed about Bender's etymology ⟨2⟩. The postulated 2nd element for its part is no less problematic. Whereas Kunama *-kul-* fits well in Ehret's etymology ⟨1⟩ (pNS **k^{h}ul*), Songay *-kor-* and Berta *-k'olo-* do not (Songay, expected *-l-* instead of *-r-*; Berta, expected *h-* instead of *k'-*). Thus, from an Ehretian point of view, those three words can be no more than "look-alike", in spite of their striking similarity.

At this point, I fell on the word *i-koŋkoro-/koŋkora* "bend" (infinitive/verbal noun) in Bender's Kunama lexicon (ref. 20), which is not represented in any of the etymologies discussed above. Semantically, this word perfectly matches Songay *haŋkoro*

and Berta *k'oŋk'olo-ŋ*. According to Ehret's phonetic "rules", this Kunama word and the Songay one could both originate from pNS ***k'oŋk^{h}oro* (neglecting vowels). The reflex of ***k'oŋk^{h}oro* in Berta, however, should be either ***k'oŋhoro-ŋ* or ***k'oŋholo-ŋ*. This apparent irregularity can be fairly explained by an assimilation of the two velars at some pre-Berta stage, hence ***k'oŋk^{h}olo-ŋ* > ***k'oŋk'olo-ŋ*. The presence of *ŋ* may also have contributed to the preservation of the following velar stop as against its expected evolution to *h*. I therefore propose that Kunama *ikoŋkoro-*, Songay *haŋkoro* and Berta *k'oŋk'oloŋ* are genetically related and cannot be split into separate elements (etymology B in Table below). In contrast, Kunama *ukunkula* (etymology A in Table below) is unrelated and may be analyzed either as a unique duplicated root (hypothesis 1) or as a compound word (hypothesis 2).
In summary:

	Etymology A	Etymology B
Kunama	Hypothesis 1: *u-ku(n)kul-a* = one partially reduplicated root (i.e., root **k^{h}ul* , Ehret ⟨1⟩) "bend" Hypothesis 2: *u-kun-kul-a* = two roots (compound) (i.e., either part of Bender ⟨2⟩ or Ehret ⟨4⟩, plus Ehret ⟨1⟩) "arm-bend"	*i-koŋkoro-* "bend"
Songay	Zerma *-kulla* (Ehret ⟨1⟩) Kaado *-kulla*	*haŋkoro* originally "bend"
Berta	*?*	*k'oŋk'olo-ŋ* originally "bend" (?)

Admittedly, the proposal for Berta is weaker than the other ones. In particular, I found no evidence in favor of "bend" being the original meaning. Therefore, one cannot exclude an alternative interpretation. For example, Berta "elbow" could be a compound made of **k'on* ~ **k'oŋ* (etymology ⟨2⟩) – if the original meaning might have been "arm" – followed by another root not yet presented here but detailed below, namely ⟨8⟩ **kol* "to be round, to curve". The postulated pNS ***k'oŋkolo-ŋ* would have become ***k'oŋk'olo-ŋ*, rather than the expected ***k'oŋholo-ŋ*, following assimilation of the 2nd velar stop to the 1st one (*k* > *k'*) in pre-Berta. In this hypothesis, the elbow would thus be "the curved part of the arm".

⟨8⟩ E988 **kol* "to be round, to curve"

Koman: Uduk *kol* "to be curved"
Saharan: Kanuri *kolo* "to swing round, spin" [NS **w* punctive (?)]
Berta: *huhulu* "egg" [partially reduplicated stem + NS **-uh* n. suff.] →
Nubian: Dongolawi *kolɛ* "waterwheel" [NS **y* n. derivative suff.]

W. Jebel: Gaam SG *kòlòd*, PL. *kòlg* "egg"
Daju: pDaju **koloŋ* "to turn" [NS **η* punctive], Nyala *kleŋ* "round" [NS **η* n./adj. suff.]
W. Nilot.: Jyang (Dinka) *kol* "wheel"
E. Nilot.: Bari *lokolo?* "curved" [< **lo-kolol*, ENil. ADJ. PREF. **lV-* + partially reduplicated stem; Bari final **l* > *?*]
→ Bender (ref. 14) reports *huuhulu* in the Undu, Mayu, and Fadasi varieties, and *hoholo* ~ *hòhòlu* in the Gebeto variety. A non-reduplicated form *holo* occurs additionally in Fadasi.

• **What about Shabo in all that?** Upon looking in the Shabo lexicon provided by Fleming in MT VII (ref. 2), one cannot miss *koggod* "elbow". Gemination of *g* in this word may suggest that we are dealing with a reduplicated **kod* stem (hence something like **kodkod*). This is a more convincing analysis than the one I suggested earlier (ref. 3). Can this stem be related to Ehret's etymology ⟨1⟩ **k^{h}ul* "to bend (intr.)"? To answer that, I first attempted to find out the outcome in Shabo of the three pNS unvoiced velar stops **k*, **k^h* and **k′* posited by Ehret. Of course, such an approach is meaningful only if Shabo does really belong among NS languages, as hypothesized here, or if Shabo shares a direct common ancestor with pNS. Thanks to the excellent set ⟨10⟩ "breast" (Appendix 1), the Shabo reflex of pNS **k* can be surely identified as *k*. Etymology ⟨13⟩ (Appendix 1) "face; mouth, tooth" suggests that pNS **k′* evolved to *k′* ~ *k* in Shabo. Unfortunately etymologies involving pNS **k^h* are of too low quality with respect to Shabo to allow any firm conclusion. As for *d* of **kod* on the other hand, I found only one etymology (Appendix 4, ⟨26⟩, Shabo *k′add*) in which Shabo *d* corresponds to pNS **l*, and that one may be deluding. Indeed, Shabo *d* is geminated here, a feature which could possibly result from the addition of a dental suffix to a root ending otherwise in *l*, e.g., **k′al-d* (cf. pNubian **ηal-t.-(V)* > Midob Nubian *kada*, E534, etymology ⟨26⟩). Thus, the argument for connecting genetically Shabo **kod* with pNS **k^{h}ul* is not strong, but this possibility can certainly not be excluded at present.

On closer inspection of Fleming's Shabo lexicon (ref. 2), I fell on the curious word *gongo:d'e* "crooked limbs". On the one hand, a cognation with Shabo **kod* and pNS **k^{h}ul*, with partial reduplication and insertion of a euphonic *n* has to be considered. However, the presence of plain *d* in one Shabo word *vs* glottalic *d′* in the other would be unaccounted for. On the other hand, *gongo:d'e* is strikingly looking like either *ukunkula*, or the triad *ikoŋkoro*, *haŋkoro*, *k'onk'oloŋ*. Is there a genetic relation with one or the others? Here too, the evidence is weak and the question remains open. As a matter of facts, I attempted in Appendix 3 to find out to which pNS consonant Shabo *g* could correspond. There are several possible and non-mutually exclusive candidates (**ŋg*, **g*, **g′*, **η*, **k^h*), but apparently neither the **k* nor the **k′* that would support cognation. The best etymologies support a pNS **ŋg* correspondence to Shabo *g* (cf. ⟨21⟩ "jaw",

⟨23⟩ "to clear throat"). In Appendix 2, I present two etymologies where Shabo *d'* (of *gongo:d'e*) might correspond to pNS **l* (as in *ukunkula*). In contrast, I found none in which Shabo *d'* would represent pNS **r* (as in *ikoŋkoro*, etc.).

Part 2. Some new lexical items linking Shabo and/or Kadu to NS

In the course of the present research, I believe that I have uncovered some possible cognates that, to the best of my knowledge, have never been reported. For full details, the reader is invited to refer to the Appendices.

⟨9⟩ Shabo: *b'akiwon* "cheek" (with metathesis of the 1st two consonants?); NS E970 **ka:p'* "lower part of face".

⟨11⟩ Shabo: *ke:ji* "thin" (?); or *c'und'e* "narrow"; NS E1018 **k^{h}ayn* "to be little, small"; Kadu: Kulishi (ref. 36) *(shi)shi?n^{y}a* "small", or more convincingly Yegang (ref. 36) *kan^{y}n^{y}a* (pl.) "children" (ref. 3).

⟨12⟩ Shabo: *akush* "skin" (?); NS E1062 **k^{h}u:r* or **k^{h}u:d* "hide, skin"; Kadu: Mudo (ref. 36) *t.iŋgòla* (PL. *kòla*) "outer bark", Krongo *kwaala* (pl. *ni-kwaala*) "bark (of tree)" (ref. 17).

⟨14⟩ Shabo: *d'unk'u* "brain"; NS E1322 **loŋk^{h}* "crown of head" (→ "crest"; "head"; "fontanelle"); Kadu: Tolibi (ref. 16) *dòŋɛ* (PL. *na-dòŋɛ*), Krongo (ref. 17) *d'od'on^{y}i* < **d'od'oŋi* (PL. *ni-d'od'on^{y}i*) "brain".

⟨16⟩ Shabo: *d'ingi* ~ *d'iŋi* "penis"; NS E730 **t.^{h}iŋk'* "stick"; Kadu: Tolibi (ref. 16) *dukwe* (PL. *ku-dukwe*) "stick, club", Krongo (ref. 17) *ti-d'ikwa* (PL. *ni-d'ikwa*) "spear (n.)".

Semantics: cf. Maban Maba *turuk* "penis", Kodoi *toru:k* "penis", Kodoi *turok* "stick, wand"; also Mimi (GD) *komis* "rod, wand, staff, penis" (ref. 8).

Note that in both etymologies ⟨14⟩ and ⟨16⟩ , Shabo *d'* corresponds to Kadu Tolibi *d* and Krongo *d'*. In contrast NS shows **l* (> *l*) in ⟨14⟩ *vs* **t.^{h}* (> *c^{h}*, *t*, etc.) in ⟨16⟩ .

⟨17⟩ Shabo: *wa:d'i* "dew, wet"; NS E1401 **weθ* "to spill onto, wet down"; Kadu: Kufo *t.i:fud'ò*, Miri *t.i-fud'ò*, Talla *t.am-bad'ò*, Tolibi *t.im-b'id'ò* ~ *thòm-bòdò* (ref. 36 ~ ref. 16), Sangali *ca(ŋ)fòd'ò* "rain" (all from ref. 36 except where indicated); Tolibi *o bidhi* (ref. 16) "wet, damp" (Kadu f_2, ref. 3).

I suspect that Fleming would add somewhat provocatively "…and English *wet*, of course!".

⟨20⟩ Shabo: *aguc'e* "to flee"; NS E508 **ŋgu:z* "to run".

⟨23⟩ Shabo: *gu:to* "to clear throat"; NS E512 **ŋgwɛd* "throat".

⟨25⟩ Shabo: *getumba* "to turn" (if < **ge-tumba* ?); NS E848 **t'amp* "to turn around.

⟨27⟩ Shabo: *dondom* "thick, fat", probably < **dom* by reduplication; NS E315 **dʸ'we:m* "fat, thick"; Kadu: Mudo (ref. 36) *v:d'vnd'v*, possibly < **v:d'vmd'vm* "thick".

Conclusions

1. While searching for the pNS precursor of the Kunama word *ukunkula* "elbow", I was struck by the very frequent occurrence in NS languages of words with the general shape *kVl* and expressing diverse meanings related to "bend" and to body anatomy. To my great astonishment, I noted that various scholars assembled such words from distinct languages into different etymologies that display only very partial overlap. A good number of those etymologies have been examined or just cited here. The observed discrepancies indicate that much additional work is needed to arrive at a consensus regarding the reconstruction of pNS.

2. The three words Kunama *ukunkula*, Songay of Gao *haŋkoro* and Berta *k'onk'oloŋ* have often been considered cognates. I propose here that while Kunama *ukunkula* is not genetically related to the other two, it is Kunama *i-koŋkoro-* that forms a cognate set with Songay *haŋkoro* and probably Berta *k'onk'oloŋ*.

Kunama *ukunkula* "elbow" may either derive from pNS **k^{h}ul* "to bend" by partial reduplication (Ehret) or consist of a compound word made up of *kon-a* "hand, arm" (Bender **KoN-* or Ehret **k'wɛy-n-*) plus Ehret **k^{h}ul* "to bend".

In contrast, Kunama *i-koŋkoro-*, Songay *haŋkoro* and Berta *k'onk'oloŋ*, supposed to mean originally "bend", are not amenable to internal analysis.

A possible cognation of Shabo *gongo:d'e* "crooked limbs" with either *ukunkula* or *ikoŋkoro*, *haŋkoro* and *k'onk'oloŋ* is considered, but awaits further testing of its validity.

It has been proposed that *ukunkula* participates in two global etymologies. To properly assess that hypothesis, a more accurate delineation of the global etymologies in question is needed.

3. I propose here some potential cognates between Shabo and/or Kadu and Nilo-Saharan that have (to the best of my knowledge) not been described yet.

4. The present work is based on the assumption that Ehret is right in reconstructing three pNS unvoiced velar stops. In addition, I could not escape making personal choices between alternative possible views. Of course, not all readers will agree. Conceivably some of them will be upset by my reliance (even though rather critical) on reconstructions proposed by Ehret. Others will find me excessively restrictive with regard to the semantic latitude I allow, particularly in the field of human anatomy. Therefore this article is open to discussion.

Acknowledgments

I am indebted to Prof. Jeffrey Heath, University of Michigan, for enriching and consolidating my data on Songay. It should be emphasized however that I would be fully responsible for any faulty interpretation or misuse of his data. I would also like to thank Mrs Isabelle de Kaenel, Mrs Anne Parrical and all their collaborators at the Bibliothèques Universitaires de Médecine et Santé Publique (BDFM) for their invaluable help in providing me with so exotic books and articles.

References (1) by number

1. Fleming 1991.
2. Fleming 2002.
3. Bürgisser 2004.
4. Blench 2004.
5. Ehret 2001.
6. Bender 1997.
7. Ducroz & Charles 1978.
8. Edgar 1991.
9. Doornbos & Bender 1983.
10. Siebert & Caudwell 2002.
11. Dimmendaal 2000.
12. Bender 1992.
13. Greenberg 1966.
14. Bender 1989.
15. Stevenson 1991.
16. Persson & Persson nd.
17. Reh 1985.
18. Bender 2000.
19. Rottland 1989.
20. Bender 2001.
21. Bengtson & Ruhlen 1994.
22. Greenberg 2002.
23. Starostin, G.S. nd.
24. Helimski Eugen et al. nd.
25. Starostin S., et al. nd.
26. Siebert, et al. 2002.
27. Prost 1956.
28. Heath 1998a.
29. Heath 1998b.
30. Heath 1998c.
31. Peace Corps nd.
32. Tersis 1968.
33. Heath Jeffrey. pc., 2005.

34. Oumarou 1993.
35. Rueck & Christiansen nd.
36. Schadeberg Thilo C. 1994.
37. Harrison et al. 1997.

References (2) by name(s)

Bender, M. Lionel. 1989. Berta lexicon. *Topics in Nilo-Saharan Linguistics,* Ed. by M.L. Bender, 271–304. Hamburg: Helmut Buske.

Bender, M. Lionel. 1992. Central Sudanic segmental and lexical reconstructions. *Afrikanistische Arbeitspapiere* (AAP) 29. 5–61.

Bender, M. Lionel. 1997. *The Nilo-Saharan Languages: A Comparative Essay*. 2nd Ed. Munich: LINCOM EUROPA.

Bender, M. Lionel. 2000. Nilo-Saharan. *African Languages. An Introduction,* Heine Bernd and Nurse Derek (eds), 43–73. Cambridge, UK: Cambridge University Press.

Bender, M. Lionel. 2001. English-Kunama lexicon. *Afrikanistische Arbeitspapiere* (AAP) 65. 201–253.

Bengtson, John D. & Merritt Ruhlen. 1994. Global etymologies. *On the Origin of Languages,* Ruhlen M., 1994, 277–336. Stanford, Calif.: Stanford University Press.

Blench, Roger. 2004. The Kadu languages and their affiliation: between Nilo-Saharan, Niger-Congo and Afro-Asiatic. Paper presented at the IXth Nilo-Saharan Conference in Khartum, 10–13th Febr., 2004, and subsequently revised. Draft available on: http://homepage.ntlworld.com/roger_blench/RBOP.htm

Bürgisser, Philippe. 2004. Some thoughts about Shabo, Ongota and the Kadu family of languages. *Mother Tongue* 9.192–215.

Dimmendaal, Gerrit J. 2000. Number marking and noun categorization in Nilo-Saharan languages. *Anthropological Linguistics* 42. 214–261.

Doornbos, Paul & Bender. L.M. 1983. Languages of Wadai-Darfur. *Nilo- Saharan Language Studies,* Ed. by M.L. Bender, 43–79. East Lansing: Michigan State University.

Ducroz, Jean-Marie & Marie-Claire Charles. 1978. *Lexique Soŋey (Songay)-Français. Parler Kaado du Gorouol*. Paris: L'Harmattan.

Edgar, John T. 1991. *Maba-group Lexicon*. Berlin: Dietrich Reimer.

Ehret, Christopher. 2001. *A Historical-Comparative Reconstruction of Nilo-Saharan*. Köln: Rüdiger Köppe.

Fleming, Harold C. 1991. Shabo: Presentation of data and preliminary classification. *Proceedings of the fourth Nilo-Saharan Linguistics Conference,* Ed. by M.L. Bender, 389–402. Hamburg: Helmut Buske.

Fleming, Harold C. 2002. Shabo: A new African phylum or a special relic of old Nilo-Saharan? *Mother Tongue* 7. 1–37.

Greenberg, Joseph H. 1966. *Languages of Africa*. 2nd revised Ed. The Hague: Mouton & Co.

Greenberg, Joseph H. 2002. *Indo-European and Its Closest Relatives. The Eurasiatic Language Family.* Volume 2. Lexicon. Stanford, Calif.: Stanford University Press.

Harrison, Byron, Annette Harrison, Michael J. Rueck & Mahaman Soumana. 1997. *Southern Songhay speech varieties. A sociolinguistic survey of the Zarma, Songhay, Kurtey, Wogo, and*

Dendi peoples of Niger. Niamey, Niger Republic: Summer Institute of Linguistics. Online at: http://www.sil.org/silesr/1999/004/zarmarpt4.pdf

Heath, Jeffrey. 1998a. *Dictionnaire Songhay – Anglais – Français. Tome I – Koyra Chiini, ou 'Songhay de Tombouctou'*. Paris: L'Harmattan.

Heath, Jeffrey. 1998b. *Dictionnaire Songhay – Anglais – Français. Tome II – Djenné Chiini, ou 'Songhay de Djenné'*. Paris: L'Harmattan.

Heath, Jeffrey. 1998c. *Dictionnaire Songhay – Anglais – Français. Tome III – Koroboro Senni, ou 'Songhay de Gao'*. Paris: L'Harmattan.

Heath Jeffrey. Personal communication, including data now published in: Heath, Jeffrey. 2005. *Tondi Songway Kiini (Songhay, Mali): Reference Grammar and TSK-English-French Dictionary*. Stanford Monographs in African Languages. Stanford, Calif.: Stanford University Press.

Helimski, Eugen, Sergei Starostin, Sergei Nikolayev, et al. nd. *Uralic database*. The Tower of Babel. http://starling.rinet.ru/Intrab.htm

Oumarou, Yaro Bourahima. 1993. *Eléments de description du Zarma (Niger)*. Thèse, 1993, Université Stendhal-Grenoble III.

Peace Corps language trainers, volunteers, and staff in Niger. nd. *Zarma Dictionary – Dictionnaire Zarma – Zarmaciine Kaamuusu*. http://www.bisharat.net/Zarma

Persson, Andrew & Janet Persson. nd. *English-Katcha dictionary from the papers of Roland Stevenson (zipped MS Word-file)*. Summer Institute of Linguistics. Editing and formatting by Roger Blench. http://homepage.ntlworld.com/roger_blench/RBOP.htm

Prost, André. 1956. *La langue SoNay et ses dialectes*. Dakar: Institut Français d'Afrique Noire.

Reh, Mechthild. 1985. *Die Krongo-Sprache (Niino Mo-Di)*. Berlin: Dietrich Reimer.

Rottland, Franz. 1989. Southern Nilotic reconstructions. *Topics in Nilo-Saharan Linguistics*, Ed. by M.L. Bender, 219–231. Hamburg: Helmut Buske.

Rueck, Michael J. & Niels Christiansen. nd. *Northern Songhay languages in Mali and Niger. A sociolinguistic survey*. Summer Institute of Linguistics. http://www.sil.org/silesr/1999/008/nsonghay.html

Schadeberg, Thilo C. 1994. Comparative Kadu wordlists. *Afrikanistische Arbeitspapiere* (AAP) 40. 11–48.

Siebert, Ralph & Simon Caudwell. 2002. *Sociolinguistic Survey Report of the Melo (Malo) and Mursi Languages of Ethiopia*. Summer Institute of Linguistics International. http://goto.glocalnet.net/maho/webresources/nilosaharan.html

Siebert, Ralph, Kati Siebert & Klaus Wedekind. 2002. *Sociolinguistic Survey Report on Languages of the Asosa – Begi – Komosha Area. Part I.* Summer Institute of Linguistics International. http://www.sil.org/silesr/abstract.asp?ref=2002–054.

Stevenson, Roland C. 1991. Relationship of Kadugli-Krongo to Nilo-Saharan: Morphology and Lexis. *Proceedings of the fourth Nilo-Saharan Linguistics Conference,* Ed. by M.L. Bender 347–369. Hamburg: Helmut Buske.

Starostin, George S. nd. *A comparative database of Dravidian languages*. The Tower of Babel. http://starling.rinet.ru/Intrab.htm

Starostin, Sergei, Anna Dybo, & Oleg Mudrak. nd. *Altaic database*. The Tower of Babel. http://starling.rinet.ru/Intrab.htm

Tersis, Nicole. 1968. *Le Dendi (Niger).* Bulletin de la SELAF (Société pour l'Etude des Langues Africaines), vol. 10. Paris.

Appendices

I have used *italics* to mark those elements of a comparison that seem to me questionable, whether for semantic or phonetic reasons. Square brackets contain Ehret's own comments. Round brackets contain additional data or comments of mine.

1. Ehret's etymologies containing pNS **k*, **k^h* and **k′* together with possible Shabo and Kadu cognates.

⟨9⟩ NS E970 **ka:p′* "lower part of face"/Koman: Uduk *kab′ash*, *kapash* 'chin' [NS **s^y* n. suff.]/Centr. Sud.: pCS **kpa* "lower part of face"/Songay: Zerma *gagaabe* "jaw"/Nubian: Kadaru *kamt.ò* < **kap′t.ò* "cheek" [**t.* n. suff.]//Shabo: *b′akiwon* "cheek" (with metathesis of the 1st two consonants?).

⟨10⟩ NS E1001 **ako*, **ko* "breast"/Koman: Uduk *ako*, Komo *ko*, Opo *kue* "breast"/Kunama: *ku-ta* "breast"/Songay: Zerma *ga-na* "breast"; *ga-n-de* "chest" (all Songay languages have *gand-*, ref. 7, 27–30, 37)/Maban: Maba *aŋu-n,* Runga *aŋgu-n,* Mimi *aku-n* "breast"/ Nubian: Nobiin *og*, Kaderu *òkò*, Dongolawi *og* "chest"/W. Nilot.: Ocolo *ko* "chest"// Shabo: *kowan* "breast" (ref. 2); *kokoŋ* "breast, chest"//*Kadu: 8 languages have -nugu or a similar form, but Mudo has just -nu "breast" (ref. 36) (→ genuine root = -nu- ?).*

⟨11⟩ NS E1018 **k^{h}ayn* 'to be little, small'/Kunama: *kennekennema* 'weak, thin' [**m* adj. affix]/Saharan: Teda *kinni* 'small'/Songay: Zerma *keyna* 'to be small; a little'/Nubian: Dongolawi *kinna* 'small; young'/Temein: *ki-kɛni-k* 'small'//Shabo: *ke:ji* 'thin'; or *c′und′e* 'narrow'//Kadu: *Kulishi (ref. 36) (shi)shi?n^{y}a* 'small', or more convincingly Yegang (ref. 36) *kan^{y}n^{y}a* (PL.) 'children' (ref. 3).

⟨12⟩ NS E1062 **k^{h}u:r* or **k^{h}u:d* 'hide, skin'/Koman: Uduk *ak^{h}ur* 'dry bark', *k^{h}ur* 'bark'/ Songay: Zerma *kuuru* 'skin, hide' (all Songay languages have *ku(:)ru*, except Dendi *kuyu* with regular shift *-r-* > *-y-*, ref. 7, 27–30, 37) //Shabo: *akush* 'skin'//Kadu: Mudo (ref. 36) *t.iŋgòla* (pl. *kòla*) 'outer bark', Krongo *kwaala* (PL. *ni-kwaala*) 'bark (of tree)' (ref. 17).

⟨13⟩ Fleming (ref. 1, 2) proposes that Shabo *k′aw-k* 'face' is related to Shabo *kaw* ~ *k′awut* ~ *kau-se* 'mouth', *kaw* ~ *k′aw* 'tooth', which in turn is related to NS E1076 **k′a* 'to take into the mouth' and its derivative E1086 **k′a-′y* 'to chew'/Koman: Uduk *k′a* 'to gnaw, chew', Gumuz **k′ën* 'to bite'/E. Centr. Sud.: **k′a* 'to bite'/Kunama: *kayo-* 'to chew'/Songay: Zerma *hɛŋ* < **hayŋ* 'tooth' (Kaado *heŋ*, Gao *hinga* ~ *hin^{y}a*, Djenné *hin^{y}e*, Dendi *hind^{y}e* ~ *hinje*, ref. 7, 27, 29, 30, 32, 37)/ Surmic: Zilmamu *kauw-* 'to bite'/E. Nilot.: Bari *kò-ja* 'to bite, sting (bee, etc.)'/W. Nilot.: pW.Nil. **kac* 'to bite'/Rub Ik *nk′ak′-* 'to eat', *ak′ak′uŋ* 'jaw'; *k′idz-* 'to bite'.

2. Could pNS **l* have evolved to *d′* in Shabo?

⟨14⟩ Shabo: *d′unk′u* 'brain'//NS E1322 **loŋk^h* 'crown of head'/Centr. Sud.: **lò* 'crown of head, crest'/Kunama: *ko-loka* 'crest' [with movable **k^h*; considered a loan by Ehret]/Songay: Zerma *loŋko* 'fontanelle antérieure (Fr.)' (Kaado idem. Certainly also Songay of Djenné *lòŋgoy* 'top of baby's head', ref. 7, 29)/Berta: *a-lu* 'head' [NS **a-* n. derivative pref.; regular **nk^h* > **h* > Ø]/W. Jebel: Gaam **lòòi* [< **logo-i-* with NS **′y* n. suff.] 'fontanelle'/E. Nilot.: Maasai *ɛ-lʊkʊnya* 'head' [NS **n^y* n. suff.]//Kadu: Tolibi (ref. 16) *dòŋɛ* (PL. *na-dòŋɛ*), Krongo (ref. 17) *d′od′on^{y}i* < **d′od′oŋi* (PL. *ni-d′od′on^{y}i*) 'brain'.

⟨15⟩ Shabo: *d'e* 'to know'/NS E1305 **le:h* 'to watch, observe'/Kunama: *li-* 'to look', *lili-k* 'to watch attentively, contemplate' [NS *k^h* iterative]/Saharan: Kanuri *li* 'to learn'/For: *la-* 'to see (past)' [NS **-a* punctive], *lauŋ-* 'to know' [NS **ŋ* punctive]/W. Nilot.: Ocolo *let.* 'to see' [NS **θ* intensive], *lyɛw* 'to investigate, spy out' [NS **w* punctive]/E. Nilot.: Maasai *-leen* 'to scout' [NS **n* durative]/Rub: Ik *i-l^{h}ye-* 'to know' [**i-* v. class pref.] //*Kadu: Miri (ref. 36) aji 'to watch'; Tolibi (ref. 16) ija 'to see'.*

pNS consonants other than **l* may also result in *d'* in Shabo: e.g.,

⟨16⟩ Shabo: *d'ingi ~ d'iŋi* 'penis'//NS E730 **t.^{h}iŋk'* 'stick'/Koman: Uduk *c^{h}iŋkira?* 'sticks, twigs' [NS **r* and **-ah* n. suffixes]/Centr. Sud.: pEast cs **ti* or **t'i* or **tri* 'log, pole'/Kunama: *tinkisha* 'rod, wand, crutch, …' [NS **t^{yh}* n. suff.]/*Saharan: Kanuri tigë 'body*//Kadu: Tolibi (ref. 16) *dukwe* (PL. *ku-dukwe*) 'stick, club', Krongo (ref. 17) *ti-d'ikwa* (PL. *ni-d'ikwa*) 'spear (n.)'.

Semantics: cf. Maban Maba *turuk* 'penis', Kodoi *toru:k* 'penis', Kodoi *turok* 'stick, wand'; also Mimi (GD) *komis* 'rod, wand, staff, penis' (ref. 8).

⟨17⟩ Shabo: *wa:d'i* 'dew, wet'//NS E1401 **weθ* 'to spill onto, wet down'/Koman: Uduk *wus* 'to wash (clothes)'/For: *wese* 'wet'/Songay: Zerma *wesi* 'to drain, scoop out (liquid)'//*Nubian: Dongolawi uss- 'to defecate'*/Daju: pDaju **osho* 'year; rainy season'/*W. Nilot.: Ocolo wɛt. 'to paint, smear on, cover with grease'*/Rub: Soo *weθit-* 'to fill' [NS **t^h* continuative suff.]//Kadu: Kufo *t.i:fud'ò*, Miri *t.i-fud'ò*, Talla *t.am-bad'ò*, Tolibi *t.im-b'id'ò ~ thòm-bòdò* (ref. 16), Sangali *ca(ŋ)fòd'ò* 'rain' (all from ref. 36 except where indicated); Tolibi *o bidhi* (ref. 16) 'wet, damp'.

3. Which pNS consonant(s) might correspond to Shabo *g* ?

⟨18⟩ Shabo: *go:-ma*, *go:-ta* 'to burn' (ref. 1)//NS E468 **go:(h)* or *g'o:(h)* 'to burn (intransitive), shine'/Centr. Sud.: pCS **gu* or **g'u* 'to burn, shine'/Saharan: Kanuri *gut* 'to warm in sun' [**t.* causative suff.]/W. Jebel: Gaam *gòon-* 'to shine' [NS **n* durative]; Aka, Molo *aguwa*, Kelo *a?uwa* 'moon'/Surmic: S. Surmic **go:* 'fire'.

⟨19⟩ Shabo: *ge:ra* 'to climb'//NS E1291 **k^{h}ur* 'hill'/Centr. Sud.: pW. CS **ku* 'hill'/Saharan: Kanuri *këri* 'hill'/Nubian: Dongolawi *kutt-* < **kurt-* 'mound' [NS **t^h* n. suff.]; Dongolawi, Nobiin *kurun^y* [NS *n^y* n. suff.] 'hamp'; Birkid *kur* 'stone'/Taman: Tama *kura* 'stone, hill'/Temein: *kurɛt.* 'stone' [**t.* SG marker]/W. Nilot.: Jyang *kur* 'stone'.

⟨20⟩ Shabo: *aguc'e* 'to flee'//NS E508 **ŋgu:z* 'to run'/ Koman: Uduk *gus* 'to run; to be afraid of; to flow of river, blood, water'/Saharan: Kanuri *ŋguza* 'to flee, run away' /*W. Jebel: Gaam gujguj 'shuffling'.*

⟨21⟩ Shabo: *ga:m(a)* 'jaw' (ref. 2)//NS E513 **ŋgwɛ:m* 'chin'/Kunama: *goma* 'chin; beard'/Saharan: Kanuri *ŋgumi* 'chin' /*Nubian: Dongolawi gumur 'neck', Nobiin gumur 'nape' [NS *r n. suff.].*

⟨22⟩ Shabo: *gutann* 'man (older than 45 years)' (borrowing from Koman?)//NS E427 **g'waθ* or **g'wa:θ* 'adult, grown up person'/Koman: Uduk *gwasan*, *gwat.^{h}an* 'males' [NS **n* n. suff. (as PL. ?)]/*Rub: pRub *g'wasat (pl. *g'was) 'woman'.*

⟨23⟩ Shabo: *gu:to* 'to clear throat'//NS E512 **ŋgwɛd* 'throat'/Kunama: *gorongora* 'larynx, Adam's apple'/Saharan: Kanuri *ŋgogolto* 'throat' [NS **t^h* n. suff.]/For: *goroŋgoroŋ* 'throat' [NS **ŋ* n. suff.]/*Maban: Maba kongorik 'cheek' [NS *kh n. pref., -k sg suff.] (see also Aiki aŋgu, ref. 8)*/ Taman: Tama *ŋòròŋòrò* 'throat'/E. Nilot.: Bari *gworo?* 'throat'.

⟨24⟩ Shabo: *go:re* 'walk slowly due to disease'//NS E523 **ŋa:r* 'sprain, injured limb'/Saharan: Kanuri *ŋgarŋgar* 'pain in foot after long march'/W. Jebel: Gaam *ŋari-* 'to sprain, twist, wrench (limb)'.

The next comparison is out of purpose, but I find it worth mentioning:

⟨25⟩ Shabo: *getumba* 'to turn' (**ge-tumba* ???)//NS E848 **t'amp* 'to turn around'/Koman: Uduk *t'am* 'to coil', *t'am is* 'to wind' [*is*: reflex]/Saharan: Kanuri *tambar* 'to roll' [NS **r* iterative].

4. Could pNS **l* lead to Shabo *d* ?

There are few data supporting such a correspondence, for example:

⟨26⟩ Shabo: *k'add* 'tongue' (ref. 2).

Kadu: Mudo *ëëdò* (PL. *niŋgò:dò*), Yegang *t.aŋvd'ò* (PL. *and'ò:-nɛ*), Kufo *ŋòd'ò* (PL. *ni-ŋòd'ò*), Miri *(ŋ)ŋad'ò* (PL. *naga-ŋad'ò*), Talla *aŋgad'uk* (PL. *ni-aŋgad'uk*), Tolibi *ŋgòd'ò* (PL. *nugu-ŋgòd'ò*), Sangali *a-ŋgòd'ò* (PL. *nò-ŋgòd'ò*), Krongo *cò:d'ò* < **k^{y}ò:d'ò* ? (PL. *ni-jò:d'ò*), Talasa *t.òŋòd'ò* (PL. *ni-d.òŋòd'ò*) 'tongue' (ref. 36). The pKadu root seems to be something like **ŋ(g)òd'ò*. However, the initial **ŋ(g)* is lacking in the Mudo sg and the Yegang pl. forms, leaving open the possibility that the actual root is limited to **òd'ò* (ref. 3).

Either NS E1098 **k'ol* 'to chew'/Centr. Sud.: pCS **k'o* 'mouth'/Kunama: *a-kalma* 'molar tooth' [NS **m* n. suff.]/Berta: *k'òl-* 'to eat'/Nubian: Hill Nubian **kol* 'to eat'.

Or NS E534 **ŋɛl* 'tongue'/Centr. Sud.: pCS **(n)drɛ* 'tongue'/Kunama: *ŋela* 'tongue' (Bender *ŋeela*, ref. 20)/Nubian: pNubian **ŋalt.-* 'tongue' [**t.* n. suff.] [Hill Nubian **jald.*, Birgid *natti*, Nobiin *nar*, Midob *kada*] plus a number of other reflexes with meanings such as 'collarbone', 'neck', 'voice', the cognation of which is difficult to accept for semantic reasons.

Or NS E1008 **k^{h}al* 'to sip, sniff, taste'/*Koman: Uduk khal 'to give an odor, smell'*/Kunama: *-kal-* 'to lick, lap'/Berta: *hala ~ hale?* 'tongue'/Nara: *kal* 'to eat'/Nubian: Dongolawi *kal* 'to eat, take food, bite (of insect)'/W. Jebel: Gaam *këlad* (PL. *këlal^{h}g*); Aka *kala*, Molo and Kelo *këla* 'tongue'/E. Nilot.: pE. Nil. **kɛla-*/**kala-* 'tooth'.

However, in most etymologies, Shabo *d* does not correspond to pNS **l* :

⟨27⟩ Shabo: *dondom* 'thick, fat', probably < **dom* by reduplication//NS E315 **d^{y}'we:m* 'fat, thick'/Koman: *dwamadwam* 'thick (of wood, rope)'/Centr. Sud.: Baledha *∂u* 'fat (n.)'/Saharan: Kanuri *jëm* 'ideophone of something very big'/W. Jebel: Gaam *d^{y}imil-* 'to be fat, thick' [NS **l* iterative as deadjective?]/E. Nilot.: Bari *dvma* 'big'//Kadu: Mudo (ref. 36) *v:d'vnd'v* 'thick'.

⟨28⟩ Shabo: *do:gu* 'to disappear', *addake* 'be absent'//NS E759 **t.'òw* 'to cease to function, stop doing'/Koman: Gule *-tòs* 'to kill' [NS **s^{y}* causative ?]/Centr. Sud.: Baledha *d?o* 'prendre fin, finir (Fr.)'/Kunama: *tu-* 'to die'/*For: tuo- 'to be unable*'/W. Jebel: Molo *took-* 'to finish' [NS **k* causative]/W. Nilot.: pW. Nil. **t.òw* 'to die'/E. Nilot.: pE. Nil. **-tua(n)-* 'to be dead' [NS **n* durative]/Rub: Ik *ts'ooniam* (PL. *ts'ooniik*) 'dead person' [NS **n* modifier suff. + number suff.], *ts'e-it-* 'to extinguish (fire)' [**t.* causative suff.]//Kadu: Tolibi (ref. 16) *udhò*, Kado (ref. 17) *t-uutu-ŋ* 'to extinguish (fire)' [*-ŋ* transifier suff.].

Note however that one could rather compare Shabo *do:gu* 'to disappear' and/or *addake* 'be absent' with Bender Centr. Sud. **leg* 'to finish' (Lokai of Madi *legi*, Mamvu *regu*, Mangbetu *-lɛku*, Kresh *lògo* and *leke*, ref. 12).

5. Unusual signs used for transcription and abbreviations

– *t* (e.g.) = alveolar
– *t.* = dental
– *t′* = glottalic
– t^h = aspirated
– t^y = prepalatal
– l^h = voiceless *l*
– *c* = pre-palatal voiceless affricate (Engl. *ch*)
– *j* = pre-palatal voiced affricate (Engl. *j*)
– *ë* = French central vowel as in *le* (definite article)
– *ò* = open *o*
– *VV* or *V:* = long vowel
– *V^* = nasal vowel
– *?* = glottal stop
– @ = vocalic symbol of unknown significance in ref. 16

As for Kadu dental stops, there is a great deal of inconsistency between sources (ref. 15–17, 36). In case of uncertainty, I have maintained the original transcription which may differ from the one generally adopted in the present article.

– adj. = adjective
– Bnnn = Bender's etymology, nb. nnn (ref. 6)
– BRnn = Bengtson and Ruhlen global etymology, nb. nn (ref. 21)
– Ennnn = Ehret's etymology, nb. nnnn (ref. 5)
– Fr. = French
– G/NSnnn = Greenberg's NS etymology, nb. nnn (ref. 13)
– GD = (Mimi of) Gaudefroy-Demombynes (ref. 6, 8)
– n. = noun
– NS = Nilo-Saharan
– pref. = prefix
– p = proto
– suff. = suffix
– v. = verb
– ⟨⟩ = etymology numbering
– [] in Ehret's (E) etymologies (ref. 5) = addendum or comment by Ehret himself

The problem of pan-African roots

Roger M. Blench
Kay Williamson Educational Foundation

A problem of establishing macrophylic relations in Africa is the existence of common lexical items that are shared between the phyla in ways which do not suggest can they be used as evidence for a genetic connection. This suggests that proposals for large-scale language classification in Africa may be flawed by a failure to consider the transphylic distribution of many roots. The paper gives some examples of such lexical items and canvasses possible explanations;

a. they are ancient loanwords
b. they are convergent because of common phonaesthemes
c. African language phyla really are all related
d. they are retained from an early stage of world language diversification

Some roots seem to have a wider distribution in Eurasia, which suggests that (d) may be relevant in some cases.

….at the time of his compilation … the desire to astonish the World by the number and variety of Languages, and to supply materials to the builders of Philological Castles in the air, based upon words, brought together, and their fancied resemblance: with this object the compiler collected words from every part of Africa, not only of Languages,, but of Dialectal Varieties of Languages, quotations from published works, or from manuscripts.

Cust (1883: 27)

1. Introduction

In a paper[1] "Is Niger-Congo simply a branch of Nilo-Saharan?" (Blench 1995) I proposed that Niger-Congo and Nilo-Saharan formed a single macrophylum and adduced lexical, phonological and morphological evidence for this hypothesis. Although two

1. Parts of this paper were originally presented as an appendix to a paper on the proposed Niger-Saharan macrophylum at the VIIth Nilo-Saharan Conference in Vienna, 2–6th September, 1998. Although it was sent to the editor, Norbert Cyffer, in 1999 and proofed in 2000, it has never appeared and will probably now not be published.

further papers (Blench 2007, in press) present additional evidence for the Niger-Saharan hypothesis, a consequence of trawling the literature has been the identification of proposed Niger-Saharan glosses are also shared with Afroasiatic and even Khoisan, which rules them out as evidence for a genetic connection. Some of the lexical items presented in Blench (1995) as evidence for Niger-Saharan turn out to have a still wider distribution in Africa. Examples are *#kulu* "skin, hide", *#kulu* "knee", *#kuru* "tortoise, turtle". This is rather surprising, since even those addicted to wide-ranging hypotheses do not usually consider all African language phyla to form a single macrophylum. Some roots also seem to have a wider distribution in Eurasia, which may point to early diffusion with human expansion. Blench (1997) presented some preliminary evidence for some of these widespread roots and proposed a type of phonaesthetic convergence related to cultural salience, in particular the earliest phases of human expansion, when gathered aquatic resources such as crabs and turtles were of considerable dietary importance. Clearly, this is relevant, but will not explain every case; why should body parts such as "knee" be more convergent than more "head"? Westermann (1927) in his pioneering identification of "West Sudanic" common lexemes (Niger-Congo in modern terms) also identified *Wanderworte*, "wander-words" that show up in widely differing language families in similar form. An example is *aku*, for grey parrot (*Poicephalus senegalensis*) recorded from Senegambia to Southern Sudan and even in diaspora communities in Brazil. Why should this word conserve its form when many more culturally salient birds have names that are replaced and diversify phonologically?

The existence of transphylic roots has the more general consequence that some elements of proposals for large-scale language classification in Africa may be flawed by a failure to consider these wider distributions. They also suggest a methodological problem for the comparativists, which depends broadly on lexical and phonological diversification. Particular lexemes that are clearly not recent loanwords, simply do not diversify like others, without clear semantic motivation. For the comparative method to work, these have to be leached out of datasets. This paper[2] gives some examples of such pan-African lexical items and considers what hypotheses can be used to explain their occurrence. Possible explanations that have been canvassed are:

a. ancient loanwords
b. convergence through common phonaesthemes
c. African language phyla really *are* all related
d. transphylic roots are retained from a very early stage of language diversification

---

2. I would here like to acknowledge the influence of Hal Fleming, for whose Festschrift this has been prepared, as well as the pioneering work in editing Mother Tongue and making available speculations on the links between world language phyla. Thanks also to John Bengtson for additional cognates.

It is important to emphasise that not all words with a transphylic distribution in Africa belong to a marked conceptual set or have an evident phonaesthetic element. The tables presented below represent preliminary datasets intended to identify common forms encountered in the search for Niger-Saharan roots.

2. Setting out the evidence

Broad transphylic comparisons require large data tables, and inevitably draw on a wide range of materials. Earlier scholars, such as Westermann and Greenberg, did not cite a reference for specific lexical items and did not give a complete bibliography of sources. This is a long, tedious task, takes up considerable space and time and may have seemed unnecessary. Moreover, those who pioneer in a field rarely heed the scholarly conventions of a less trustful age. But science is nothing if not about repeatability; we should be able to check the claims of historical linguists just as much as those of laboratory scientists. Language citations therefore provide sources, and proto-languages, marked by starred forms, must be carefully treated. Where I have been unable to confirm other authors' cited attestations in the data tables below these are silently omitted. Semantically, the datasets given below are extremely conservative, since permitting a wide range of comparanda is the most common critique of "world etymologies".

A common but problematic practice in this area is the citation of starred forms to represent families or even phyla. If such forms are based on the intensive reconstruction of a small group of well-studied languages this may have some validity, but generally these forms are quasi-reconstructions based on rapid inspection of purported cognates, and often the source is frankly mysterious. Thus Ehret (2001) cites starred forms for proto-Central Sudanic, the source of which is unclear. Only one reconstruction of Central Sudanic has been published (Bender 1992) but Ehret does not cite this and his forms do not agree with it. Bender (1997: 131 ff.) has a section titled "items linking N-S and N-C" where he cites a number of forms for *N-C. These eschew the two major published sources (Westermann 1927; Mukarovsky 1976–77) and list forms not recognisable to scholars of Niger-Congo. Starred forms must therefore be treated with a considerable degree of scepticism unless their pedigree is well-established.[3]

3. Another common problem is that the "Common Bantu" reconstructions of Guthrie (1967–1971) are frequently cited as Proto-Bantu. Guthrie did not intend this to be the case and many of these forms are demonstrably *not* Proto-Bantu.

3. Data tables: Pan-African roots

The following tables present datasets for pan-African roots and include some examples of extra-African cognates. The forms preceded by # are common forms created for the purpose of reference rather than reconstructions. A gloss is only given when the meaning differs from the head gloss.

1. #ɓwoN come

Phylum	Group	Language	Attestation	Comment	Source
NS	ES	proto-Daju	*ɓoN		RCS
NS	ES	proto-Nilotic	*ɓun	PL. *pɔ	D
NS	ES	Nuer	**ben**		RCS
NS	ES	Padang Dinka	**bɔ**		RCS
NS	ES	Mabaan	**bɛɛd**		RCS
NS	ES	Nara	**wɔ**		E
NS	Fur	Fur	**bo-**	reach	Jakobi (1990)
NC	Dogon	Walo	**wá**		RMB
NC	Mande	Guro	ɓɔ		ALMCI
NC	Atlantic	Mankanya	**bi, ban**	*arriver*	GS
NC	Ịjọ	Proto-Ịjọ	*ɓó		KW
NC	Gur	Deg	**bà**		ALGCI
NC	Adamawa	Waja	**ba-**		Kleinewillinghöfer (1991)
NC	Kwa	Twi	**bà**		Chr
NC	WBC	Ọnịca Igbo	**-bɪa**		Williamson (1972)
NC	EBC	Pe	**bén**		RMB
AA	W Chadic	Yiwom	**bɛ̀l**		JI
AA	W Chadic	Kulere	**bo**		JI
AA	W Chadic	Tsagu	**bàà**		JI
AA	C Chadic	Tera	**ɓa**		JI
AA	E Chadic	Mawa	**ooboŋ**		Jim Roberts (p.c.)

Commentary: Not attested in Khoesan and Afro-Asiatic outside Chadic. Westermann (p. 209) noted that this word frequently shows up as a future auxiliary in Niger-Congo languages. Dimmendal (1988: 35) notes that the irregular plural *pɔ must be reconstructed to Proto-Nilotic and forms with initial p- appear elsewhere, hinting at a still greater time-depth. Palatalisation is scattered throughout Niger-Congo but nowhere forms a consistent pattern. cf. also some Indo-European forms, e.g., French *viens!*

Ref: W:209, E:563

2. #keri to split, cut, break

Phylum	Group	Language	Attestation	Gloss	Source
NS	Kuliak	Ik	**ka-uƙot-**	cut	Heine (1999)
NS	Koman	Uduk	**kwal**	split	Bender (1983)
NS	Berta	Undu	**kíírà**	split hard substance	RCS
NS	ES	Dongola Nubian	**gor**	cut grass	RCS
NS	ES	Nyimang	**kɪr**	cut	RCS
NS	Fur	Fur	**karr-**	split	Jakobi (1990)
NS	Saharan	Daza	**kɔr**	*couper*	Le Coeur (1950)
NS	Saharan	Beria	**kɔ́t**	*couper*	JC
NS	Songhay	Zarma	**kŏrtù-**	tear, split	BKW
NC	Dogon	Ana	**keɗele**	cut	RMB
NC	Ijoid	Proto-Ijọ	**pɛ́lɛ́**	cut	KW
NC	Mande	Bambara	**kárí**	cut	Ba
NC	Atlantic	Bullom	**keth-**	*couper*	GS
NC	Kru	Nyabwa	**cei**	cut	ALKrCI
NC	Ubangian	Ndunga-le	**-kèlé-**	cut	Mo
NC	WBC	Ọnịca Igbo	**-cà**	cut	Williamson (1972)
NC	EBC	Tarok	**ca**	break, cut	RMB
AA	Omotic	proto-Dizoid	**k'aat'-**	cut	Bender (2003: 302)
AA	Central Cushitic	Arbore	**k'uur-**	cut	LS
AA	Agaw	Bilin	**kər**	break	LS
AA	Cushitic	Dahalo	**k'eer-**	chop	LS
AA	South Cushitic	Iraqw	**qeet**	be broken	MK
AA	Berber	Kabyle	**cerref**	*couper*	Dallet (1982)
AA	Chadic	Masa	**káná**	cut	JI
AA	Chadic	Dghwede	**kɬà**	cut	JI
AA	Chadic	Warji	**kɔ́n**	cut	JI
AA	Chadic	Ngas	**can**	cut	JI
KS	Southern	!Xóõ	**t'kx'āla**	open meat to flatten it	Traill (1994)
KS	Central	Proto-East Khoe	***kade**	cut flesh in strips	Vo97

(*continued*)

2. #keri to split, cut, break (Continued)

Commentary: This lexeme has been proposed as a "world etymology", and the cognate set would presumably then including English "cut" (see Bender 1997: 122 for more Nilo-Saharan examples). In Cushitic and Omotic these forms are very widespread (see examples under **k'er** "split" and **kaal-ta** "axe" in Lamberti & Sottile 1997: 411, 435).

Refs: Gr:80; G:97,135,154; B:122, 133

3. #kulu knee

Phylum	Family	Language	Attestation	Comment	Source
NS	Shabo	Shabo	**hutu/kutti**		Teferra (p.c.)
NS	Kuliak	Ik	**kútúŋ**		Heine (1999)
NS	Koman	Kwama	**dugʊl**		RCS
NS	Berta	Berta	**guʃuŋ**		RCS
NS	Kunama	Kunama	**tùgà**	?C	Bender (ms.)
NS	Maba	Mesalit	**kàdíŋó**		Ed
NS	Fur	Fur	**kùrù**		Jakobi (1990)
NS	ES	Kenzi	**kur(ti)**		RCS
NS	CS	Mangbetu	**nɛ̀-káátì**	PL. ɛ̀-	Demolin (p.c.)
NS	Saharan	Beria	**kórú**		JC
NS	Kadu	Katcha	**kúúgɛ**	PL. **nu-gúúgi**	Sch94
NC	Dogon	Tebul Ure	**kundugɔ**		RMB
NC	Ijoid	Biseni	**íkónḯée**		KW
NC	Mande	Soninke	**xuruŋi/o**	PL. **-onu**	VV
NC	Atlantic	Bijogo C	**kunu**	PL. **ŋanu** [*genou*]	GS
NC	Kordofanian	Tima	**kuruŋa**		RCS
NC	Ubangian	Yakoma	**lì-kũ̃ũ̃**		Mo
NC	Kwa	Ewe	**kòlí**		Ro
NC	Bantu	Zones C, H etc.	**#-koto, -godo**	knee	BLR3
NC	Bantu	Zones E, F etc.	**-kónò**	forearm, leg	BLR3
AA	Omotic	Wolaytta	**gulba-ta**		LS
AA	Cushitic	*PC	***gulb-/ *gwilb-**		Ehret (1987: 24)
AA	S Cushitic	Asax	**ngulu-et**		MK
AA	Cushitic	Dahalo	**gilli**		LS
AA	East Cushitic	Sidamo	**gulube**		HECD
AA	Agaw	Bilin	**gərəb**		LS

(*continued*)

3. #kulu knee (**Continued**)

Phylum	Family	Language	Attestation	Comment	Source
AA	Semitic	Amharic	**gulbät**		HECD
AA	Chadic	Sukur	**kɪrɪm**		JI
AA	Chadic	Tera	**xulukti**		JI
KS	Southern	!Xóõ	**g‖xúũ**		Traill (1994)
KS	Central	Kxoe- ⸢Anda	**kúdù**		Vo97
KS	Central	Shua-Cara	**(kú)kúdù**		Vo97
KS	Northern	Ju‖'hoan	**g!xòà**		Dickens (1994)

Commentary: A preliminary version of this dataset appears in Blench (1997). Gregersen (1972) treats these as two distinct sets for "leg" and "knee" but they are probably to be put together and the more doubtful cognates discarded. There are two potential cognates, in Bantu, the more widespread #-**kónò** which generally means "forearm" and the less common #–**koto/godo** which is knee directly. It is more probable that the rare forms are genuine cognates and #-**kónò** just a chance resemblance. Bender (1997: 133) pursues linkages that include a purported PNC root *$\mathbf{k^hon}$ for "knee" and brings in Mende **kon** "head" because the "knee as head of the leg". This analysis is not used here. A rather different form, ***BU(N)KA** is proposed as a "world etymology" in Bengtson & Ruhlen (1994).

Refs: B81: 261, Gr.:82,84, G.:101,123, M.:II:223, B:134

4. #kuru Tortoise, turtle

Phylum	Family	Language		Attestation			Gloss	Source
—	Sandawe	Sandawe		**k^hú**	**rú**		tortoise	Sands p.c.
—	Hadza	Hadza		**k'õ**	**ló**		tortoise	Sands p.c.
—	Hadza	Hadza		**k'úú**		**tá-**	turtle	Sands p.c.
—	Laal	Laal		**kú**		**nán**	*petite tortue*	Boyeldieu ms.
NS	Kuliak	Ik	**ro**	**ki**	**ro**	**k**	tortoise	Heine (1999)
NS	Koman	Kwama		**k'**	**u**	**kiʃ**	turtle	RCS
NS	Songhay	Songhay	**ń**	**kúú**	**r**	**á**	small tortoise	BWK
NS	Saharan	Kanuri		**kó**	**ro**	**wú**	tortoise	Cy
NS	Maba	Maba	**fa**	**k**	**ruu**	**n**	tortoise	Ed
NS	Surmic	Didinga	**bo-**	**ko**	**l**		tortoise	RCS
NS	ES	Dinka	**le-**	**ku**	**r**		tortoise	RCS
NS	CS	Asua	**ùn**	**gú**	**lú**		tortoise	Demolin (p.c.)
NS	CS	Maʔdi	**o**	**kù**			tortoise	RCS
NS	Kadu	Krongo		**-kó**	**ò**	**ŋ (ní-)**	tortoise	Reh (1985)
NC	Dogon	Tebul Ure	**aŋguŋ**	**gu**	**ru**		tortoise	RMB
NC	Ijoid	Furupagha	**ɔbʊ**	**kɔ**	**rɔ**		turtle	KW
NC	Mande	Yaure		**kú**	**lú**		tortoise	ALMCI
NC	Atlantic	Serer		**xɔ**		**m**	*tortue*	GS

(*continued*)

4. #kuru Tortoise, turtle (Continued)

Phylum	Family	Language		Attestation			Gloss	Source
NC	Kordofanian	Masakin		**(k)ə**	**rə**		tortoise	RCS
NC	?	Pre		**k**	**ru**	**wɛ**	tortoise	Creissels (p.c.)
NC	Senufo	Nabaj		**xu**	**ru**		tortoise	ALGCI
NC	Ubangian	Geme		**kú**	**lō**		turtle	Mo
NC	Kwa	Mbatto	**ő́mő́**	**k**	**rõ**	**ɛ̃́**	tortoise	ALKCI
NC	Kwa	Ewe		**k**	**lo**		tortoise	Ro
NC	WBC	Ịṣẹ̀kiri	**ólu**	**kú**	**rú**	**mɛ̀**	tortoise	BCCW
NC	EBC	Doka	**a-**	**ku**	**l**		tortoise	BCCW
NC	Bantu	CB	**–**	**kú**	**dù**		tortoise	BLR3
AA	S Cushitic	Gorowa		**ka**	**n**	**keʾ**	tortoise	MK
AA	E Cushitic	Burji		**ko**		**cʾáa**	tortoise, turtle	Sasse (1982)
AA	Beja	Beja	**se**	**ku**	**ur**		tortoise	Hudson (n.d.)
AA	W. Chadic	Hausa	**kùŋ**	**ku**	**ruu**		tortoise	A49
AA	W. Chadic	Mwaghavul		**kú**	**r**		tortoise	JI
AA	C. Chadic	Huba	**kwà**	**kú**	**rù**	**m**	tortoise	Kraft (1981)
AA	Masa	Lame		**gù**	**rè**	**i**	tortoise sp.	Sachnine (1982)
AA	E. Chadic	Toram	**kùn**	**gù**	**rù**		turtle	Jungraithmayr (p.c.)
AA	Berber	Kabyle	**tafe**	**k**	**ru**	**rt**	tortoise	Dallet (1982)
KS	North	Auen		**!gu**	**ru**		tortoise-shell	Bleek (1956)
KS	Central	Naro		**‖go**	**e**		tortoise	Traill (1986)
KS	Central	Mohissa		**cu**	**ru**		tortoise	Bleek (1956)

Commentary: An early version of this table was presented in Blench (1997) where it was argued that the importance of turtles and tortoises in the gathering phase of human history had made this word particularly salient. The diversity of forms attested may reflect the fact that different species may have compound names (see the Kanuri and Aiki forms). Examples of possible extra-African cognates have been recorded; Sora (Muṇḍā, Austroasiatic) **kola** and Tamil (Dravidian) **kuruḷai**, both for tortoise.

Refs: C.:321, Gr.:88, G.:159

5. #kulu "skin, hide"

Phylum	Family	Language	Attestation	Gloss	Source
NS	Kunama	Kunama	**agala**		RCS
NS	ES	Nuer	**kul**		RCS
NS	ES	Murle	**kween**		RCS
NS	Saharan	Teda	**koro-ta**		Le Coeur (1950)

(*continued*)

5. #kulu "skin, hide" (Continued)

Phylum	Family	Language	Attestation	Gloss	Source
NS	Songhay	Songhay	**kúurú**		DC
NC	Dogon	Tebul Ure	**gudugo**		RMB
NC	Ijoid	Defaka	íkpá		KW
NC	Mande	Kpelle	**kɔlɔ**	*peau*	Creissels (1981)
NC	Atlantic	Gola	**ókɔ́lɔ́**	peau	GS
NC	Atlantic	Nalu	**makũ**	PL. **akũ**	GS
NC	Kru	Kuwaa	**kũ̀**		ALKrCI
NC	Gur	Bieri	**kwanu**	*peau*	Ma
NC	Ubangian	Mundu	**kɔ̀nɔ̀**		Mo
NC	Kwa	Siwu	**ɔ-kó**	PL. **sì-kó**	RMB
AA	Omotic	Wolayta	**galba**	skin	LS
AA	Beja	Beja	**kurbe**	skin	LS
AA	Agaw	Kwaräsa	**korbe**	skin/leather	LS
AA	East Cushitic	Ba'iso	**galba**	skin	LS
AA	South Cushitic	Alagwa	**kaari**	hide, skin	MK
AA	Semitic	Amharic	**qurbät**	hide	HECD
AA	E Chadic	Ndam	**gərέ**	skin	JI
AA	C Chadic	Kotoko	**ŋkɛ̀nɛ**	skin	JI
AA	W Chadic	Tala	**kuur**	skin	JI
KS	Southern	!Xóõ	**ǂgúnu**	dried skin	Traill (1994)
KS	Central	Naro	**khò**	*Haut/Fell*	Vo97

Commentary: A preliminary version of this dataset appears in Blench (1997). Greenberg (1963: 21) initially identified this root for Niger-Congo. He later (p. 157) quotes Krongo, but his form does not correspond to that in Reh (1985) which is not evidently cognate. Blench (1997) represents a preliminary compilation of this gloss for Africa. Creissels (1981: 316) points out the Songhay cognate and adds further citations for Niger-Congo. Bender (1997: 129) gives further examples for Nilo-Saharan, although he includes "basket" in his semantic set. Other commentators include "bark", for example Uduk (Eastern Sudanic) **khur** "bark".

Refs: G.:21, Gr.:84, B:129, E:491

6. #mor- "fat, oil, grease

			Attestation			
Phylum	**Group**	**Language**	**I**	**II**	**Gloss**	**Source**
NS	ES	Murle	**morɛ**			RCS
NS	ES	*PN	***mɔ-r**			D

(*continued*)

6. #mor- "fat, oil, grease (Continued)

			Attestation			
Phylum	**Group**	**Language**	**I**	**II**	**Gloss**	**Source**
NS	Maba	Masalit		**ɲuŋgi**	*oil, fat*	Ed
NS	Songhay	Zarma	**máání**		*graisse*	BWK
NC	Dogon	Walo		**nù**	oil	RMB
NC	Dogon	Ana		**nìì**	oil	RMB
NC	Ijo	Furupagha	**imɛɛlɪ**		fat	KW
NC		#PWS	**-mì**			W
NC	Mande	Kpelle		**ŋulo**	oil	VV
NC	Atlantic	Bassari		**nĭl**	*graisse*	GS
NC	Atlantic	Balanta		**-ngelēe**	*graisse*	GS
NC	Kordofanian	Talodi		**ŋ-aag**	(?C)	RCS
NC	Kordofanian	Moro		**ŋela**	grease	RMB
NC	EBC	Eten	**mos**		fat	RMB
AA	Omotic	Zayse	**moora**		fat	LS
AA	Cushitic	Oromo	**moora**		fat	HECD
AA	Semitic	Amharic	**mora**		fat	HECD
AA	Chadic	Tala	**mɔl**		oil	JI
AA	Chadic	Gudu	**mar**		oil	JI
AA	Chadic	Masa	**mùl**		oil	JI
KS	Southern	!Xóõ		**ʼǁnə̀hã**	fat	Traill (1994)
KS	Khoe	PNKhoe		***//˜nui**	*Fett/Öl*	Vo97
KS	Khoe	PWKhoe		***//˜núì**	*Fett/Öl*	Vo97

Commentary: At first sight these might appear to be two separate roots. However, the assumption here is that the m- is originally a class affix, signifying liquids or mass nouns which has become fused to these stem. Forms for mass nouns with m- affixes correspond to Kordofanian ŋ- classes in other branches of Niger-Congo (Schadeberg 1989). The Khoisan forms with their combination of high back and front vowels suggest the reason for their alternation in other language families. The "original' form (if that has a meaning in this context) would then have been something like **#muri** or **#ŋuri**.

Refs: D.:40, W.:257

7. #(dw)isi fire

Phylum	**Group**	**Language**	**Attestation**	**Comment**	**Source**
NS	ES	Meidob	**ussi**		RCS
NS	Kadu	Miri	**issi**		RCS
NS	CS	Miza	**a(t)si**		Bender (1992)

(*continued*)

7. #(dw)isi fire (Continued)

Phylum	Group	Language	Attestation	Comment	Source
NS	CS	Shemyar	**duʃu-n**		Bender (1992)
NS	Saharan	Beria	**ɟíé**		JC
NC	Mande[4]	Beng	**síé**		VV
NC	Kordofanian	Moro	**isia**		RMB
NC	Dakoid	Nnakenyare	**yísi**		RMB
NC	Tivoid	Tiv	**úsu**		Abraham (1940)
AA	Semitic	Akkadian	**iʃaat-**		
AA	South Cushitic	Iraqw	**ʾaɬa**		MK
AA	Chadic	Ron Kulere	**wùʃ**		JI
AA	Chadic	Karekare	**ʔèsî**		Schuh (p.c.)
AA	Chadic	Miya	**osi**		Schuh (p.c.)

Commentary: Not attested in Khoisan, Atlantic. Only recorded in some subgroups of West Chadic and thus probably a local loanword. If the Akkadian form is genuinely related, then it is tempting to assume this an old AA root loaned into NS and thence into eastern NC. Bender (1992: 43) reconstructs Proto-Central Sudanic #co, but as an areal loan, reconstruction is probably not a meaningful exercise.

Refs: Bender (1991: 5)

8. #-si. dog

Phylum	Group	Language	Attestation	Source
NS	Fur	Fur	**asà**	Jakobi (1990)
NS	ES	Proto-Daju	***iise**	RCS
NS	ES	Nara	**wos**	RCS
NS	CS	Baka	**ísì**	RCS
NS	CS	Lugbara	**atsí**	RCS
NS	CS	Lendu	**kazź**	RCS
NS	Koman	Anej	**kas**	RCS
NS	Maba	Masalit	**wasi**	Ed
NS	Kadu	Katcha	**ìs(s)ì**	RCS
NS	Songhay	Kaado	**hánsì**	DC
NC	Atlantic	Manjaku	**û-bûs**	GS
NC	Adamawa	Dza	**iicwá**	UKW
NC	BC	E. Ogbia	**ìsíə̀**	KW
NC	BC	Nupe	**eʃi**	Ban
NC	Bantoid	Ndoro	**síɛ**	RMB
AA	Semitic	Amharic	**wɨʃa**	HECD
AA	Cushitic	PHEC	**waʃa**	HECD

(*continued*)

4. Widespread in South Mande (see cognates in Vydrine ined.)

8. #-si. dog (Continued)

Phylum	Group	Language	Attestation	Source
AA	Cushitic	Beja	**yas**	Hudson ms.
AA	Chadic	Mwaghavul	**as**	JI
AA	Chadic	Kola	**hàzà**	JI

Commentary: Although originally cited by Greenberg (1963: 120), more complete evidence was marshalled by Bender (1981: 258) with attestations in Fur, Sudanic, Kordofanian and possibly Ari [Omotic]. Not attested in Khoesan, Mande and most branches of Afro-Asiatic, which have variants of kVr/n-. This root is extremely widespread in Central Africa and yet does not form a convincing pattern. In Central Sudanic, for example, it is attested in almost every language (Bender 1992: 40, 48). Bender separates the roots with initial b-, but it is likely that the two forms go together.

Refs: G:120

4. Roots with extra-African distributions

Some pan-African roots have possible extra-European cognates, for example "fly" (given in 10.), which may be explained as ideophonic convergence, the sound of beating wings giving rise to the verb. But other lexical items are more difficult to account for in this way, notably "crab". The table shows the African attestations of this word. Because "crab" does not occur in standard wordlists and is associated with wetter regions, it is simply not recorded in many lexical sources.

9. #kala crab

African crabs are highly speciated but can be divided into three categories; marine, freshwater and land crabs. The taxonomy of freshwater crabs is given in Cumberlidge (1999). Material on land crabs is not easily available but they are widespread throughout the continent and are frequently culturally important because of their role in divination systems. Crab divination is reviewed in Blench and Zeitlyn (1989/90) which shows that the words for "spider" and "crab" are etymologically interconnected in the Bantu borderland because of their comparable significance in divination systems. The table presents a sample of crab names. The evidence here is weaker than for other lexical items, largely because words for crab are much more rarely cited. Many of the sources used for the other two tables simply lack the lexeme "crab".

Phylum	Family	Language	Attestation	Source
–	Hadza	Hadza	**goma:**	Sands (p.c.)
NS	C. Sudanic	Mbay	**kə́-bàr**	Keegan (1997)

(*continued*)

Phylum	Family	Language	Attestation	Source
NC	Ijoid	Nembe	**à-kàngà**	Kaliai (1964)
NC		PWS	**#-ka(l)-**	W
NC	Mande	Gban	**klá**	VV
NC	Unclassified	Pre	**kamu**	Creissels (p.c.)
NC	Atlantic	Temne	**a-kara**	W
NC	Gur	Mõõre	**garã́-ga**	Canu (1976)
NC	Kwa	Ewe	**à-gálã̀**	Ro
NC	WBC	Nupe	**kara**	Ban
NC	WBC	Obolo	**úkà**	Co91
NC	Mambiloid	Mambila	**kaab**[21]	PM
AA	W. Chadic	Hausa	**ƙáágwáá**	A49
AA	C. Chadic	Bana	**kwérekwérekiŋ**	Gravina (p.c.)

Commentary: Mukarovsky (1976: 144) adds further Niger-Congo cognates. The Niger-Congo roots are discussed in Williamson & Shimizu (1968: 92).

Refs: M:144; W:230

"Crab" also has widespread Eurasian cognates (Blench 1997). The table below sets out some attestations and reconstructions that have been proposed for "crab" in Old World language phyla.

Phylum	Family	Language	Attestation	Source
Japonic		Modern Japanese	**kani**	
Altaic		Modern Korean	**ke**	
Austroasiatic	Proto-Mon-Khmer		***kə(n)taam**	Diffloth (1994)
	Proto-North Bahnaric		***katam**	Smith (1972)
Austronesian	Proto-Austronesian		***kaRaŋ**	Blust (ined.)
	Proto-Nuclear Micronesian		***karika**	Marck (p.c.)
Andamanese	Great Andaman	Aka Biada	**kátta-da**	Portman (1887: 22)
	Little Andaman	Onge	**tekandue**	Dasgupta & Sharma (1982)
Sino-Tibetan	Tibeto-Burman	Tamang	**khakre**	Rana (2005)
Kusundic		Kusunda	**kakchi**	Rana (2005)
Dravidian	Common Dravidian		**kup(p)i**	Burrow & Emeneau (1984: 158)
Indo-European		Greek	**karkinos**	
Vasconic		Basque	**karramorro**	Trask (p.c.)

Clues to the processes at work in African languages can be gained by comparison with Indo-European associations of words for "crab". Indo-European has a root #*kar-* meaning "hard" which has a complex association with words for "crab".[5] Latin *cancer* and Greek *karkinos* are both derived from reduplications of the original root, the image apparently being the hardness of the crab's shell (Watkins 1982). A similar association also exists in Niger-Congo; Westermann (1927: 240) reconstructs #*kual-* for "to be hard" in Proto-Western Sudanic, and #*-kal-* for "crab".

10. #pur- "to fly, jump"

This root presents an intriguing methodological problem. The citations are so similar that they must be related in some way. Is this simply a case of ideophonic convergence, where similar languages come to similar conclusions about the sound of beating wings? Not all languages do this as many datasets for "fly" are omitted on the grounds of non-cognacy. Similarly, the semantic link with "jump" which seems to be typical for Africa is not generally found elsewhere, where the common association appears to be "flee".

Phylum	Family	Language	Attestation	Gloss	Source
NS	Maba	Mesalit	**fir**		Ed
NS	Berta	Berta	**hɔ'rɔŋ**		Bender (1989)
NS	ES	Gaam	**pərd-**	fly	Bender & Ayre (1980)
NS	ES	*PN	***pär**	fly, jump	D
NS	Songhay	Djenné Chiini	**firri**	fly	Heath (1998)
NS	Saharan	Kanuri	**fàr**	to jump, fly	Cy
NC	Dogon	Bunɔgɛ	**pile**	to fly	RMB
NC	Ijoid	Nkoro	**fĩĩ**	fly	KW
NC		*PWS	***pi, pil-**	to fly, flutter	W
NC	Mande	Bamana	**pā́**	fly, jump	VV
NC	Atlantic	Temne	**fal(ar)**	*voler*	GS
NC	Kordofanian	Moro	**abəro**	to fly	RMB
NC	EBC	Lokə	**fiilɔ́**	to fly	JS
AA	Beja	Beja	**biir**	fly	Hudson (p.c.)
AA	Agaw	Awngi	**pərr-**	jump	Applewyard (p.c.)
AA	Proto-East-Cushitic		***bar(ar)**	fly	Sasse (1982)
AA	East Cushitic	Burji	**burr-**	fly	HECD

(*continued*)

5. I am grateful to Václav Blažek for this observation.

Phylum	Family	Language	Attestation	Gloss	Source
AA	South Cushitic	Gorowa	**furuu'**	fly	MK
AA	Chadic	Hausa	**ɓùráá**	jump	JI
AA	Chadic	Yedina	**fə́r**	jump	JI
AA	Chadic	Kwang	**bɔ́ré**	jump	JI
AA	Berber	Kabyle	**fferfer**	*voler*	Dallet (1982)
KS	Khoe	PNKhoe	***pe**	*springen*	Vo97
KS	Khoe	PEKhoe	***pe**	*springen*	Vo97

Commentary: The meanings of "fly" and "jump" are regularly intertwined throughout African language phyla. Bender (1997: 121) lists more Nilo-Saharan cognates in elliptical style.

Refs: Gr.:83, D.:42, W.:275, B: 121, E: 382, Ehret (1987: 26)

Swadesh (1971) derived a similar form ideophonically as a world gloss, although he spreads the semantic net wider than is included here. This etymology is reprised in Bengtson & Ruhlen (1994) with a proposed proto-form ***par**. Some of the more clearly cognate forms drawn from these sources are;

Phylum	Family	Language	Attestation	Gloss
Kartvelian		Georgian	**p'er**	
North Caucasian		Abkhaz	**pir**	
Indo-European	Iranian	Persian	**parr–**	
Indo-European		English	**fly**	
Uralic	Ugric	Khanty	**por**	
Dravidian		Tamil	**paRa**	fly, hover, flutter
Isolate		Gilyak	**parpar**	fly, hover about
Austroasiatic	Mon-Khmer	Khmer	**par**	
Sino-Tibetan	Sinitic	Chinese	**fei**	

Conclusions

The tables given above provide *a priori* evidence for a series of Pan-African roots that occur across the language phyla of Africa in forms sufficiently similar as to exclude their use as evidence for genetic affiliation. There is probably no one explanation for these forms, but it is striking that so many confirm the general formula k + back vowel + r/l. "Turtle", "skin" and "knee" have no obvious link and are only joined by the phonaesthetic qualities of their canonic form. Even "cut" and "crab" have kVr- structures although the vowels are more diverse. Exactly how this works is unclear, but it seems that some canonic forms prevent words from undergoing

the usual diversification typical of language genesis. Some of these are apparently confined to Africa, while others are more global. Another canonic form, not collated here, is *#koro* for "round, wheel, circular', which has cognates across Eurasia, the Pacific and the New World as well as throughout Africa.

The other transphylic lexemes have a narrower distribution, and cultural salience may be at work. "Fat" may not be culturally salient today, but in an era when hunting and gathering was the basis of subsistence, both oils and animal fats were highly valued and terms for them would thus become deeply embedded. Similarly, "fire", essential to hunting and cooking, would have the same deep significance as "water", which appears to be marked throughout Africa and Eurasia by a bilabial nasal. The case of "dog", may be somewhat different, as the dog is an ancient but datable introduction into Africa, hence these wandering words (Blench 2000). Dogs appear in rock-paintings at a defined point and gradually appear in excavations over the last few thousand years. The forms of the *#iʃi* frame are more obviously coherent geographically, and may well reflect the diffusion of the dog across Central Africa.

The existence of widespread roots that prove to be scattered across African language phyla but that have been used for the identification and classification of language families, should make us wary. Although more obvious ideophones are easily discarded, it is clear that many pan-African roots cannot be predicted on this basis. Their deep embedding may be related to cultural salience in prehistory, or simply transferred homophony; frames that mark key terms are retained in homophonous but less marked concepts. This points to an important research lacuna in our understanding of phonaesthemes and their broader significance for historical linguistics.

Acronyms and terminology

I have adopted the convention for reconstructions used in the Niger-Congo volume edited by Bendor-Samuel (1989), distinguishing those established by regular sound-correspondences from those derived by quick inspection of cognates. By this criterion, most major reconstructions are "quasi-reconstructions" (inevitably). The effect of this is to translate the starred forms of various writers to hache "#".

* Reconstruction established from complete analysis of sound-change

\# "Pseudo-reconstruction" established from quick inspection of cognates

Acronym	Expansion or source	Reference or language treated
A49	Abrahams (1949)	Hausa
AA	Afroasiatic	
ALGCI	Mensah & Tchagbale (1983)	Gur
ALKCI	Hérault (1983)	Kwa
ALKrCI	Marchese (1983)	Kru

Acronym	Expansion or source	Reference or language treated
ALMCI	Halaoui, Tera and Trabi (1983)	Mande
B	Bender (1997)	Nilo-Saharan
B79	Bender (1979)	Gumuz
Ba	Bailleul (1996)	Bambara
Ban	Banfield (1914)	Nupe
BC	Benue-Congo	
BCCW	Williamson & Shimizu (1968) & Williamson (1973)	Benue-Congo
BWK	Bernard & White-Kaba (1994)	Zarma
C	Consonant	
CB	Common Bantu	Guthrie 1967–71
Chr	Christaller (1933)	Twi
Co91	Bruce Connell (p.c.)	Lower Cross
Cy	Cyffer (1994)	Kanuri
D	Dimmendaal (1988)	Proto-Nilotic
DC	Ducroz & Charles (1978)	Songhay Kaado
Demolin	Didier Demolin (p.c.)	Central Sudanic
E	Ehret (2001)	Nilo-Saharan
Ed	Edgar (1991)	Maba group
ES	Eastern Sudanic	
G	Greenberg (1963)	African languages
Gr	Gregersen (1972)	Kongo-Saharan
GS	Guillaume Segerer (p.c.)	Atlantic
Gt	Guthrie (1967–1971)	Bantu
HECD	Hudson (1989)	Highland East Cushitic
JI	Jungraithmayr & Ibriszimow (1995)	Chadic
JC	Jakobi & Crass	Zaghawa
JS	Jan Sterk (p.c.)	Upper Cross
KW	Kay Williamson (†) (p.c.)	Ijoid, Cross River
LS	Lamberti & Sottile (1997)	Cushitic and Omotic
M	Mukarovsky (1976/7)	Proto-Western Nigritic [= Proto-Atlantic-Congo]
Ma	Manessy (1975)	Oti-Volta
MK	Mous & Kießling (2004)	Southern Cushitic
Mo	Moñino (1988)	Ubangian
N	Nasal	
NC	Niger-Congo	
NS	Nilo-Saharan	
P	Prost (1953)	South Mande
PB	Proto-Bantu	
PD	Proto-Daju	Thelwall 1989

(*continued*)

(Continued)

Acronym	Expansion or source	Reference or language treated
PI	Proto-Ịjọ	Williamson, in prep.
PM	Perrin & Mouh (1995)	Mambila
PN	Proto-Nilotic	
PNC	Proto-Niger-Congo	
PNS	Proto-Nilo-Saharan	
PWN	Proto-Western Nigritic	Mukarovsky 1976/77
PWS	Proto-West Sudanic	Westermann 1927
RCS	Roland Stevenson mss.	Nilo-Saharan, Kordofanian
RMB	Author's fieldwork	
Ro	Rongier (1995)	Ewe
Sands	Bonny Sands (p.c.)	Hadza, Sandawe
Sch94	Schadeberg (1994)	Kadu
V	Vowel	
Vo82	Voßen (1982)	Eastern Nilotic
Vo88	Voßen (1988)	Maa
Vo97	Voßen (1997)	Khoisan
VV	Valentin Vydrine (p.c.)	Mande
W	Westermann (1927)	Western Sudanic

References

Abraham, R.C. 1940. *A dictionary of the Tiv language.* London: Crown Agents.

Abraham, R.C. 1949. *Dictionary of the Hausa language.* London: University of London Press.

Bailleul, Charles 1996. *Dictionnaire Bambara-Français*. Bamako: Donniya.

Banfield, A.W. 1914. *Dictionary of the Nupe Language*. Shonga: The Niger Press.

Barreteau, D. & Le Bléis, Y. 1990. *Lexique Mafa.* Paris: Geuthner.

Beaton, A.C. 1968. *A grammar of the Fur language*. Linguistic Monograph Series, 1. Khartoum: Khartoum University.

Bender, M.L. 1979. Gumuz: a sketch of grammar and lexicon. *Afrika und Übersee*, LXII: 38–69.

Bender, M.L. 1983. Proto-Koman Phonology and Lexicon. *Afrika und Übersee*, LXVI: 259–297.

Bender, M.L. 1989. Berta lexicon. Topics in Nilo-Saharan Linguistics. Ed. by M.L. Bender, 271–304. Hamburg: Helmut Buske Verlag.

Bender, M.L. 1991. Genetic Classification of Nilo-Saharan. Paper for the Table Ronde du réseau 'Diffusion lexicale en zone Sahélo-Saharienne'. Nice.

Bender, M.L. 1992. Central Sudanic Segmental and Lexical Reconstructions. Afrikanische Arbeitspapiere, 29: 5–61.

Bender, M.L. 1997. [2nd Edn.] *The Nilo-Saharan languages: a comparative essay.* Munich: Lincom Europa.

Bender, M.L. 2003. *Omotic lexicon and morphology.* Carbondale: Lionel Bender.

Bender, M.L. 2005. *The East Sudanic languages: lexicon and phonology.* Carbondale: Lionel Bender.

Bender, M.L. ms. *Kunama lexicon*. Computer printout.
Bender, M.L. & Ayre, M.A. 1980. *Preliminary Gaam-English-Gaam dictionary*. Available from the author.
Bengtson, John D. & Merritt Ruhlen. 1994. Global etymologies. *On the origin of languages: studies in linguistic taxonomy*. Merritt Ruhlen, 277–336. Stanford: Stanford University Press.
Bernard, Yves & Mary White-Kaba. 1994. *Dictionnaire Zarma-Français*. Paris: ACCT.
Bleek, Dorothea. F. 1956. *A Bushman Dictionary*. New Haven: American Oriental Society.
Blench, Roger M. 1995. Is Niger-Congo simply a branch of Nilo-Saharan? *Proceedings of the Fifth Nilo-Saharan Linguistics Colloquium, Nice, 1992,* Ed. by R. Nicolai & F. Rottland, 68–118. Köln: Rudiger Köppe.
Blench, Roger M. 1997. Crabs, turtles and frogs: linguistic keys to early African subsistence systems. *Archaeology and Language, I,* Ed. by R.M. Blench & M. Spriggs, 166–183. London: Routledge.
Blench, Roger M. 2007. Niger-Saharan III. *Proceedings of the VIIIth Nilo-Saharan Conference in Hamburg 2–6th September, 2001*, Ed. by D. Payne & M. Reh. Koln: Rudiger Köppe.
Blench, Roger M. 2000. Minor Livestock species in Africa. *The origin and development of Africa liverstock*, Ed. by R.M. Blench & K.C. MacDonald, 314–338. London: University College.
Blench, Roger M. in press. Furthr evidence for Niger-Saharan and the problem of Pan-African roots. Proceeding of the VIIth Nilo-Saharan Conference in Vienna, 2–6th September, 1998, Ed. by N. Cyffer. Köln: Rudiger Köppe.
Blench, R.M. & Zeitlyn, D. 1989/90. A web of words. *Sprache und Geschichte in Afrika* 10/11: 171–186.
Boyeldieu, P. ms. n.d. *Presentation du láà:l ou "Gori" (Moyen-Chari, Tchad).* Paris: CNRS.
Burrow, T. & Emeneau, Murray B. 1984. [2nd Edn.] *A Dravidian etymological dictionary*. Oxford: Clarendon Press.
Canu, Gaston. 1976. *La langue Mò:rē*. Paris: SELAF.
Christaller, J.G. 1933. *Dictionary of the Asante and Fante language called Twi*. Basel: Basel Evangelical Missionary Society.
Creissels, Denis. 1981. Songhay et Niger-Congo (mandé). *Nilo-Saharan,* Ed. by T.C. Schadeberg & M.L. Bender, 307–327. Holland: Foris Publications.
Creissels, Denis. ined., *Liste de mots Pre.* ms.
Cust, R.N. 1883. *The modern languages of Africa*. 2 vols. London: Richard Bentley.
Cumberlidge, Neil. 1999. *The freshwater crabs of West Africa*. Paris: IRD.
Cyffer, Norbert. 1994. *English-Kanuri Dictionary*. Köln: Rudiger Köppe.
Dallet, J-M. 1982. *Dictionnaire Kabyle-Français*. SELAF, Paris.
Dasgupta, D. & Sharma, S.R. 1982. *A handbook of the Onge language.* Anthropological Survey of India, Linguistic Series, 5. Calcutta.
Dempwolff, Otto. 1938. *Vergleichende Lautlehre des austronesischen Wortschatzes, vol. 3: Austronesisches Wörterverzeichnis*. Zeitschrift für Eingeborenen-Sprachen, Supplement 19. Berlin: Reimer.
Dickens, Patrick. 1994. *English- Ju|'hoan, Ju|'hoan-English dictionary*. QKF8. Köln: Rüdiger Köppe.
Diffloth, Gerald. 1994. The lexical evidence for Austric, so far. *Oceanic Linguistics* 33: 2. 309–321.
Dimmendaal, Gerrit J. 1988. The lexical reconstruction of Proto-Nilotic: A first reconnaissance. *Afrikanische Arbeitspapiere* 16. 5–67.
Ducroz, Jean-Marie & Marie-Claire Charles. 1978. *Lexique Soŋey (Songay) Français*. Paris: Harmattan.

Edgar, John T. 1991. *Maba-group lexicon*. Berlin: Reimer.
Ehret, Christopher. 1987. Proto-Cushitic reconstruction. *Sprache und Geschichte in Afrika* 8. 7–180.
Ehret, Christopher. 2001. *A historical-comparative reconstruction of Nilo-Saharan.* Köln: Rudiger Köppe.
Greenberg, Joseph H. 1963. *The Languages of Africa*. Published for Indiana University by Mouton, the Hague.
Gregersen, Edgar A. 1972. Kongo-Saharan. *Journal of African Languages* 11: 1. 69–89.
Guthrie, Malcolm. 1967–71. *Comparative Bantu*. 4 vols. Farnham: Gregg International Publishers.
Halaoui, Nazam, Kalilou Tera & Monique Trabi. 1983. *Atlas des languages mandé-sud de Côte d'Ivoire*. Abidjan: ACCT and ILA.
Heath, J. 1998. *Dictionnaire Songhay-Anglais-Français*. Paris: l'Harmattan.
Heine, Bernd. 1999. *Ik dictionary.* Köln: Rüdiger Köppe.
Hérault, G. 1983. *Atlas des Langues Kwa de Côte d'Ivoire, Tôme 2*. Abidjan: ACCT and ILA.
Hudson, Grover. 1989. *Highland East Cushitic dictionary*. Hamburg: Helmut Buske.
Hudson, R. n.d. *Beja dictionary*. ms.
Jakobi, Angelika. 1990. *A Fur grammar*. Hamburg: Helmut Buske Verlag.
Jakobi, Angelika & Joachim Crass. 2004. *Grammaire du beria (langue saharienne)*. Köln: Rüdiger Köppe.
Jungraithmayr, H. & Dymitr Ibriszimow. 1995. *Chadic lexical roots.* [2 vols.] Berlin: Reimer.
Kaliai, M.H.I. 1964. *Nembe-English dictionary*. Ibadan: Institute of African Studies.
Keegan, John M. 1997. *Dictionary of Mbay*. München: Lincom Europa.
Kraft, C. 1981. *Chadic wordlists*. (3 vols.) Berlin: Reimer.
Lamberti, M. & Sottile, R. 1997. *The Wolaytta language.* Köln: Rüdiger Köppe.
Le Coeur, Charles. 1950. *Dictionnaire Ethnographique Teda.* Dakar: IFAN.
Manessy, Gabriel. 1975. *Les langues Oti-Volta*. Paris: SELAF.
Marchese, Lynell. 1983. *Atlas Linguistique Kru*. Abidjan: ACCT and ILA.
Mensah, E.N.A. & Tchagbale, Z. 1983. *Atlas des languages gur de Côte d'Ivoire*. Abidjan: ACCT and ILA.
Moñino, Yves, Ed. 1988. *Lexique comparatif des langues oubangiennes*. Paris: Geuthner.
Mous, M. & Kießling, R. 2004. *Reconstruction of proto-West Rift.* Köln: Rüdiger Köppe.
Mukarovsky, Hans G. 1976–1977. *A study of Western Nigritic*. 2 vols. Wien: Institut für Ägyptologie und Afrikanistik, Universität Wien.
Perrin, M.J. & Mouh, M. 1995. *Lexique mambila-français*. Yaoundé: SIL, Cameroun.
Portman, M.V. 1887. *Manual of the Andamanese languages.* Government Printer, Port Blair.
Prost, R.P.A. 1953. *Les language Mandé-Sud du groupe Mana-Busa*. Dakar: IFAN.
Rana, B.K. 2005. New materials on the Kusunda language. *Mother Tongue*, X: 109–122.
Reh, Mechthild. 1985. *Die Krongo-Sprache (nìino mó-dì).* Berlin: Reimer.
Rongier, Jacques. 1995. *Dictionnaire français-éwé*. Paris: ACCT-Karthala.
Sasse, Hans-Jurgen. 1982. *An etymological dictionary of Burji*. Hamburg: Buske.
Schadeberg, Thilo C. 1981a. The classification of the Kadugli language group. *Nilo-Saharan,* Ed. by T.C. Schadeberg & M.L. Bender, 291–306. Holland: Foris Publications.
Schadeberg, Thilo C. 1981b. *A Survey of Kordofanian. Volume 1: The Heiban Group*. Hamburg: Helmut Buske.
Schadeberg, Thilo C. 1994. Comparative Kadu wordlists. *Afrikanische Arbeitspapiere* 40. 11–48.
Schadeberg, Thilo C. & Marvin L. Bender, eds. 1981. *Nilo-Saharan*. Holland: Foris Publications.

Smith, Kenneth D. 1972. *A phonological reconstruction of proto-North-Bahnaric.* Asian-Pacific Series, No. 2. SIL, Santa Ana, California.

Swadesh, Morris. 1971. *The Origin and Diversification of Language.* Ed. by Joel Sherzer. Chicago, IL: Aldine/Atherton.

Teferra, Anbessa. ined. *Shabo wordlist*. Typescript.

Thelwall, Robin. 1981. *Subgrouping and Reconstruction of Daju*. Schadeberg & Bender 1981. 167–184.

Tosco, M. 1991. *A Grammatical Sketch of Dahalo*. Hamburg: Buske.

Traill, A. 1986. Do the Khoi have a place in the San? New data on Khoisan linguistic relationships. *Sprache und Geschichte in Afrika* 7: 1. 407–430.

Traill, A. 1994. *A !Xóõ dictionary*. QKF9. Köln: Rüdiger Köppe.

Voßen, Rainer. 1982. *The Eastern Nilotes: Linguistic and Historical Reconstructions.* Kölner Beitrage zur Afrikanistik, Volume 9. Berlin: Reimer.

Voßen, R. 1988. Khoe linguistic relationships reconsidered: the data. *New Perspectives on the study of Khoisan,* Ed. by R. Voßen, 67–108. Hamburg: Buske Verlag.

Voßen, Rainer. 1997. *Die Khoe-Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas.* Köln: Rüdiger Köppe.

Westermann, Diedrich. 1927. *Die Westlichen Sudansprachen und ihre Beziehungen zum Bantu.* Berlin: De Gruyter.

Williamson, Kay, Ed. 1972. *Igbo-English dictionary: Based on the Onitsha dialect.* Benin: Ethiope Publishing Corporation.

Williamson, Kay. 1973. *Benue-Congo comparative wordlist: Vol. 2.* Ibadan: West African Linguistic Society.

Williamson, Kay. in prep. *Comparative Ịjọ.* Electronic ms. left by the late Professor Williamson.

Williamson, Kay. & Kiyoshi Shimizu. 1968. *Benue-Congo comparative wordlist: Vol. 1*. Ibadan: West African Linguistic Society.

PART IV

Languages of Eurasia, Oceania, and the Americas

Some thoughts on the Proto-Indo-European cardinal numbers

Allan R. Bomhard
Charleston, SC, USA

This paper takes a fresh look at the reconstruction of the Proto-Indo-European cardinal numbers. Evidence is presented that the earliest form of the number "two" was (in traditional reconstruction) **do-* and that the form usually given in the standard handbooks, namely, **duwō*/**dwō*, was a later borrowing, probably from Northwest Caucasian. Moreover, the earliest form of the number "four" was **Hok*^{h}*-t*^{h}*o-*. It is preserved in the number "eight", **Hok*^{h}*-t*^{h}*oH(w)*, which is a later dual form meaning "two fours". The forms **meyu-* "four", found in the Anatolian branch, and **k*^{wh}*ét*^{h}*-wor-* "four", found in the non-Anatolian daughter languages, are both later innovations. Finally, the number "ten", *de-km̥t-*, is taken to be an old compound meaning "two hands", as often suggested in the literature.

Though there are problems with the reconstruction of a common form for the number "one" (see below), the following cardinal numbers "one" to "ten" are traditionally reconstructed for later Proto-Indo-European (for additional information, cf. Blažek 1999: 141–324; Meillet 1964: 409–413; Beekes 1995: 212–213; Szemerényi 1960; Gamkrelidze – Ivanov 1995. I: 740–744; Sihler 1995: 404–433; Adrados 1975. II: 871–877; Adrados – Bernabé–Mendoza 1995–1998. III: 127–131):

The numbers in Anatolian are, for the most part, not known inasmuch as they are written ideographically (cf. Luraghi 1997: 27). The number "seven" occurs in Hittite in the ordinal (DAT.) *ši-ip-ta-mi-ya* "seventh" (cf. Sanskrit *saptamá-ḥ* "seventh"; Latin *septimus* "seventh") (cf. Sturtevant 1951: 30, 44, 60, 63, 77, and 87; Kronasser 1956: 152; Benveniste 1962: 83). The number "three" is also represented in Hittite in (adv.) *te-ri-ya-an-na* "for the third time", and the military title *te-ri-ya-al-la*, *tar-ri-ya-na-al-li* "third-in-command, officer of the third rank" (cf. Kronasser 1956: 151; Benveniste 1962: 82; Blažek 1999: 186–187), apparently to be read **tri-* "three" (cf. Benveniste 1962: 86), while "two" is found in Hittite in the military title *du-ya-na-al-li* "second-in-command, officer of the second rank", the compound *ta-a-i-ú-ga-aš*, *da-a-i-ú-ga-aš*, *ta-a-ú-ga-aš* "two years old" (*da-*/*ta-* "two" + *i-ú-ga-aš* "yearling"), *da-a-an*, *ta-a-an* "a second time; second", and (NOM. SG. C.) *da-ma-a-(i-)iš* "second, other" (cf. Sturtevant 1951: 34, 58, 61, 67, and 110; Kronasser 1956: 151; Benveniste 1962: 81), and in Hi-

	Brugmann (1904: 363–365)	Szemerényi (1996: 222–224)	Meier-Brügger (2003: 233–234)	Fortson (2004: 131)
1.	*oi-no-s	*oinos	(*Hoi̯-)	*oi-no-
	*oiu̯o-	*oi-u̯o-	*oikos	*oi-ko-
	*sem-	*sem-	*sem-	*sem-
2.	*d(u)u̯ō(u)	*duwō/*dwō	*d(u)u̯o-	*d(u)u̯oh$_1$
3.	*trei̯-, *tri-	*treyes	*tréi̯-es	*tréi̯es
4.	*qᵘ̯etu̯or-	*kʷetwores	*kʷétu̯or-	*kʷétu̯ores
5.	*peŋqᵘ̯e	*penkʷe	*penkʷe	*pénkʷe
6.	*s(u̯)ek̂s	*s(w)eks	*s(u)u̯éḱs	*su̯ék̂s
7.	*septm̥	*septm̥	*septḿ	*septḿ
8.	*ok̂tō(u)	*oktō	*otḱ-	*ok̂tō(u)
9.	*neu̯n̥ *enu̯n̥	*newn̥	*h$_1$néu̯n̥	*neu̯n̥
10.	*dek̂m̥	*dekm̥t/*dekm̥	*déḱm̥	*dek̂m̥

eroglyphic Luwian *tu-wa/i-zi* "two" (cf. Laroche 1960: 206; Meriggi 1962: 136; Blažek 1999: 164). All three of these forms agree with what is found in the non-Anatolian Indo-European daughter languages. The forms in the Anatolian languages for the number "four", however, differ from those that are found elsewhere: Proto-Anatolian **meyu-* "four" > Hittite (NOM. PL.) *mi-e-(ya-)wa-aš*, (ACC. PL.) *mi-e-ú-uš*, (GEN. PL.) *mi-i-ú-wa[-aš]* "four", Luwian *mauwa-* "four" (instr. PL. *ma-a-u-wa-a-ti*) (cf. Benveniste 1962: 81; Laroche 1959: 70; Blažek 1999: 201–202).

Two basic stems may be reconstructed for the number "one": **Hoy-* and **sem-*(cf. Sihler 1995: 404–407; Fortson 2004: 131). The underlying meaning of the first stem appears to have been "single, alone", while that of the second stem appears to have been "together (with)" (cf. Szemerényi 1996: 222; Blažek 1999: 155). The first stem only occurs with various suffixes: (1) **Hoy-no-*(cf. Latin *ūnus* "one" [Old Latin *oinos*]; Old Irish *óen*, *óin* "one"; Gothic *ains* "one"; Old English *ān* "one"; Old High German *ein* "one"; Old Church Slavic *inъ* "one" – it is also found in Greek οἴνη, οἰνός "roll of one [in dice]"); (2) **Hoy-wo-*(cf. Avestan *aēva-* "one"; Old Persian *aiva-* "one" – it is also found in Greek οἶος "alone, lone, lonely" [Cyprian οἶϜος]); (3) **Hoy-kʷʰo-* or **Hoy-kʰo-*(cf. Sanskrit *éka-ḥ* "one"; Mitanni ["Proto-Indic"] *aika-* "one"). The second stem is found in Greek: Attic (NOM. SG. m.) εἷς "one", Doric ἧς "one, Cretan ἔνς (< *ἕνς < *ἕμς < **sems*) "one"; Attic (f.) μία (< *σμ-ια) "one". It is also found in Armenian *mi* "one". To complicate matters, the various forms of the ordinal found in the daughter languages are based upon yet another Proto-Indo-European stem: **pʰer(H)-/*pʰr̥(H)-* "first" (> **pʰr̥H-wo-*, **pʰr̥H-mo-*, **pʰrey-mo-*, **pʰrey-wo-*, **pʰroH-tʰo-*, **pʰroH-mo-*, etc. [for details, cf. Blažek 1999: 141–162; see also Szemerényi 1996: 228; Sihler 1995: 427–428]).

There was a variant form **t'w-i-*(traditional reconstruction **dw-i-*) "two" in Proto-Indo-European that was used in compounds (cf. Gamkrelidze – Ivanov 1995. I: 742) and in the adverbial form **t'w-i-s* "twice" (cf. Latin *bis* "twice" [Old Latin *duis*]; Sanskrit

d(u)ví "twice"; Avestan *biš* "twice"; Greek δίς "twice"; Middle High German *zwir* "twice"). The regular form for the number "two" is traditionally reconstructed as a dual **duwō/*dwō* (Szemerényi's reconstruction), though the dual forms may have arisen in the early prehistory of the individual daughter languages themselves (cf. Sihler 1995: 408). This view is quite attractive, and I would reconstruct **t'(u)w-o-* as a plural (originally indeclinable) and not as a dual at the Proto-Indo-European level (the plural is still found, for example, in forms such as Greek [NOM. PL.] δύο, [NOM.-ACC. PL.] δυοῖν). Attempts to come up with an etymology within Indo-European itself for this number have met with little success (cf. Blažek 1999: 175–179). That the core form was **t'(u)w-*(cf. Blažek 1999: 178; Villar 1991: 136–154; Ernout – Meillet 1979: 187–188) is shown by the fact that the thematic vowel **-o-* could be added directly to the that form, on the one hand, to yield the form traditionally reconstructed for the independent word for the number "two", while, when used in compounds or to express "twice", the extension **-i-* could be added directly to the core form instead. Thus, we get **t'(u)w-o- ~ *t'(u)w-i-* "two".

There are several forms in Hittite that point to an alternative form for "two" in Proto-Indo-European – these are: the compound *ta-a-i-ú-ga-aš*, *da-a-i-ú-ga-aš*, *ta-a-ú-ga-aš* "two years old" (*da-/ta-* "two" + *i-ú-ga-aš* "yearling"), *da-a-an*, *ta-a-an* "a second time; second", and (NOM. SG. c.) *da-ma-a-(i-)iš* "second, other". These forms point to a Proto-Indo-European **tʾe-/*tʾo-*(earlier **tʾe-/*tʾa-*) "two" (cf. Sturtevant 1951: 61 [Sturtevant reconstructs Proto-Indo-Hittite **do-* "two"]; Benveniste 1962: 78–86 [Benveniste brings in data from non-Anatolian Indo-European daughter languages to support his views]). There is absolutely no way to reconcile **tʾe-/*tʾo-* with **t'(u)w-o/i-* phonologically so that they can be convincingly combined in a single reconstruction (Adrados – Bernabé – Mendoza 1995–1998. III: 138 note the problems involved and discuss proposed solutions). Consequently, two competing forms must be reconstructed for the number "two" in Proto-Indo-European. If the Proto-Indo-European number "ten" were originally a compound meaning "two hands", that is, **tʾe-* "two" + **k^hṃ(t^h)-* "hand", as some have suggested (cf. Szemerényi 1960: 69 and 1996: 224, fn. 16; Markey 1984: 284–285; Justus 1988: 533; Gamkrelidze – Ivanov 1995. I: 747; Adrados – Bernabé – Mendoza 1995–1998. III: 131; but rejected by Blažek 1999: 295–296), it would provide additional evidence for reconstructing two separate forms for the number "two".

This situation raises the question as to why there should be two alternative forms for the number "two" in Proto-Indo-European. A possible answer is that **tʾe-/*tʾo-* may have been the native form, while **t'(u)w-o/i-* may have been a borrowing. Given the geographical location of the Indo-European homeland in the vicinity of the Black Sea near speakers of early Northwest Caucasian languages, these languages might have been a possible source for the **t'(u)w-o/i-* form. Indeed, there is a striking resemblance between Proto-Indo-European **t'(u)w-o/i-* "two" and similar forms for this number in Northwest Caucasian: Proto-Northwest Caucasian **tʾqʾo-* "two" > Proto-Circassian **tʾʔʷə* "two", Proto-Ubykh **tʾqʾʷə* (>**tʾqʾʷa*) "twice", Proto-Abkhaz-Abaza **tʾʕʷə* "two"

(cf. Colarusso 1992: 45). Kuipers (1975: 19) reconstructs Proto-Circassian **Tqʾ°(a)* "two" (> Bžedux *tʾ°(a)/tʾ(a)w*, *-tʾ(a)* "two [twice]"; Kabardian *-tʾa* only in *məzamətʾa* "more than once, repeatedly", literally, "not-once-not-twice"). Colarusso (1992: 45) derives the Proto-Indo-European form for the number "two" from **tʾʔʷə*, which he claims first became **tʾəʔʷ* and then **tʾ(u)w-o-* [traditional **d(u)w-o-*]. Colarusso (1992) documents many other similarities between Proto-Indo-European and Northwest Caucasian. These similarities lead Colarusso to think about possible genetic relationship. I prefer to see the similarities to be due to the fact that the Indo-Europeans occupied territory north of and between the Black and Caspian Seas that was originally inhabited by speakers of early Northwest Caucasian languages. We can further speculate that **tʾ(u)w-o/i-* "two" eventually replaced the native Proto-Indo-European word for "two", which survived only in relic forms and in the word for the number "ten" (**tʾe-kʰm̥(tʰ)*).

The Proto-Indo-European word for the number "three" is completely straightforward and can be reconstructed **tʰr-ey-/*tʰr-i-*. Sanskrit (NOM.-ACC.) *tisrá* and related forms in other Indo-European daughter languages are dissimilated from **tʰri-sr-* (cf. Sihler 1995: 410).

The word for the number "four" is traditionally reconstructed **kʷetwores* (so Szemerényi; Brugmann reconstructs **qᵘetu̯or-*). The most convincing etymology is that offered by Burrow (1973: 259) (see also Beekes 1987: 219):

> 4.This number is formed on the basis of a root *kwet* which seems originally to have meant something like "angle" (cf. Lat. *triquetrus* "triangular"), whence "square" and from that "four". In the masc. and neut. (*catvāras*, *catvvāri*, Lat. *quattuor*, etc.) the stem is formed by means of the suffix *-var*, with adjectival accent and vṛddhi in the nominative. In the other cases (ACC. *catúras*, etc.) the suffix has the weak form according to the general rule. A neuter noun **cátvar*, or its IE prototype, is presupposed by the thematic extension *catvara-* "square, crossroads". Elsewhere the simple *r*-suffix may appear (Gk. Dor. τέτορες, Lat. *quarter*), or the elements of the suffix may be reversed (Av. *čaθru-*).

In accordance with Burrow's views, the form **kʷʰetʰ-wor-* "square" may be reconstructed for later Proto-Indo-European. It was preserved in Sanskrit in the thematic derivative *catvará-m* "quadrangular place, square, crossroads" (cf. Mayrhofer 1956–1980. I: 371). It was this form that served as the basis for the number "four" found in the non-Anatolian daughter languages: (NOM. PL.) **kʷʰetʰ-wṓr*. Curiously, the suffix **-wor-* is replaced by **-sor-* in the feminine (cf. Sanskrit *cátasraḥ*). Thus, the root was **kʷʰetʰ-*, to which different suffixes could be added. It is intriguing to speculate that **kʷʰetʰ-wor-* may have replaced an earlier form for "four", which is preserved in Anatolian. On the other hand, some have suggested that the original form for the number "four" was **H₂okʰtʰo-* and that "eight" was simply the dual of this stem, whose underlying meaning was "two fours" (cf. Gamkrelidze – Ivanov 1995. I: 747; Burrow 1973: 260). This suggestion finds support in Kartvelian (cf. Blažek 1999: 268).

The number "four" is reconstructed as **otxo-* in Proto-Kartvelian, and this is generally taken to be a loan from Proto-Indo-European (cf. Klimov 1998: 145–146; Fähnrich – Sardshweladse 1995: 269; Gamkrelidze – Ivanov 1995.I: 775 [Gamkrelidze – Ivanov reconstruct Proto-Kartvelian **(o)ŝt(o)-*]). I favor this explanation and consider **H*$_2$*ok*^{h}*t*^{h}*o-* to be the original form of the number "four" in Early Proto-Indo-European. It was replaced by **meyu-* in Anatolian, while, in the Proto-Indo-European antecedent of the non-Anatolian daughter languages it was replaced by **k*^{wh}*et*^{h}*-wor-*. It only survives in the later Proto-Indo-European form for the number "eight", **H*$_2$*ok*^{h}*t*^{h}*oH*$_1$*(w)*, a dual formation originally meaning "two fours". No doubt, this replaced an earlier form for the number "eight", which, unfortunately, can no longer be recovered.

One final comment may be made here: in Etruscan, there is a number *huθ*. Its exact meaning is uncertain – it could be "six", or it could be "four" (cf. Cristofani 1991: 77; Blažek 1999: 235; Bonfante – Bonfante 2002: 94–95). If it is "six", then the number *śa* is "four". On the other hand, if it is "four", then the number *śa* is "six". Without going into the whole question here of whether Etruscan and Proto-Indo-European are ultimately genetically related, we can say that *huθ* more closely resembles Proto-Indo-European **H*$_2$*ok*^{h}*t*^{h}*o-* "four", while *śa* more closely resembles Proto-Indo-European **s(w)eks* "six" (Szemerényi's reconstruction). As noted by Blažek (1999: 211 and 235) and Briquel (1994: 329), support for considering the meaning of *huθ* to be "four" comes from the identification of *huθ* in the Pre-Greek name Υττηνία Υττηνία for the city Tetrapolis (Τετράπολις, composed of τέτρα- "four" and πόλις "city") in Attica. This may provide another piece of evidence in support of considering **H*$_2$*ok*^{h}*t*^{h}*o-* to have been the original form for the number "four" in Proto-Indo-European.

The number "five" was **p*^{h}*enk*^{wh}*e* (Brugmann **peŋqᵘe*) in Late Proto-Indo-European. It is usually identified with words for "fist" and "finger": (1) Proto-Indo-European **p*^{h}*ŋ̍k*^{wh}*-st*^{h}*i-* "fist" > Proto-Germanic **fuŋχstiz* > West Germanic **fūχsti-* > **fūsti-* > Old English *fȳst* "fist"; Old Frisian *fest* "fist"; Middle Low German *fūst* "fist" (Dutch *vuist*); Old High German *fūst* "fist" (New High German *Faust*) (cf. Mann 1984–1987: 968 **pŋ̍kstis* [**pŋ̍qu̯stis* ?] "fist"; Onions 1966: 358; Kluge – Mitzka 1967: 187; Kluge – Seebold 1989: 205); Serbian Church Slavic *pęstъ* "fist"; (2) Proto-Indo-European **p*^{h}*enk*^{wh}*-ró-* "finger" > Proto-Germanic **fiŋgraz* "finger" > Gothic *figgrs* "finger"; Old Icelandic *fingr* "finger"; Old English *finger* "finger"; Old Frisian *finger* "finger"; Old Saxon *fingar* "finger"; Old High German *fingar* "finger" (New High German *Finger*) (cf. Feist 1939: 150; Lehmann 1986: 114; De Vries 1977: 120; Kluge – Mitzka 1967: 198; Kluge – Seebold 1989: 215). Though not without problems from a phonological point of view, the above comparisons can hardly be questioned. Ultimately, all of these forms may indeed go back to a verbal stem **p*^{h}*enk*^{wh}*-* "to take in hand, to handle", as suggested by Horowitz (cited by Blažek 1999: 228), though it should be mentioned that this putative verb stem is not attested in any of the daughter languages. Blažek (1999: 229) notes that the meanings "fist", etc. are primary.

Several different reconstructions are possible for the Proto-Indo-European word for the number "six": **sek^{h}s*, **swek^{h}s*, **k^{h}sek^{h}s*, **k^{h}swek^{h}s*, **wek^{h}s* (for details, cf. Blažek 1999: 234–242; see also Sihler 1995: 413). This number was also borrowed by Kartvelian: Proto-Kartvelian **ekšw-* "six" (cf. Klimov 1998: 48 **eks$_{1}$w-*; Fähnrich – Sardshweladse 1995: 125–126 **eks$_{1}$w-*; Schmidt 1962: 107 **ekšw-/*ekšu*; Gamkrelidze – Ivanov 1995. I: 775 **ekŝw-*). Sihler (1995: 413) takes **wek^{h}s* (he writes **weḱs*) to be the original form and considers the initial **s-* to be a secondary development (imported from the number "seven") (Szemerényi 1996: 222 and Beekes 1995: 213 express the same view). Thus, in accordance with Sihler's views, the earliest form of the Proto-Indo-European number "six" should be reconstructed as **wek^{h}s*. As Sihler notes, when **s-* was merely added to **wek^{h}s*, the result was **swek^{h}s*, but when it replaced the initial consonant, the result was **sek^{h}s*. The Iranian forms pointing to original **k^{h}swek^{h}s* (cf. Avestan *xšvaš* "six") appear to be due to developments specific to Iranian and should not be projected back into Proto-Indo-European (cf. Sihler 1995: 413).

The Proto-Indo-European word for the number "seven", **sep^{h}t^{h}m̥* (Brugmann **septṃ*), is sometimes considered to be a loan from Semitic (cf. Blažek 1999: 256–257; Gamkrelidze – Ivanov 1995. I: 747). That this number is ancient in Indo-European is clear from the fact that it is found in Hittite.

We have already discussed the number "eight". For "nine", Proto-Indo-European most likely had **newn̥* (cf. Szemerényi 1996: 223). Other possible reconstructions are *n̥/*h$_{1}$néwṃ* and **H$_{1}$newn̥/ṃ* (cf. Brugmann 1904: 365 **neu̯n̥*, **enu̯n̥*; Blažek 1999: 283 */ṃ*H$_{1}$newṃ*; Meier-Brügger 2003: 234 */n̥ *h$_{1}$néu̯n̥,*; Watkins 1998: 67 */n̥ *h$_{1}$néwn̥,*; Haudry 1979: 68 **néwṃ/n̥*; Burrow 1973: 260; Sihler 1995: 415 **H$_{1}$néwn̥*; Gamkrelidze – Ivanov 1995. I: 744 **neu(e)n*; Buck 1933: 230 [Buck takes Greek ἐννέα to be "a blend of *ἐνϜα and *νεϜα"]; Rix 1992: 172 **h$_{2}$néu̯n̥*).

As noted above, the Proto-Indo-European number "ten" must originally have been a compound meaning "two hands", that is, **t(e)-* "two" + **k^{h}m̥(t^{h})-* "hand".

The Proto-Indo-European word for the number "hundred" is traditionally reconstructed as **(d)ĥm̥tóm* – it is usually considered to be a derivative of **deḱm̥(t)* "ten" and meant something like "ten tens" (cf. Szemerényi 1996: 226; Watkins 1998: 67; Meier-Brügger 2003: 235; Beekes 1995: 213; Gamkrelidze – Ivanov 1995. I: 744).

Though there was probably no common Proto-Indo-European word for "thousand", the form **g^{h}eslo-* served as the basis for the Indo-Iranian, Greek, and Latin terms (cf. Szemerényi 1996: 227; Beekes 1995: 216; Meier-Brügger 2003: 235; Meillet 1964: 414; Brugmann 1904: 368). According to Bengtson (1987: 260–261), this form is to be derived from Proto-Indo-European **g^{h}es-* "hand" (he writes **ĝhes-*) plus a suffix **-lo-*.

We may now summarize our findings. The numbers "one" to "ten" may be reconstructed as shown below for the earliest stage of Proto-Indo-European. However,

inasmuch as Anatolian corroboration is lacking for several of the numbers, the following reconstructions must be considered provisional:

1. *H_2*ay-*(later *H_2*oy-*), **sem-*, **p^{h}er(H)-/*p^{h}r̥(H)-*
2. (earliest form) **t'e/a-*; (later also) **t'(u)w-a-*(still later **t'(u)w-o-*), **t'(u)w-i-*; though originally a plural form, this was later reinterpreted as a dual.
3. **t^{h}r-ey-/*t^{h}r-i-*
4. **H_2ok^h-t^{h}a-*(perhaps with original, non-apophonic *-o-* in the first syllable) (later **H_2ok^h-t^{h}o-*); replaced by **meyu-* "four" in the Anatolian languages and by **k^{wh}et^h-wor-* "four-sided, square" in the non-Anatolian daughter languages.
5. **p^{h}enk^{wh}e*
6. **wek^{h}s* (later also **swek^{h}s* ~ **sek^{h}s*)
7. **sep^{h}t^{h}m̥*
8. Original unknown; replaced by (dual) **H_2ok^{h}t^{h}oH_1(w)* "two fours" = "eight" in the non-Anatolian daughter languages.
9. **newn̥*
10. **t'e-k^{h}m̥(t^h)* (original meaning "two hands").

References

Adrados, Francisco R[odríguez]. 1975. *Lingüística indoeuropea [Indo-European Linguistics]*. 2 vols. Madrid: Gredos.

Adrados, Francisco R., Alberto Bernabé & Julia Mendoza. 1995–1998. *Manual de lingüística indoeuropea [Handbook of Indo-European Linguistics]*. 3 vols. Madrid: Ediciones Clásicas.

Beekes, Robert S.P. 1987. "The Word for 'Four' in Proto-Indo-European." *Journal of Indo-European Studies* 15: 1/2. 215–219.

Beekes, Robert S.P. 1995. *Comparative Indo-European Linguistics: An Introduction*. Amsterdam & Philadelphia: John Benjamins.

Bengtson, John D. 1987. "Notes on Indo-European '10', '100', and '1000'," *Diachronica* IV: 1/2. 257–262.

Benveniste, Émile. 1962. *Hittite et indo-européen [Hittite and Indo-European]*. Paris: Adrien-Maisonneuve.

Blažek, Václav. 1999. *Numerals. Comparative-Etymological Analysis and their Implications*. Brno: Masarykova Univerzita v Brně.

Bonfante, Giuliano & Larissa Bonfante. 1983. *The Etruscan Language: An Introduction*. New York, NY: New York University Press. [2002] [Revised Edition. Manchester: Manchester University Press.]

Briquel, Dominique. 1994. "Étrusque et indo-européen" [Etruscan and Indo-European]. *Langues indo-européennes [Indo-European Languages],* Ed. by Françoise Bader, 319–330. Paris: CNRS Éditions.

Brugmann, Karl. 1904. *Kurze vergleichende Grammatik der indogermanischen Sprachen. Auf Grund des fünfbändigen "Grundriß der vergleichenden Grammatik der indogermanischen Sprachen von K. Brugmann und B. Delbrück" [Concise Comparative Grammar of the Indo-European Languages. Based upon the Five Volume "Elements of the Comparative Grammar of the Indo-European Languages by K. Brugmann and B. Delbrück"].* Berlin: Walter de Gruyter. Reprinted 1970.

Buck, Carl Darling. 1933. *Comparative Grammar of Greek and Latin*. 10th impression 1966. Chicago & London: University of Chicago Press.

Burrow, Thomas. 1973. *The Sanskrit Language.* 3rd edition. London: Faber & Faber.

Colarusso, John. 1992. "Phyletic Links between Proto-Indo-European and Proto-Northwest Caucasian." *The Non-Slavic Languages of the USSR: Linguistic Studies (Second Series),* ed. by Howard I. Aronson, 19–54. Chicago, IL: Chicago Linguistic Society. [Reprinted in *Mother Tongue* (Newsletter) 21. 8–20 (1994).]

Cristofani, Mauro. 1991. *Introduzione allo studio dell'etrusco [Introduction to the Study of Etruscan]*. Leo S. Olschki Editore.

De Vries, Jan. 1977. *Altnordisches etymologisches Wörterbuch [Old Norse Etymological Dictionary]*. Reprint of 2nd Edition. Leiden: E.J. Brill.

Ernout, Alfred & Antoine Meillet. 1979. *Dictionnaire étymologique de la langue latine: Histoire des mots [Etymological Dictionary of the Latin Language: History of Words].* 4th Edition. Paris: Klincksieck.

Fähnrich, Heinz & Surab Sardshweladse. 1995. *Etymologisches Wörterbuch der Kartwel-Sprachen [Etymological Dictionary of the Kartvelian Languages]*. Leiden: E.J. Brill.

Feist, Sigmund. 1939. *Vergleichendes Wörterbuch der gotischen Sprache [Comparative Dictionary of the Gothic Language]*. 3rd Edition. Leiden: E.J. Brill.

Fortson, Benjamin W., IV. 2004. *Indo-European Language and Culture: An Introduction*. Oxford: Blackwell.

Gamkrelidze, Thomas V. & Vjačeslav V. Ivanov. 1995. *Indo-European and the Indo-Europeans: A Reconstruction and Historical Typological Analysis of a Protolanguage and a Proto-Culture.* 2 vols. English translation by Johanna Nichols. Berlin, New York & Amsterdam: Mouton de Gruyter.

Justus, Carol F. 1988. "Indo-European Numerals and Numeral Systems", *A Linguistic Happening in Memory of Benjamin Schwartz: Studies in Anatolian, Italic, and Other Indo-European Languages,* ed. by Yoël L. Arbeitman, 521–541. Louvain: Peeters.

Klimov, G[eorgij] V. 1998. *Etymological Dictionary of the Kartvelian Languages*. Berlin & New York: Mouton de Gruyter.

Kluge, Friedrich & Walther Mitzka. 1967. *Etymologisches Wörterbuch der deutschen Sprache* [Etymological Dictionary of the German Language]. 20th Edition. Berlin & New York: Walter de Gruyter.

Kluge, Friedrich & Elmar Seebold. 1989. *Etymologisches Wörterbuch der deutschen Sprache* [Etymological Dictionary of the German Language]. 22nd Edition. Berlin & New York: Walter de Gruyter.

Kronasser, Heinz 1956. *Vergleichende Laut- und Formenlehre des Hethitischen* [Comparative Phonology and Morphology of Hittite]. Heidelberg: Carl Winter.

Kuipers, A[ert] H. 1975. *A Dictionary of Proto-Circassian Roots*. Lisse: Peter de Ridder Press.

Laroche, Emmanuel. 1959. *Dictionnaire de la langue louvite [Dictionary of the Luwian Language]*. Paris: Adrien-Maisonneuve.

Laroche, Emmanuel. 1960. *Les hiéroglyphes hittites* [The Hittite Hieroglyphs]. Vol. I. Paris: Centre International de la Recherche Scientifique.

Lehmann, Winfred P. 1986. *A Gothic Etymological Dictionary.* Leiden: E.J. Brill.

Luraghi, Silvia. 1997. *Hittite.* Munich: Lincom Europa.

Luraghi, Silvia. 1998. "The Anatolian Languages." *The Indo-European Languages,* ed. by Anna Giacalone Ramat & Paolo Ramat, 169–196. London & New York: Routledge,

Mann, Stuart E. 1984–1987. *An Indo-European Comparative Dictionary*. Hamburg: Helmut Buske Verlag.

Markey, Thomas L. 1984. "The Grammaticalization and Institutionalization of Indo-European Hand." *Journal of Indo-European Studies* 12: 3/4. 261–292.

Mayrhofer, Manfred. 1956–1980. *Kurzegefaßtes etymologisches Wörterbuch des Altindischen* [Short Etymological Dictionary of Old Indic]. 4 vols. Heidelberg: Carl Winter.

Meier-Brügger, Michael. 2003. *Indo-European Linguistics*. Berlin: Walter de Gruyter.

Meillet, Antoine. 1964. *Introduction à l'étude comparative des langues indo-européennes [Introduction to the Comparative Study of the Indo-European Languages]*. University, AL: University of Alabama Press. Reprint of 8th Edition (1937).

Meriggi, Piero. 1962. *Hieroglyphisch-Hethitisch Glossar [Hieroglyphic Hittite Glossary]*. 2nd edition. Wiesbaden: Otto Harrassowitz.

Onions, C.T., Ed. 1966. *The Oxford Dictionary of English Etymology*. Oxford: Clarendon Press.

Rix, Helmut. 1992. *Historische Grammatik des Griechischen: Laut- und Formenlehre [Historical Grammar of Greek: Phonology and Morphology]*. 2nd Edition. Darmstadt: Wissenschaftliche Buchgesellschaft.

Schmidt, Karl Horst. 1962. *Studien zur Rekonstruktion des Lautstandes der südkaukasischen Grundsprache [Studies on the Reconstruction of the Sound Structure of the South Caucasian Parent Language]*. Wiesbaden: Kommissionsverlag Franz Steiner GMBH.

Sihler, Andrew L. 1995. *New Comparative Grammar of Greek and Latin*. New York & Oxford: Oxford University Press.

Sturtevant, Edgar H. 1933. *A Comparative Grammar of the Hittite Language*. Baltimore, MD: Linguistic Society of America. [1951] [Revised Edition. Vol. I. New Haven, Conn.: Yale University Press.]

Szemerényi, Oswald. 1960. *Studies in the Indo-European System of Numerals*. Heidelberg: Carl Winter.

Szemerényi, Oswald. 1996. *Introduction to Indo-European Linguistics*. Translated from the 4th edition (1990) of *Einführung in die vergleichende Sprachwissenschaft* [Introduction to Comparative Linguistics], with additional notes and references. Oxford: Clarendon Press.

Villar, Francisco. 1991. "The Numeral 'two' and its Number Marking." *Perspectives on Indo-European Language, Culture, and Religion: Studies in Honor of Edgar C. Polomé*, vol. 1, pp. 136–154. McLean, VA: Institute for the Study of Man,

Watkins, Calvert. 1998. "Proto-Indo-European: Comparison and Reconstruction." *The Indo-European Languages,* ed. by Anna Giacalone Ramat & Paolo Ramat, 25–73. London & New York: Routledge.

Some Old World experience of linguistic dating

J.A. Janhunen
University of Helsinki

This paper addresses the problem concerning linguistic depth and its measurement on the basis of empirical knowledge from selected Eurasian language families and areal complexes, including Indo-European, Ural-Altaic, and Palaeo-Siberian. There are clear differences between language families and their individual members with regard to the speed of evolution (innovative vs. conservative languages) and lexical openness (open vs. closed languages). Even so, the change of natural languages takes place within certain limits determined by extralinguistic factors, and for assessing these limits the methods of glottochronology and linguistic paleontology retain their relevance. The differentiation of the modern language families of Eurasia is the result of splits which correspond to actual prehistorical and protohistorical periods of cultural innovation, including the Neolithic and the Iron Age. Prospects of identifying deeper-level genetic affiliations are diminished by the relatively restricted comparative corpus available from the oldest reconstructable protolanguages, such as Uralic.

One of the most problematic parameters in diachronic linguistics is the speed of linguistic change. The comparative method gives us tools to identify the mechanisms of change and to estimate the relative chronological positions of languages sharing either a common origin (genetic cognates) or a history of mutual contacts (borrowings), but it says little definite of the absolute chronological depth of the shared linguistic phenomena. Even so, empirical knowledge tells us that the speed of linguistic change is not random. We know that no natural language remains unchanged for any significant period of time, but we also know that no language changes so rapidly that its functions as a means of communication would be severed.

The fact that the evolution of languages involves a constant balancing between stability and change has always been well known to practitioners of diachronic and comparative linguistics. It was also known to Morris Swadesh, whose glottochronological method was the first major attempt to formalize the impact of time on languages (Swadesh 1952). The two principal claims made by Swadesh were, first, that all languages incorporate an inherent tendency to change at a fixed rate,

and, second, that this rate affects the relative proportions of lost and preserved basic vocabulary. Conversely, the amount of shared basic vocabulary between any two mutually related languages allows, according to Swadesh, to give an absolute dating to the point of their divergence, that is, to the breakup of their common protolanguage.

Although glottochronology was immediately criticized (Hoijer 1956, and many others), we may today conclude that the critique was never able to invalidate the fundamental principles of the method. Rather, the critics only pointed out problems in how the method was being applied to actual linguistic data. A typical argument against Swadesh was that the individual members of many language families can actually be shown to have evolved at different rates, with some languages being more innovative than others. A case in point is offered by the Scandinavian languages, among which Icelandic is clearly closer to Proto-Scandinavian than its closest relative Norwegian (Bergsland & Vogt 1962). Similar differences are observed between the Romance languages, which, like Germanic, had been among the type examples for Swadesh when he calculated his original formula of lexical half-life.

1. Revising the glottochronological method

It is probably correct to say that the critique that was presented against glottochronology revealed three areas of problems which even today continue to call for a revision of the method. The three areas are: (1) the role of non-lexical innovations, (2) the need for etymological critique, and (3) the impact of extralinguistic circumstances. It may be taken for certain that none of these areas of problems went unnoticed by Swadesh himself, though he was unable to present a definitive solution to them. The basic issue is that no successful method can be created without making generalizations. The reality is always more complex than the rules designed to describe it, but without any rules the complexity of the reality is totally beyond description.

1. *The role of non-lexical innovations.* Swadesh implied that the lexicon is the most reliable indicator of linguistic change, but he certainly turned to the lexicon also for the simple reason that it is the part of linguistic structure easiest to quantify. However, innovations do take place also in the grammar, and it may be assumed that in most cases a lexically innovative language is also grammatically innovative, while conservative languages are conservative in all respects. To take an example from the Mongolic family, the number of phonological innovations separating Khamnigan Mongol (in Manchuria) from Proto-Mongolic is less than half of that exhibited by the neighbouring Mongolic languages (Janhunen 1992). Khamnigan Mongol is also

morphologically and lexically conservative, though these aspects of the language are more difficult to evaluate due to the mutual closeness of all Mongolic languages. On the other hand, Mangghuer (in the Amdo Qinghai region), which is lexically among the most innovative Mongolic languages, has also grammatically undergone a profound transformation (Slater 2003).

In principle, it should be possible to measure the speed of linguistic change also by counting other than lexical innovations. Under otherwise stable conditions, linguistic innovations are conditioned by the constant replacement of earlier generations of speakers by new ones. There is no reason why the new generations of speakers would make more innovations in some particular part of linguistic structure as compared to others. Under unstable conditions this can happen, however, as is illustrated by the phenomenon of the so-called creoles, or creolized languages, which tend to combine a relatively stable lexicon with a relatively innovative grammatical structure. This is, incidentally, also exemplified by English, a strongly creolized variety of Germanic, which in spite of its structural transformation retains a basic vocabulary consistent with its genetic origin.

Creolization is, however, a gradual phenomenon, and there is no sharp distinction between creoles and non-creoles. Creolization always involves asymmetric diachronic change, but the asymmetry need not be confined to the distinction between lexicon and grammar, for the intensity of innovation can also be different for different parts of the grammar. The attempts made to classify creoles as a special type of languages with "nongenetic transmission" (Thomason & Kaufman 1988) are therefore destined to lead to a dead end. Even strongly creolized languages, like the Romance type creoles, belong to their genetic lineages, though it happens that their lineages are most obvious in the lexicon. This has an important implication for the theory of linguistic dating, in that it seems to corroborate the idea of Swadesh concerning the diagnostic value of the lexicon. The relative stability of the basic vocabulary even in strongly creolized languages correctly gives them shallow glottochronological datings, while a focus on the grammar might give the false impression of a longer history of independent evolution.

While creolization involves a combination of grammatical innovation with lexical stability, there must also be languages that are lexically open and grammatically closed. A possible example is Japanese, which is famous for its ability to incorporate large numbers of foreign lexical elements, including items of basic vocabulary, into native grammatical patterns. Moreover, this property of lexical openness seems to be a relatively permanent, or at least recurrent, feature of Japanese, as is evident from its two principal layers of foreign vocabulary, Sino-Japanese (late 1st millennium) and Anglo-Japanese (late 2nd millennium). The point is that the grammatical markers are nevertheless of a native origin in Japanese, which means that there is a section even in the lexicon that is not immediately open to borrowing.

2. *The need for etymological critique.* It was already evident from the earliest applications of the glottochronological method by Swadesh that he often tended to ignore the critical experience of comparative linguistics, entering the world of random distant comparisons. The worst applications of glottochronology in the past are, in fact, cases in which the method has been applied to languages which are not even known to be genetically related. This has lead to a situation in which both accidental lookalikes and mutual borrowings have been used as evidence for genetic relationships, and as a basis for "dating" protolanguages which actually never existed.

The problem was noted by the late Sergei Starostin, who introduced his own version of glottochronology, which he called "etymostatistics" (Starostin 1989a). Unfortunately, although this was probably the most significant development in glottochronology since its beginnings, Starostin's model has not been widely followed. Stressing the importance of critical etymological work, Starostin correctly pointed out that glottochronological analysis should always be preceded by a careful elimination of lookalikes and borrowings. This is especially important in the comparison of closely related languages, whose primary separation may have been followed by a phase of secondary interaction. It goes without saying that only primary cognates should be considered when making glottochronological calculations.

Even more significantly, Starostin elaborated on the problems arising from the semantic approach to basic vocabulary, as proposed by Swadesh. It was immediately obvious that the so-called basic word lists, or lists of basic meanings (ranging from 100 to 215 items), compiled by Swadesh are not universally valid, and the items in them are not necessarily of equal value, for some meanings seem to be more likely to undergo lexical replacement than others. This was also noted by Gerhard Doerfer, who established that different items of basic vocabulary, such as the various terms for body parts, can have very different rates of diachronic resistance, as can be deduced from the time required for native speakers to recall them (Doerfer 1988). In principle, each item in the Swadesh lists should therefore have an individual coefficient expressing its idiosyncratic probability of being lost or preserved. Only a word list with correctly differentiated statistical weights of the individual items can serve as a basis for reliable glottochronological calculations.

In the absence of any simple solution to the problem, Starostin did not develop the idea of individual coefficients any further. He proposed, however, that word lists should be replaced by texts. Since the most basic and, hence, oldest lexical items tend to have the highest text frequency, we can, in principle, take any text from any given language, or even a word list treated as a text, and calculate the number of cognates that any fixed number of etymons (e.g., 100 or 200) extracted in a running order from the text have in one or more related languages. This is, without a doubt, Starostin's most important contribution to the issue, for the idea of operating with texts will allow glottochronological comparisons to be made also between languages for which

no synchronic word lists can be collected. Moreover, the method inherently makes evident the importance of proper etymological critique.

3. *The impact of extralinguistic circumstances.* The many extant examples of innovative and conservative languages imply that there must be extralinguistic factors that influence the speed of linguistic change. Some of these factors are connected with demographic parameters, such as birth and mortality rates, length of generations, and the overall age structure of the speech community. These parameters, in turn, are influenced by social, cultural, economic, and ecological circumstances, such as climate and vegetation, means of subsistence, patterns of habitation, and systems of kinship. In general, it may be assumed that a small but rapidly growing population with a large proportion of children and few old people is likely to be linguistically innovative. Demographic bottle necks created by natural disasters, including epidemics and famine, can presumably also promote rapid linguistic change.

Another sphere of extralinguistic factors is connected with the intensity of interaction with other speech communities. There is no doubt that contacts stimulate change, while isolation favours the *status quo.* The conservativeness of Icelandic, for instance, is quite certainly due to its marginal position and geographical isolation. On the other hand, the high degree of innovativeness of strongly creolized languages is connected with their history as interethnic mediums between people most of whom are originally native speakers of other languages. Ultimately, it may be the proportion of native and non-native speakers involved in the process that determines whether the innovations are concentrated in the lexicon or in the grammar.

The impact of extralinguistic factors shows that glottochronological calculations always require calibration. The original analogy to radioactive carbon dating is obvious, but it has to be recognized that calibration is a more serious problem in glottochronology than it is in physics. We simply do not know how the different components influencing the intensity of contacts could be objectively measured, and how the results of this multivariable analysis could be incorporated into glottochronological calculations. Although this sounds like an unsurmountable obstacle, the situation is not hopeless. Empirical experience suggests that languages diverging from a common source most often evolve at comparable speeds, but even when this is not the case, the difference remains within certain limits. To define these limits is perhaps the most acute challenge to glottochronology today.

2. Linguistic paleontology and chronological depth

While we may confidently conclude that glottochronology is not dead, the unsolved problems connected with the method make its application in the form proposed

by Swadesh precarious for the time being. Meanwhile, there are also other methods for dating protolanguages. A traditional, but intuitively reliable, method is offered by linguistic paleontology, which is based on the knowledge that lexical reconstruction provides information on the chronological and geographical context of the underlying protolanguage. The size of the reconstructable lexical corpus is also indicative of the time level of the protolanguage. Irrespective of the possibility that lexical replacement may have been more rapid in some branches of a language family than in others, a protolanguage for which very few lexical items can be reconstructed must, in general, be older than one with a larger lexical corpus.

In spite of frequent attempts to link protolanguages with actual archaeological cultures, the information provided by the linguistic data is rarely sufficient to allow definitive conclusions in this respect. In most cases, the archaeological evidence is simply too fragmentary, and the cultures established by archaeologists too diffuse, to make linguistic identifications possible. A typical mistake made in this connection is to forget that linguistic boundaries are not necessarily cultural boundaries. Even many modern cultures are multilingual, and many speech communities are multicultural. However, what the lexical evidence does allow us to establish is the cultural stage of the speech community that once spoke a given protolanguage.

Taking Northern Eurasia as an example, and judging by the size of the reconstructable lexicon, the oldest uncontroversially established protolanguage in the region would seem to be Proto-Uralic, which, even after recent additions (Aikio 2002), has a comparative lexical corpus of less than 200 items, including pronominal roots. A qualitative analysis of the Uralic lexical corpus would suggest a Mesolithic dating, which in absolute terms could mean anything between 5,000 and 10,000 years before present, for there are words for pre-Neolithic but probably post-Paleolithic innovations in hunting ("bow" and "arrow"), mobility ("ski", "to row"), and handicrafts ("glue", "to twine"), but no words connected with agriculture, domestication of animals, or trade. Another protolanguage with a similar lexical profile is Proto-Eskaleutic (Eskimo-Aleut), for which some 330 lexical items have been reconstructed, again including pronominal roots (Bergsland 1986).

The situation may be compared with that of other protolanguages in the region. For Kamchukotic (Chukchee-Kamchadal), for instance, a corpus of about 800 common words has been reconstructed (Mudrak 2000), a generous figure which in a more critical framework might turn out to be smaller, but which nevertheless suggests a time level less ancient than that of Proto-Uralic and Proto-Eskaleutic. Samoyedic, one of the two principal branches of Uralic, also has a an etymological corpus of just about 800 items (Janhunen 1977, with later additions). It is, however, noteworthy that there is nothing in the lexical evidence suggesting that the cultural level of the Proto-Kamchukotic speakers would have substantially differed from the Mesolithic, while the lexical corpus of Proto-Samoyedic suggests at least some familiarity with metal

working ("iron") and domestic animals ("horse", "domestic reindeeer"). In fact, external criteria (Bulghar Turkic loanwords of the Hunnish period) unambiguously give Proto-Samoyedic an Iron Age dating.

It has often been noted that the small size of the Proto-Uralic lexical corpus is partly due to the technical circumstance that the Uralic language family is normally analyzed as being composed of only two primary branches, Samoyedic and Finno-Ugric, which together comprise fewer than 40 distinct languages, all of them with a very short history of documentation. This is markedly different from Indo-European, which has almost a dozen "separate" branches and a total number of more than 400 languages, including some early ones that are amply documented in sources dating as far as 3500 years back. The unwillingness, or unability, to operate with a binary taxonomy on the Indo-European side increases the traditional lexical corpus of Proto-Indo-European (Pokorny 1959) to a size several times larger than that of Proto-Uralic. At the same time, the historical Indo-European languages function like a magnifying glass, bringing Proto-Indo-European one "generation" closer to us than would be the case if only contemporary languages were available.

It happens that there are also typological factors which make the results of lexical reconstruction in Uralic and Indo-European mutually difficult to compare (Janhunen 1997). The discrepancy between the two language families can, however, be reduced in two alternative ways. On the one hand, the Indo-European lexical corpus is substantially decreased if it is assumed that Indo-European (Indo-Hittite) also involves a binary division, between Anatolian and non-Anatolian, a solution supported by phonological and morphological isoglosses. On the other hand, the Uralic lexical corpus is considerably increased if it is assumed that Uralic is actually composed of a non-binary "comb," or "rake", with several branches very much like Indo-European, a hypothesis at least superficially supported by the small number of known non-lexical innovations between Proto-Uralic and Proto-Finno-Ugric (Salminen 2002). Either way, Uralic and Indo-European become more compatible, suggesting that the absolute ages of the two protolanguages may not be so different, after all.

However, the idea of a non-binary division of Uralic contradicts the lexical evidence, which clearly supports the status of Samoyedic as a separate primary branch (Michalove 2002). This conclusion can be disregarded only by assuming *ad hoc* that Samoyedic has been lexically exceptionally innovative, as compared with all other branches of Uralic. This, on the other hand, is in contradiction with the fact that Samoyedic, located in the eastern periphery of the Uralic family, is both phonologically and morphologically relatively conservative, just as the Finnic branch (Balto-Finnic) is in the west. The combination of lexical innovativeness with grammatical conservativeness would place Samoyedic in the category of languages of the Japanese type. In the case of Japanese we can, however, historically verify the situation and identify the external sources of lexical replacements (Chinese and English), while the

circumstances underlying the diachronic position of Samoyedic are unknown. Most probably, therefore, Samoyedic is simply so distantly related to Finno-Ugric that the number of common words has been reduced to its current level.

Attempts to bring Proto-Uralic up to a more shallow chronological level have also been made on the basis of a qualitative analysis of the lexical corpus. In particular, it has been pointed out that Proto-Uralic is normally reconstructed as having had one basic metal term ("metal", later attested in the meanings "copper", "iron"), which would mean that the language was spoken in the late Neolithic (Kallio 2006), at a time when native copper and meteorite iron started to be exploited. Unfortunately, not much weight can be placed on this argument, for the term in question (Proto-Uralic **wäckä*) involves multiple problems, including both irregular variation within Uralic and possible external connections in several other language families (Aalto 1959).

It is more illuminating to take a look at the numeral systems of the Uralic languages. These suggest a systematically binary westward-branching division of the language family (Janhunen 2000), with the decimal system becoming complete only at the level of the individual shallow-level branches. For Proto-Uralic, only the items for "two" and "five" can be reconstructed, while the Proto-Finno-Ugric numerals cover the whole range from "two" to "six". The rest of the items for the first decade were built successively in the shallow-level branches from either native or borrowed elements. The situation strongly suggests that Proto-Uralic speakers, unlike, for instance, Proto-Indo-European speakers, had only an incipient numeral system. The alternative assumption that Proto-Uralic had a more or less full set of numerals which were selectively replaced by innovations and borrowings in all branches of the family is simply not credible. It is well known that cases of numeral borrowing normally imply a simultaneous adoption of an entire numeral system ("seven" and "ten" in Uralic, Chinese numerals and classifiers in Japanese and Korean).

At least relatively speaking, Uralic therefore seems to remain a good candidate for the oldest extant language family in Eurasia. What this means in absolute terms is more difficult to say, since cultural levels at any given time have varied widely across the continent, and Mesolithic communities in one region may have been contemporaneous with Neolithic and post-Neolithic communities in other regions. In any case, the crucial cultural innovations denoted on the Indo-European side by words representing the protolanguage-level of reconstruction ("horse", "cart", "wheel", and others) were unknown to Proto-Uralic speakers, and remained so to Proto-Finno-Ugric speakers. It is, however, relevant to note that the two language families otherwise show a parallelism in their lines of development, in that their expansion seems to have involved two major stages: the primary stage corresponding to the breakup of the original protolanguage, and a secondary stage corresponding to the breakup of the principal branches. This is a pattern observed also in several other language families.

It is not difficult to see that the language families of Northern Eurasia represent basically three levels of chronological depth:

1. *The deep level,* datable to the Mesolithic or Neolithic stages of cultural evolution, in absolute terms perhaps 8,000 to 5,000 years ago. This was a period when several old and expansive language families were formed in Eurasia thanks to the population growth stimulated by innovations in food production, mobility, social structures, and warfare. Apart from Uralic and Indo-European, this chronological level comprises the large and diversified families of Semitic and Sino-Tibetan, and apparently also Austro-Asiatic and Austronesian. In this company, Uralic, in view of the small size of its lexical corpus, stands out as the potentially most ancient uncontroversially established entity, followed by Eskaleutic and the first stage of Indo-European (Indo-Hittite). The second stage of the Indo-European expansion is often, and probably correctly, connected with the Pastoral Revolution, which was a late Neolithic to early Metal Age innovation complex particularly increasing the mobility of non-settled populations in otherwise difficult regions (the steppe belt of Central Eurasia).
2. *The medium level*, datable to the Iron Age, some 3,000 to 1,000 years ago. Conspicuously many secondary branches of Indo-European (Celtic, Romance, Germanic, Balto-Slavic, Iranian, Indo-Aryan, and Tokharian), Sino-Tibetan (Sinitic, Bodic, and others), and most branches of Uralic (Saamic, Finnic, Mordvinic, Mansi-Hungarian, and Samoyedic) underwent a major breakup in exactly this period, resulting in the modern diversity of shallow subbranches and languages. Several other language families in Northern Eurasia with no known further affiliations also date from this period. These include Turkic (Turko-Bulgharic), Mongolic (with Para-Mongolic), Tungusic, Japonic (with Para-Japonic), and Yeniseic (Keto-Kottic). The origination of these language families is connected with political and military developments, some of which are documented in Western and Chinese sources of the time.
3. *The shallow level*, datable to Modern Era, in absolute terms hardly more than 500 years ago. This level typically comprises entities which exhibit very little internal diversification, and which therefore are often classified as isolates. Examples in Northern Eurasia include Ainu(ic), Amuric (Ghilyak or Nivkh), Yukaghir(ic), and Korean(ic). At least some of these cases probably involve the last surviving traces of originally larger language families, which had undergone a phase of diversification before the diversity was extinguished by internal or external factors. Shallow-level entities are also well known from the context of the extant larger language families, where they may similarly represent the last survivors of originally more diversified major branches (like Latin in the Italic branch of Indo-European).

Isolates naturally offer a tantalizing challenge to comparative linguistics, for their shallow depth makes it appear likely that they might still have living relatives among

other extant languages. Very often, isolates or shallow families are compared with each other, which is, however, a risky operation since the chronological step from the shallow synchronic data to the inevitably very deep hypothetical protolanguage is so steep. A warning example is offered by the comparisons of Japonic and Korean(ic), for which a common lexical corpus (apart from mutual and shared loanwords) can only be created by postulating correspondences that are so complicated as to make them unlikely to be true (Martin 1966). On the other hand, the comparison of isolates with larger language families involves the dangers of anachronism (incompatibility of the time levels compared) and omnicomparativism (simultaneous comparison of an entity with an entire range of other entities).

In any case, the evidence from real language families tells us that the more distant the protolanguage is, the smaller the reconstructed lexical corpus can be expected to be. This has always been a serious problem for the so-called Altaic comparisons, which have produced the incredible number of almost 3,000 lexical parallels (Starostin, Dybo & Mudrak 2003). Current estimates concerning the date of "Proto-Altaic" place it some 7,000–8,000 years before present, but this chronology is no proportion to the size of the reconstructed lexicon. It also has to be said that the Common Altaic lexical corpus comprises terms for cultural innovations, such as "stirrup", which are thousands of years younger than the proposed level of the protolanguage. In fact, these items are among the most convincing proofs against the Altaic Hypothesis (Róna-Tas 1973), since they prove that a large part of the Altaic comparative corpus consists of relatively late protohistorical loanwords.

Although long-range comparisons should always be given a chance, it is difficult to avoid the impression that their proponents are, as a rule, all too optimistic about the quantity of the comparative material that can remain of hypothetical relationships that would predate the extant language families by several millennia. To take only one example, Sergei Starostin, whose person involved a curious combination of rigorous critique and uncontrolled fantasy, supported his conception of a connection between Nostratic and Sino-Caucasian (both of them hypothetical constructions) by a comparative word list comprising as many as 213 items, plus 22 pronominal roots (Starostin 1989b). Whatever is thought of the general validity of his hypothesis, the figure is much too high. If the Uralic comparative corpus has a size of less than 200 items, we should not expect to be able to reconstruct more than a fraction of this for any entity representing an even deeper chronological level.

Ultimately, it is a question of whether there were major linguistic expansions before the Mesolithic and Neolithic developments that produced the extant deep-level language families. It is more likely that the Paleolithic stage of cultural development was characterized by a slower rate of linguistic diversification than later periods, simply because population growth was so slow and the means of mobility so limited. Of course, linguistic evolution as such cannot have taken place any slower

in the Paleolithic than in later times, but the conditions underlying the expansion of certain languages at the expense of others were absent for most of the time. Large language families, as known today, are descendants of protolanguages whose speakers had a consistent and recurrent cultural or demographic edge. We do not know whether such an edge could be formed in the Paleolithic. This is not to say that Paleolithic language families would be a theoretical impossibility. Their likelihood is, however, greatest in regions which, like the Americas, were first inhabited relatively late during the human expansion. In the Old World, the post-Paleolithic linguistic expansions have quite probably erased any traces of older language families.

3. Dating homelands and linguistic continuities

One of the most unexpected results of modern human genetics has been that human populations tend to have very old local roots. Although the methods of absolute dating in genetics continue to be even less accurate that those in diachronic linguistics, the fact remains that many modern populations can be shown to involve a long history of local continuity. This is perhaps most evident in marginal regions, such as the British Isles and Lapland in Europe, where traces of the original settlement in the early Holocene are still discernible in the genetic composition of the local people. This continuity can also be followed in local archaeology, which rarely suggests a complete replacement of older cultures. Rather, the pattern is one of gradual accumulation of new influences upon a continuous cultural heritage.

The observed continuities in archaeology and human genetics, in combination with political considerations, have led to a reinterpretation of the linguistic prehistory of many regions of the world. While it used to be assumed in the past that ethnic groups have "migrated" from their "homelands", arriving in new territories and carrying with them their languages, cultures, and genes, many scholars have recently started to claim that the prehistory of languages, like that of cultures and genes, involves considerable local continuity. These scholars seriously suggest that the direct ancestors of many modern languages, including those classified as Uralic (Wiik 2002) and Indo-European (Alinei 1996–2000), have arrived in their current areas of distribution much earlier than was assumed before, possibly already in the Paleolithic. Typically, the new approach is being marketed as a "crossdisciplinary" synthesis between linguistics, archaeology, and human genetics (Saukkonen 2006).

If the assumption of an early expansion of the extant large language families were true, several revisions would be necessary in the chronological framework of diachronic linguistics as a whole. For one thing, the speed of linguistic change would have to have been much slower than is conventionally assumed, for otherwise the early geographical breakup of protolanguages would have led to a much larger differentiation

of their descendants than is actually the case. Alternatively, the entire structure of language families, and the system of genetic relationships between languages, would have to be interpreted in a way substantially different from that of traditional comparative linguistics. It could be assumed, for instance, that the divergent model for explaining similarities between languages is wrong and should be replaced with a model based on convergence. In fact, the proponents of the linguistic "continuity theory" have proposed exactly such revisions to the established methods of diachronic and comparative linguistics (cf. also Pusztay 1995).

The reality is, however, very different. One only needs to take a look at some historically documented regions in order to become convinced of the fact that the connection of languages with geographical locations is generally very recent and loose. Take, for instance, Anatolia, a marginal peninsula in the Eastern Mediterranian. We know that Anatolia still 2,000–4,000 years ago was the home to a variety of ancient local languages, none of which survives today. Some of these languages belonged to the Anatolian branch of Indo-European, while others did not. By the time of the Byzantine Empire, all of these languages had been replaced by new ones, mainly by Greek, but marginally also by others (Armenian, Iranian, Arabic). However, even Greek was replaced by Turkish in the course of a political reorientation which took only 500 years to complete, ending with the fall of Constantinople (1453).

One asks where the Byzantine Greek speakers of Anatolia are now. Certainly, they did not move to Greece, nor were they eradicated by an act of genocide. What happened was that they changed their language and survive today in the genetic composition of the modern Anatolian Turks. The same had happened when the speakers of the earlier languages of the region had changed their speech to Greek. There was never any major immigration of an entirely new population to Anatolia, or a total replacement of the earlier population, a circumstance which explains why the local genetic composition and cultural patterns are so resistant. What was changed was the language, as well as some parts of the cultural orientation (including religion), and this happened basically by way of adaptation and assimilation.

The situation is very similar in other marginal regions of Eurasia, including the British Isles in the west, Fennoscandia (Scandinavia and Finland) in the north, and Korea and the Japanese Islands in the east. For all of these regions, we can postulate, and to some extent document, a recurrent replacement of earlier languages by newcomers. Most of the expansions directed towards these marginal regions were generated in the more centrally located parts of the continent, where, of course, the linguistic boundaries have also undergone movements in the course of time. Indeed, we do not have to go further back than the Middle Ages to find a linguistic map of Eurasia very different from that of the present day.

Linguistic boundaries do not, of course, move totally without human migration, but in the typical case the introduction of a new language takes place with a minimal

influx of new genes. In Eurasia, the proportional weight of the immigrant component in the genetic composition of the modern population is perhaps particularly high in Japan, where the arrival of the Japanese language, some 2,000 years ago, was connected with a total cultural reorientation from the local Jomon heritage with a basically Neolithic profile to the continental Yayoi complex with metal technology and agriculture (Hudson 1999). Even so, the genes of the earlier inhabitants of Japan also continue to live on in the composition of the modern Japanese.

It may be concluded that languages typically do not involve prolonged local continuities. Unlike genes and cultures, languages, as defined by their genetic lineages, are not anchored in localities. The few documented cases in which a single linguistic lineage is known to have been in more or less the same location for a period of 2,000–3,000 years, as in certain parts of Greece, Italy, and China, are absolute exceptions from the normal pattern, and even in these cases we know that the languages were ultimately injected from the outside into an environment where entirely different languages had been spoken before. It is a different matter that the earlier languages can have left traces in the typological orientation of the newcomers, though it is a considerable challenge to identify exactly those typological properties that are most likely to survive the process of language replacement (Nichols 1992).

In this connection, it is crucial not to ignore the factor of linguistic extinction. Even the deep-level language families that exist today have reached their full territorial extension relatively recently. The Neolithic and post-Neolithic expansions that have led to the diversification of the extant language families have covered an unknown number of relatively more "aboriginal" languages and language families. This is, in particular, true of Eurasia, where the synchronic language family density is conspicuously smaller than, for instance, in North America (Austerlitz 1980). The native populations in large parts of Eurasia, especially along the sparsely-inhabited arctic belt (Lapland, Arctic Russia, and Northern Siberia) have only recently adopted the languages they speak today (Helimski 2000).

Although we will never know how many languages were lost due to the expansion of the extant language families, it is reasonable to assume that the world has been developing towards decreasing linguistic diversity since the Neolithic. The population size of the pre-Neolithic speech communities is likely to have been more or less invariant under any given ecological circumstances. An idea of what the size of these speech communities may have been is given by those parts of the world that still retain, or have until recently retained, at least a token of the earlier diversity. For Northern Eurasia, the most reliable sample region is Siberia, where the size of the typical speech community in historical times has ranged from 500 to 5,000 individuals (Dolgix 1960). Taking this size as a guideline, even a relatively small territory like Finland may still have had as many as ten distinct languages before the arrival of

Finnish, which expanded less than 1,000 years ago mainly thanks to the slash-and-burn technology of agriculture (Janhunen 2005). For pre-Yayoi Japan, a diversity of up to 300 distinct languages may be postulated. Of these, only one, Ainu(ic), survives, and it has also moved from its earlier "homeland" in Central Japan to its historical location on Hokkaido, Sakhalin, and the Kurile Islands (Janhunen 2002).

4. Conclusions for long-range comparisons

With this paper, I wish to pay my homage to Hal Fleming, the indefatigable field linguist, enthusiastic long-ranger, and widely respected initiator of the *Mother Tongue* movement. I fully share his view that distant comparisons should always remain an important part of comparative linguistics. In a world with more than 400 language families, many of which are still diachronically little studied, there is always a chance that new cases of actual genetic relationship can be discovered. These chances are the greater the less well investigated linguistic region we are dealing with, while for well-documented parts of the world, like Eurasia, more caution is probably in place. Even in the absence of positive results, long-range comparisons can contribute to the understanding of the universal, typological, and areal properties of languages.

In Northern Eurasia, the most celebrated cases of potential distant relationship proposed so far are the Altaic, Ural-Altaic, Nostratic, and Eurasian hypotheses. While I agree with the critics of the Altaic hypothesis (Vovin 2005, and others) that the comparative material on which this particular framework is conventionally based is not genetically diagnostic, it is more difficult to take distance from some of the arguments presented in favour of the Ural-Altaic and Nostratic hypotheses. At least, the syntactic and morphosyntactic parallels across the Ural-Altaic belt (Fokos-Fuchs 1962), and the undeniable material correspondences in the pronominal systems of the Nostratic "phylum" (Dolgopolsky 1984) call for an explanation, which may, of course, turn out to be something different from a genetic relationship.

However this may be, distant comparisons should always be conducted with consideration of the experience available from well-established language families. This experience can perhaps be gathered under the following three points:

1. *Quality before quantity.* In all comparative work, small is beautiful, but this is nowhere more important than in long-range work. If we make distant comparisons between extant deep-level language families, we should not expect to be able to reconstruct a corpus of more than some dozens of good etymologies. In fact, this will be enough for the demonstration of a distant relationship, provided that the corpus is phonologically regular and paleolinguistically consistent, that is, that it does not include items connected with secondary cultural innovations.

2. *The extralinguistic explanation of expansions.* Since language families are formed by way of linguistic expansion and diversification, and since expansions always have a concrete cultural or demographic background, this must also be so with the language families established by distant comparisons. When postulating Paleolithic ancestors for a set of Neolithic or Mesolithic families, such as Uralic and Indo-European, an attempt should be made to explain the circumstances under which the underlying linguistic expansion could have taken place. Although a good comparative corpus can always speak for itself without extralinguistic evidence, many actual cases of distant relationship, involving only a fragmentary corpus, can be made more credible if the extralinguistic background of the postulated relationships can be explained. In this connection, the general reconstruction of the cultural, demographic, and linguistic conditions of the Paleolithic remains a challenge that should be taken by long-rangers and conventional comparativists alike.

3. *The shallowness of the linguistic map.* Even if distant comparisons can yield positive results, it should be recognized that languages move on the map, and they move fast. For this reason, the chances of locating homelands and protolanguages with any exactitude are equal to nil. The recent trend of dating linguistic continuities with the help of human genes and archaeology involves a serious misunderstanding of the processes that govern linguistic expansions. This is perhaps particularly relevant for long-rangers, for the prolonged local continuities of languages postulated on the basis of genes and cultures presuppose even earlier protolanguages than would otherwise be required, a circumstance that only decreases the credibility of the distant comparisons being proposed. In this respect, long-rangers and conventional comparativists have a common cause: we should actively promote an understanding of the fact that languages can spread, and do spread, irrespective of genes and cultures.

References

Aalto, Pentti. 1959. 'Ein alter Name des Kupfers.' *Ural-Altaische Jahrbücher* 31. 33–40.

Aikio, Ante. 2002. 'New and old Samoyed etymologies,' *Finnisch-Ugrische Forschungen* 57. 9–57.

Alinei, Mario. 1996–2000. *Origine delle lingue d'Europa,* vols. 1–2. Bologna: Il Mulino.

Austerlitz, Robert. 1980. 'Language family density in North America and Eurasia.' *Ural-Altaische Jahrbücher* 52. 1–10.

Bergsland, Knut. 1986. 'Comparative Eskimo-Aleut phonology and lexicon.' *Journal de la Société Finno-Ougrienne* 80. 63–137.

Bergsland, Knut & Hans Vogt. 1962. 'On the validity of glottochronology.' *Current Anthropology* 3. 115–153.

Doerfer, Gerhard. 1988. *Grundwort und Sprachmischung: Eine Untersuchung an Hand von Körperteilbezeichnungen.* Münchener ostasiatische Studien, vol. 47. Stuttgart: Steiner Verlag.

Dolgix, B.O. 1960. *Rodovoi i plemennoi sostav narodov Sibiri v. XVII v.* Trudy Instituta etnografiï, vol. 55. Moskva.

Dolgopolsky, Aron P. 1984. 'On personal pronouns in the Nostratic languages.' *Philologia Germanica 6* Linguistica et philologica: Gedenkschrift für Björn Collinder, 65–112. Wien: Wilhelm Braumüller.

Fokos-Fuchs, D.R. 1962. *Rolle der Syntax in der Frage nach Sprachverwandtschaft mit besonderer Berücksicht auf das Problem der ural-altaischen Sprachverwandtschaft.* Ural-Altaische Bibliothek, vol. 11. Wiesbaden: Otto Harrassowitz.

Helimski, Eugene. 2000. 'Before the Uralians came: White spots on the historical language map of Northern Eurasia and the Uralic languages.' *Congressus Nonus Internationalis Fenno-Ugristarum*, vol. 2, pp. 72–73. Tartu.

Hoijer, Harry. 1956. 'Lexicostatistics: A critique.' *Language* 32. 49–60.

Hudson, Mark J. 1999. *Ruins of Identity: Ethnogenesis in the Japanese Islands.* Honolulu: University of Hawai'i Press.

Janhunen, Juha. 1977. *Samojedischer Wortschatz: Gemeinsamojedische Etymologien.* Castrenianumin toimitteita, vol. 17. Helsinki.

Janhunen, Juha. 1992. 'On the position of Khamnigan Mongol.' *Journal de la Société Finno-Ougrienne* 84. 115–143.

Janhunen, Juha. 1997. 'Kriticheskie zamechaniya k indo-uraljskim sravneniyam poslednego vremeni.' *Balto-slavyanskie issledovaniya* 1988–1996, 178–183.

Janhunen, Juha. 2000. 'Reconstructing Pre-Proto-Uralic: Spanning the millennia of linguistic evolution.' *Congressus Nonus Internationalis Fenno-Ugristarum*, vol. 1, 59–76. Tartu.

Janhunen, Juha. 2002. 'On the chronology of the Ainu ethnic complex.' *Bulletin of the Hokkaido Museum of Northern Peoples* 11. 1–20.

Janhunen, Juha. 2005. 'När kom finnarna till Finland?' [When did the Finns come to Finland?] *Sphinx.* Societas Scientiarum Fennica Yearbook 2004–2005, 77–91.

Kallio, Petri. 2006. 'Suomen kantakielten absoluuttista kronologiaa [On the absolute chronology of the proto-languages of Finnish].' *Virittäjä* 2006: 1. 2–25.

Martin, Samuel E. 1966. 'Lexical evidence relating Korean to Japanese.' *Language* 42. 185–251.

Michalove, Peter A. 2002. 'The classification of the Uralic languages: Lexical evidence from Finno-Ugric.' *Finnisch-Ugrische Forschungen* 57. 58–67.

Mudrak, O[leg] A. 2000. *Etimologicheskii slovarj chukotsko-kamchatskix iazykov.* Moskva: Iazyki russkoi kuljtury.

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

Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch.* Bern: Francke Verlag.

Pusztay, János. 1995. *Diskussionsbeiträge zur Grundsprachenforschung Beispiel: das Proto-Uralische*. Veröffentlichungen der Societas Uralo-Altaica, vol. 43. Wiesbaden: Harrassowitz verlag.

Róna-Tas, András. 1973. 'Did the Proto Altaic people know the stirrup?' *Studia Mongolica* 1: 13. 169–171.

Salminen, T[apani]. 2002. 'Problems in the taxonomy of the Uralic languages in the light of modern comparative studies.' In: *Lingvisticheskii bespredel.* Sbornik statei k 70-letiyu A.I. Kuznecovoi, 44–55. Moskva.

Saukkonen, Pauli. 2006. *Suomalais-ugrilaisten kansojen ja kielten alkuperäongelma* [The Problem Concerning the Origin of Finno-Ugric Peoples and Languages]. Helsinki: Helsinki University Press.

Slater, Keith. 2003. *A Grammar of Mangghuer: A Mongolic Language of China's Qinghai-Gansu Sprachbund.* RoutledgeCurzon Asian Linguistic Series, vol. 2. London: RoutledgeCurzon.

Starostin, S[ergei] A. 1989a. 'Sravniteljno-istoricheskoe iazykoznanie i leksikostatistika.' In: *Lingvisticheskaia rekonstrukciia i drevneishaia istoriia vostoka*, vol. 1, 3–39. Moskva.

Starostin, S[ergei] A. 1989b. 'Nostratic and Sino-Caucasian.' In: *Lingvisticheskaia rekonstrukciia i drevneishaia istoriia vostoka*, vol. 1, 106–124. Moskva.

Starostin, Sergei [A.], Anna Dybo & Oleg [A.] Mudrak. 2003. *Etymological Dictionary of the Altaic Languages.* Handbuch der Orientalistik 8/1–3. Leiden: Brill.

Swadesh, Morris. 1952. 'Lexicostatistic dating of prehistoric ethnic contacts.' *Proceedings of the American Philosophical Society* 96. 452–463.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics.* Berkeley & Los Angeles: University of California Press.

Vovin, Alexander. 2005. 'The end of the Altaic controversy.' *Central Asiatic Journal* 49. 71–132.

Wiik, Kalevi. 2002. *Eurooppalaisten juuret* [The Roots of Europeans] Jyväskylä: Atena Kustannus Oy.

The languages of northern Eurasia

Inference to the best explanation

John D. Bengtson
Vice President, Association for the Study
of Language in Prehistory

This paper discusses the development of hypotheses of classification of the languages of northern Eurasia, from the early "Scythian" hypothesis to the later Nostratic, Eurasiatic, Sino-Caucasian, and Dene-Caucasian proposals. The concept of scientific "proof" is discussed and contrasted with an alternative concept of "best explanation." Eurasiatic and Dene-Caucasian can then be viewed as testable and fruitful hypotheses that, so far, provide the best explanations for language diversification in northern Eurasia.

1. Language families and macrofamilies of northern Eurasia

Linguists have been using modern methods of classifying the languages of northern Eurasia for the past four centuries. By the 1700's major families such as Indo-European, Uralic, and Altaic had been recognized (however sketchily), and not much later some began to speculate about even deeper families. Rasmus Rask, for example, a pioneer in the description of Indo-European, proposed a "Scythian" language family that encompassed (in roughly west-to-east order) Basque, Uralic, Caucasian, Turkic, Mongolian, Manchu-Tungus, Chukchi-Kamchatkan, and Eskimo (1834).[1] Later on other scholars (Holger Pedersen, Björn Collinder, V.M. Illich-Svitych, Aharon Dolgopolsky, Allan Bomhard, Joseph H. Greenberg, et al.) refined the "Scythian" hypothesis, calling it instead "Nostratic" or "Eurasiatic."[2] All of the latter linguists agree in adding Indo-European to, and removing Basque and Caucasian (at least North Caucasian)[3] from, the family. Some also add Gilyak (Nivkh), Kartvelian, Dravidian, Elamite, and Afro-Asiatic,

---

1. As cited by Ruhlen (1987: 128).

2. Blažek (p.c.) adds Heinrich Koeppelmann (*Die euroasiatische Sprachfamilie.* Heidelberg 1933).

3. During the twentieth century, especially its latter half, there has developed a growing consensus that North Caucasian is genetically distinct from South Caucasian, along with a tendency to use the names Caucasian and Kartvelian, respectively (cf. Fleming 1987: 163–164; Blažek 1997: 161).

though at varying degrees of remoteness.[4] Just within the past decade there seems to be a growing consensus that there is a "core" **Eurasiatic** family consisting of Indo-European, Kartvelian, Uralic, Yukaghir, Altaic (including Korean and Japanese), Chukchi-Kamchatkan, and Eskimo-Aleut.[5] Harold Fleming (1987) has called this macrofamily Mitian.[6] Dravidian and Elamite may be further outliers, and many paleolinguists now agree that Afro-Asiatic can be considered another macrofamily roughly coordinate with Nostratic/Eurasiatic, rather than a part of it.[7]

Contemporaneous with the gradual development of the Eurasiatic hypothesis a few linguists began to get inklings of another major macrofamily in northern Eurasia. As we saw above, Basque and (North) Caucasian were increasingly viewed as distinct from Eurasiatic, and relatives for them were sought in other languages, usually Burushaski, Sino-Tibetan, and Yeniseian. The German Karl Bouda was probably the most assiduous in this regard, publishing a series of articles (1938–1968) connecting all five of the genetic units named above.[8] The American anthropologist Edward Sapir proposed a genetic connection between Sino-Tibetan ("Indo-Chinese") and the Na-Dene family of North America, and the Sino-Tibetanist Robert Shafer independently proposed a similar relationship.[9] In the 1980's these hypotheses were re-examined by the Russian linguist Sergei A. Starostin, who found them plausible (at least as regards Caucasian, Sino-Tibetan and Yeniseian) and proposed a "Sino-Caucasian" family with etymologies and sound correspondences. Starostin's Muscovite colleague Sergei L. Nikolayev

---

4. For a comparison of different versions of Nostratic/Eurasiatic, see Ruhlen (1987: 257–259), Blažek (1997: 165).

5. Gilyak (Nivkh) and Ainu should be mentioned. Greenberg (2000) includes both in his Eurasiatic. Dolgopolsky (1984) and Bomhard (1996) include only Gilyak. Some paleolinguists think Ainu belongs rather to the Austric macrofamily. (See Ruhlen [1987: 127ff.]; Fleming [1987: 173–174]; Blažek [1997: 163].)

6. From the characteristic personal pronouns, *mi* "me" and *ti* "thee". Blažek (p.c.) informs me that the term "Mitian" was used by the Rumanian scholars Marius Sala and Ioana Vintilă-Rădulescu (*Les langues du monde,* Bucureşti/Paris, 1984).

7. Henceforth the term Eurasiatic will be used to designate the "core" Eurasiatic as defined by Greenberg (2000) and Bomhard (1996). Starostin (1999) used Eurasiatic in a much wider sense, for a "superfamily" encompassing Dene-Caucasian, Afro-Asiatic, and Nostratic. Robert Shafer (of Sino-Tibetan fame) had a similar concept he called "Eurasial." See Renfrew & Nettle (1999), especially the articles by Bomhard and Starostin, for evolving ideas on the subgrouping of Nostratic and Eurasiatic.

8. For a list of references see especially Blažek (1997), Blažek & Bengtson (1995). Note that for Bouda Caucasian included North Caucasian + Kartvelian, and Sino-Tibetan included Tai.

9. For a discussion of, and additional support for, Sapir's and Shafer's "Sino-Dene," see Bengtson (1994).

then revived Sapir's Sino-Dene proposal by comparing Na-Dene and Caucasian (with an appendix listing Starostin's Sino-Tibetan and Yeniseian cognates).[10] After further contributions by Václav Blažek, Merritt Ruhlen, and this writer, the present-day **Dene-Caucasian** hypothesis includes Basque, Caucasian, Burushaski, Sino-Tibetan, Yeniseian, and Na-Dene.[11]

So at the present time paleolinguists, insofar as they are willing to commit to remote linguistic relationships, generally agree that there are two major macrofamily-hypotheses to account for the origins of the languages of northern Eurasia: **Eurasiatic** and **Dene-Caucasian**.[12] One of them, Eurasiatic, has generated quite a bit of interest and discussion among historical linguists (see, e.g., Renfrew & Nettle [1999]; Hegedűs & Sidwell [2004]). On the other hand, Dene-Caucasian (DC), outside of the journal *Mother Tongue,* has hardly been discussed at all by linguists, and even among "long rangers" DC seems to be less accepted than Eurasiatic, a fact that I discussed in an article several years ago (Bengtson 1998). There I presumed some of the causes of this lack of acceptance to be as follows:

> … Starostin and Nikolaev, in their earlier articles, compared reconstructions without listing attested forms … ; … the Caucasian reconstruction is forbiddingly complex, and (until recently [1994]) its full methods and materials were not published; … the classification of Caucasian languages has not been resolved to everyone's satisfaction; and … the Caucasian and Sino-Tibetan [and Yeniseian] peoples, in particular, are thought to be too distant from each other, geographically, historically, racially, and linguistically, to be seriously considered as kin. (Bengtson 1998: 35)

Now, some eight years later, DC is still in need of serious discussion, let alone acceptance. There has never been a Dene-Caucasian Symposium to compare with the Nostratic Symposium held at Cambridge University in 1998, or the Nostratic Centennial Conference (University of Pécs 2003). But there have been some positive developments: DC has been one of the major foci of the Evolution of Human Languages Project sponsored by the Santa Fe Institute (2001 through the present), in which

---

10. Starostin's and Nikolayev's articles, originally in Russian, were published in English in Shevoroshkin (ed.) (1991).

11. Possibly also Sumerian (Bengtson 1997a). Shevoroshkin (1999, 2003, 2004) includes other North American families besides Na-Dene: Salishan, Wakashan, Algic, Siouan, but other paleolinguists have not yet concurred in this. Blažek & Bengtson (1995) included Kusunda data in some Dene-Caucasian etymologies. This needs to be re-evaluated *vs.* the Indo-Pacific affiliation proposed by Whitehouse et al. (2004).

12. Afro-Asiatic in the southwest, and Austric in the southeast, while of great interest in their own right, will be disregarded in this paper, which focuses on northern Eurasia.

Sergei Starostin and this writer (and to a lesser extent also George Starostin) have collaborated to collect, evaluate, and revise the evidence for DC – lexical, phonological, and morphological. Recently, before his untimely death, Sergei Starostin worked on Burushaski and discovered many new Burushaski-DC etymologies, which, along with earlier etymologies, now appear on the EHL/Tower of Babel website.[13]

2. Questions of method

In this paper I would like to go to the crux of the matter: what is necessary to demonstrate the validity of a macrofamily – for example, the Dene-Caucasian macrofamily? Some linguists have gone so far as to demand that a language family or relationship be "proved beyond a reasonable doubt." It hardly needs saying that this requirement may be appropriate to a court of law, but few scientists – whether in the "hard" sciences of physics, chemistry, etc., or in the "softer" social sciences – would take it seriously as a guideline. Other linguists understand the relativity of proof, and they, the "relativists," "seem to treat probable [linguistic] affiliations as differing in degree, some being more probable than others," while the "absolutists appear to regard some genetic connections as indisputable and others as inconceivable" (Wescott 1999). The following quotes (with emphasis added) epitomize what seems to me to be a reasonable "relativist" position:

> In all empirical sciences … all that we *can* get are results so close to certainty that for all practical purposes we can consider them true, that is, a hypothesis which is overwhelmingly better than any other in accounting for the facts. (Greenberg 1995: 207)
>
> Rather than define "proof" in linguistic investigations of genetic relatedness by proposing pseudo-specific percentages of various sorts, I would like to replace this notion with the fuzzier, but pragmatically more realistic, concept of "superior hypothesis." … Once the historical linguist has identified a diagnostic core of traits that cannot plausibly be coincidental, are too extensive to be explained by known language contact phenomena, and clearly do not exist together in any other languages, the hypothesis of genetic relatedness becomes the best explanation. (Vajda 1999: 88–90)
>
> The job of the comparative linguist is to provide the best explanation possible consistent with the facts. In proposing a classification, it is *not* necessary that the linguist "prove" that the classification is absolutely certain by the presentation of

13. http://ehl.santafe.edu or http://starling.rinet.ru.

> conclusive evidence. In response to widely speculative classifications … many careful, empirically based linguists jumped to the opposite extreme and took the position that all languages should be treated as unrelated unless and until proved otherwise. … However, on closer inspection, this requirement turns out to be untenable and not in keeping with standard scientific procedures. (Newman 2000: 261)
>
> The goal of an historical linguist should not be to demand an explanation for everything before he believes anything [about remote linguistic relationships such as Amerind and Eurasiatic] … Rather a scientist should attempt to explain non-random phenomena, for example, the prevalence of the N/M "I/thou" pronominal pattern in the Americas, and a different pronominal pattern, M/T "I/thou" in northern Eurasia. … if common origin is not responsible for these different pronominal patterns, what is? (Ruhlen 2005: 348)

Another important concept is that of negative evidence, i.e., the usually unspoken assumption that *other* languages are *not* a part of the family we are trying to demonstrate. "But in our eagerness to look at the inclusive aspect of the relationships we lose sight of the exclusive – yet more than anything else it is the lack of evidence for the affiliation to the languages excluded that makes the proposal truly meaningful. This lack of evidence is never demonstrated, however, and so, in effect, the most important evidence for each proposal is routinely omitted." (Whitehouse 2003)

Greenberg (1995: 207) mentions an important characteristic of a *best explanation* hypothesis of linguistic relatedness. "Such hypotheses have a further important characteristic, which we may call fruitfulness, That is, they can be built upon and lead to further discoveries and explanations of hitherto puzzling phenomena."

In what follows I would like to emphasize the following questions (a) What is the *negative evidence*, in the Whitehouse sense, i.e., the evidence that makes Eurasiatic "not Dene-Caucasian" and Dene-Caucasian "not Eurasiatic"? This can be stated positively by outlining diagnostic differences between the two macrofamilies; (b) the fruitfulness question: is Dene-Caucasian a fruitful hypothesis?; (c) The question of the *best explanation* for language diversification (and thus ethnogenesis) in northern Eurasia: are the *superior hypotheses* Eurasiatic and Dene-Caucasian, or are other hypotheses preferable, or even possible?

3. Differences between Eurasiatic and Dene-Caucasian

There is a number of diagnostic features in which Eurasiatic and Dene-Caucasian differ. The lexical differences are numerous and will not be discussed here. It will be

more illuminating, as well as much more succinct, to concentrate on some of the differences in *phonology* and *morphology*.

3.1 Phonological differences

3.1.1 *Laterals*

A remarkable phonetic feature of Dene-Caucasian is an abundance of lateral affricates, the sounds we will designate here as *dl* (lenis), *tl* (fortis), and *tl'* (glottal). All three affricates persist in most Na-Dene languages, for example, Navajo *dlõõ'* "prairie dog", *-tlé/-tlèè'* "socks", *tl'éé'* "night". The next best preservation is in some East Caucasian languages, especially Avar, Andian, and Tsezian. Note, for example, some of the cognates I have proposed for the Navajo words just mentioned: Akhwakh *tl':ã-k'a* "hare", *itl':e-l* "stocking", *ratla* "night", respectively, though with much reorganization of the original laryngeal qualities.[14] In all the other DC languages the original lateral affricates have been replaced by clusters or single phonemes. In a paper yet to be published I propose about 100 etymologies involving the Proto-DC lateral affricates *dl, tl, tl',* and their reflexes.[15] An interesting aspect of the latter is that both Basque and Burushaski turned lateral affricates *between vowels* into clusters that reverse the original phonetic sequence (TR > RT).[16] In Basque the cluster is *r̄d* or *r̄t,*[17] in Burushaski *lt or ld* (see below for etymological examples).

The presence of lateral affricates is a diagnostic feature that distinguishes Dene-Caucasian from Eurasiatic. Starostin (1989), in his comparison of Dene-Caucasian and Eurasiatic, lists several etymologies in which DC lateral affricates match plain laterals in Eurasiatic:[18]

---

14. As far as I know, no modern Caucasian language preserves the threefold contrast of Proto-Caucasian (**dl ~*tl ~ tl'*). Northern Akhwakh, for example, has a four-fold contrast of fortis (*tl*), tense fortis (*tl:*), glottal (*tl'*), and tense glottal (*tl':*).

15. "The Dene-Caucasian Lateral Affricates." Of course, many of these etymologies were already proposed by S.A. Starostin and others.

16. Where T = coronal (stop or fricative) and R = resonant liquid.

17. The symbol *r̄* denotes the long or trilled *r* of Basque, written *rr* between vowels, e.g., *berri* [beri] "new, young".

18. These comparisons imply that Eurasiatic and Dene-Caucasian spring, ultimately, from a common source – Hal Fleming's "Boreal." Reconstructions have been revised, and cognates listed, in accord with the EHL/Tower of Babel database (http://ehl.santafe.edu or http://starling.rinet.ru). I have added Basque comparanda where relevant. The symbol (~) indicates "compared with, possibly cognate."

1. DC ***Htl'wĭnV*** "winter, year" (Avar *tl':in* "winter", Burushaski **téne* "last year"[19]) ~ EUR ****luńgV*** "snow" (Finnish *lumi,* Tungus *lune* "wet snow", Korean *nun* "snow")
2. DC ****tl'ĕp'V*** "flat (thing)" (Avar *tl'eb* "stone", Burushaski **tápi* "stony terrace", Old Chinese **lēp* "tablet", etc.) ~ EUR ****lVp'V*** "flat (thing)" (Evenki *lāpčā* "fish fins", Russian *lapa* "paw", Norwegian *love* "palm of the hand", *labb* "paw", etc.)
3. DC ****tlHwémV*** "liquid, soak" (Avar *tl:amí-* "liquid", Burushaski **tam* "bathe, swim, wash", Basque *limuri* "slippery, humid", Old Chinese **ləm* "soak", etc.) ~ EUR ****lVmV*** "wet" (Georgian *lump'-* "wet", Japanese *náma* "fresh, raw", etc.)
4. DC ****tl'ăpĭ*** "leaf" (Lak *č'ap'i,* Burushaski **ltap* "leaf, petal", Basque *lapar* "bramble", Old Chinese **lap* "leaf, foliage", Kott *ďipi* "leaf", etc.) ~ EUR ****lVp'V*** "leaf, bark" (Hungarian *levél* "leaf", Turkish *yaprak,* Georgian *lapan-* "bast", Lithuanian *lāpas* "leaf", Greek λοπός "husk, peel, bark", etc.)
5. DC ****[m]ĥétl'V*** "milk, butter" (Archi *nåtl':* "milk",[20] Chechen *nalχa* "butter", Burushaski **maltáṣ* "butter", etc.) ~ EUR ****mälgi*** "breast, udder" (English *milk,* Saami *mielgâ* "breast, chest", Yukaghir *mełu(t)* "chest", Yupik *mulik* "nipple", etc.)[21]
6. DC ****tl'ănpV*** "tongue; lip" (Tsez *tl'epu* "lip", Hatti *alip* "tongue", Jingpo šiŋ-*lep* "tongue", Kott *alup*) ~ EUR ****l[a]pV*** "to lick, lip" (Mongolian *lab-si-* "eat greedily", English *lip,* Latin *labium,* etc.)
7. DC * = ***iwtl'E*** "die, kill" (Bezhta *–itl'-* "to kill", Chechen = *al-* "to die", Basque **hil* "die, dead", Burushaski *[d-]*-l-* "to hit, sting, beat, kill", Haida *tl'a*-dáa "to kill [several things]") ~ EUR ****HwelV*** "to fight, kill" (Turkish *öl-* "die", Hittite *hulla-/-i-* "fight, beat down", Old Norse *valr* "dead on the battlefield", etc.)

Proto-Afro-Asiatic also had lateral affricates, e.g., **PAA **tlap-*** "leaf" (in Chadic, Cushitic) ~ DC ****tl'ăpĭ*** "leaf", EUR ****lVp'V*** "leaf, bark"; **PAA **tlV(n)f-*** "lip" (Semitic, Egyptian, Cushitic) ~ **DC **tl'ănpV*** "tongue; lip", **EUR **l[a]pV*** "to lick, lip".

3.1.2 *Velars/postvelars*

Another difference between Eurasiatic and Dene-Caucasian is the presence of a uvular (postvelar) series in the latter but not in the former. As Starostin (1989: 45) pointed out, the reconstruction of uvular consonants in Eurasiatic (his "Nostratic") was based only on Kartvelian evidence, which could be the result of areal influence (cf. uvular

---

19. Hunza-Nager *téndili* "last year", Yasin *téne* "year before last". The association of "winter" and "year" in northern Eurasia is commonplace.

20. *å* denotes a pharyngealized vowel. The lateral affricates in Archi have a velar tinge, thus the word for milk might be more accurately written *nåkl':*. In all the other Lezgian languages lateral affricates have become plain velars (or uvulars), e.g., Lezgi *nek,* Rutul *näk,* Udi *naq:* "milk", etc.

21. Cf. Ruhlen (1994c).

consonants of quite recent origin in the Turkic languages of the Caucasus).[22] After the breakup of Dene-Caucasian uvulars remained in Caucasian, Burushaski, Yeniseian, and Na-Dene, but were lost (merged with velars or 0) in Basque and Sino-Tibetan. Note some of the comparisons adduced by Starostin (1989):

8. DC ***č'HəqwV*** "thick, big, abundant" (Dargwa *čaqw-* "high", Burushaski **śóq-*um "wide, broad", Basque **aśko* "much, many", **aśki* "enough", Tibetan *ćhog* "be enough", Eyak-Athabaskan **ćə̀xw* "big", etc.) ~ EUR ****č'ok'V*** "big, many" (Turkish *çok* "many, very", Hungarian *sok* "many", etc.)
9. DC ****qwänV*** "woman" (Khinalug *χin-* "woman", Chechen *qin* "brother-in-law's wife", Ket *qi:m* "woman", Burushaski **γen-* "queen",[23] etc.) ~ EUR ****k/u/nV*** "woman" (Azerbaijani *günü* "co-wife [in polygamy]", Proto-Japanese **kuanami* "first wife [in polygamy]", Armenian *kin* "wife", Persian *zan,* English *queen,* etc.)
10. DC ****χqwīrhV*** "horn" (Lak *qi,* Ubykh *q´a,* Ket *qå',* etc.) ~ EUR ****KirV*** "horn, top of head" (Finnish *kiire* "crown of the head", Hittite *karawar* "horn", Greek **κέρας**, etc.)
11. DC ****nV(x)qwV*** "back, behind" (Avar *náq:a* "behind", Eyak tli-*naq-*t, Burmese *nok* "space behind", etc.) ~ EUR ****nVkV*** "neck, vertebra" (Hungarian *nyak* "neck", Turkish *yaka* "collar", Evenki *nikimna* "neck", English *neck,* etc.)

Afro-Asiatic also has a uvular series, though it does not always agree with Dene-Caucasian in the placement of uvulars, e.g., AA ****kVn-*** "co-wife, sister-in-law" (Semitic, Berber, Chadic, Agaw) ~ DC ****qwänV*** "woman", EUR ****k/u/nV*** "woman"; AA ****k'ar-*** "horn" (Semitic, Egyptian, Omotic) ~ DC ****χqwīrhV*** "horn", EUR ****KirV*** "horn, top of head", etc.

3.2 Morphological differences

3.2.1 *Noun class/gender*

A salient difference between Eurasiatic and Dene-Caucasian is in their respective systems of noun classes or genders. There is abundant evidence that the original DC morphological system included noun classification into a number of different classes, based on the oppositions of living/non-living and masculine/feminine. Such systems are still operational in most of the Caucasian languages (2–8 classes), Burushaski (4 classes), and Yeniseian (Ket: 3 classes). In the remaining Eurasian DC families, Basque and Sino-Tibetan, the morphological function of noun classification was lost, but the former existence of such a system can be inferred from what appear

---

22. As Whitehouse (p.c.) points out, an alternative explanation could be that Proto-Eurasiatic had uvulars, and only Kartvelian (as the first family to split off) retained them.

23. Burushaski /γ/ is a voiced uvular fricative. The attested forms are (Y) *yéndeṣ*, (H, N) *yéniṣ,* PL. *yénaŋ*.

to be fossilized prefixes on nouns. Such prefixes (Greenberg's "stage-II articles") also appear sporadically in Caucasian and Yeniseian:

- **Basque** **be-hats* "toe, thumb", **be-laři* "ear", **be-lhaun* "knee", **bi-si* "life", etc.: cf. **Caucasian:** Ubykh *b-La* "eye", Tindi *b-etl':u* "stomach", *b-atl':i* "in the middle, between"; **Sino-Tibetan:** Tibetan *b-źin* "face", Garo *bi-kha* "liver", *bi-tši* "egg", etc.
- **Basque** **e-lhuī* "snow", **e-guī* "firewood", **e-puřdi* "rump", **i-sářdi* "sweat", **i-se-ba* "aunt", etc.: cf. **Yeniseian** **'ə-mas* "thorn", **'ï-Gïn* "tree-bark", **'ï-pVl* "aspen tree" (cf. Basque **e-s-pel* "box tree"), etc.
- **Basque** **o-dol* "blood", **o-śa-ba* "uncle", **u-s-ki* "anus", etc.: cf. **Caucasian:** Rutul *u-xun,* Tsakhur *wu-xun* "belly"; **Yeniseian** **'u-lVk* "bladder, bubble"
- **Basque** **a-ćo* "old woman", **a-bere* "cattle", **a-ho* "mouth", etc.: cf. **Yeniseian** **'a-lit* "woman", **'a-lVp* "tongue"
- **Caucasian:** Avar *mi-łir* "wing", *ma-xá* "abomasum", *me-géž* "beard": cf. **Sino-Tibetan:** Tibetan *m-ćhin* "liver", *m-gul, m-gur* "throat", *m-khal* "kidney", etc.
- **Caucasian:** Ubykh *t-χamə́* "skin, fur", *t-χwa* "ashes", etc.: cf. **Sino-Tibetan:** Tibetan *d-gun* "winter", *d-bu* "head", *d-kyil-ma* "center", etc.

Na-Dene, in isolation from the rest of DC, developed its own variant of a noun-classification system, which was marked on verb forms. For example, Eyak had a system in which all nouns fall into two categories: unclassified (including higher animals and some other nouns) and classified (other nouns falling into a large number of classes).[24]

In Eurasiatic languages we find quite a different state of affairs. The "core" Eurasiatic families, Uralic and Altaic, lack grammatical gender altogether, as does Kartvelian. "I consider it probable, since the other branches of Eurasiatic do not have grammatical gender, that it is an Indo-European innovation and that the first step was the distinction of common [animate] from neuter [inanimate] gender" (Greenberg 2000: 185). Vladimir Illich-Svitych (see Dybo 2004) postulated an opposition of two noun classes, "agent nouns" *vs.* "object nouns," in Proto-Nostratic. Bomhard (2004) essentially agrees in assuming an active/stative opposition in Proto-Nostratic, along with a complementary opposition of animate/inanimate.

In Afro-Asiatic it is fairly well established that the proto-language had a binary gender system (masculine/feminine), also derived, according to the Nostraticists, from the active/stative contrast.

3.2.2 *Personal pronouns*

In personal pronouns there is again a clear contrast between Eurasiatic and DC. It is well known that Eurasiatic is characterized by the pronoun stems **mi* "I, me" and

24. Krauss (1968).

**ti* "thou, thee". The paradigm is found in most of the Eurasiatic families, though Kartvelian, Turkic, and Tungusic have instead **mi/*si* or **bi/*si,* where **si* is identified with the IE second person singular verbal affix *–s(i).* The **mi /*ti* paradigm is so characteristic of Eurasiatic that some have called the family "Mitian" (Fleming 1987). The original paradigm of DC is still unclear, but the three centrally-located families, Caucasian, Burushaski, and Yeniseian, show traces of suppletive first and second person singular pronouns that can be schematized as follows (Table 1):[25]

Table 1. Suppletive personal pronouns in reconstructed Dene-Caucasian languages

	Caucasian	**Burushaski**	**Yeniseian**
1st person singular (1)	*nɨ́[26]	*a-[27]	*b-/*'ab-/
1st person singular (2)	*zō[28]	*dźa[29]	*ŋ[30] *'adz[31]
2nd person singular (1)	*ʁw V̄[32]	*gu-/*go-[33]	*kV-/*'Vk-[34]
2nd person singular (2)	*u̯ō[35]	*un[36]	*'aw/*'u[37]

These forms seem to indicate an underlying suppletion in proto-DC, where the first person singular had two different stems, something like **(ʾa)ŋV vs. *(ʾa)dzu,* and

---

25. This scheme owes much to the version offered by Blažek (1995: 109).

26. This stem is restricted to Lak and Dargwa: Lak *na,* Dargwa *nu* "I".

27. (H,N,Y) "first person singular prefix" in verbal forms and nouns ("my-"). Initial DC **ŋ>Ø* or *h* is regular in Burushaski.

28. Khinalug *zï*, Tsakhur *zu*, Chechen *so*, Avar *dun*, Kabardian *sa*, etc. "I"; Lak *t:u-* "I" (oblique base), etc.

29. Yasin *dźa,* Hunza, Nager *dźe* "I" (absolutive first person singular pronoun). (Commonly written *ja, je.*)

30. *ŋ* persists unchanged in Kott *–aŋ* (1st. PERS. SG. subject marker); according to S.A. Starostin PDC **ŋ > *m > b* in Ket *–ba-,-bo-*(1st PERS. SG. dative marker), Arin *bi-, be-*(possessive prefix "my"), etc.

31. Ket *āt,* Kott *ai,* Pumpokol *ad* "I", etc.

32. **ʁ* represents a voiced uvular fricative. PEC **ʁwV¯* "thou" > Chechen *ħo,* Dargwa *ħu,* Tsakhur *Gu* [ʁu], etc.

33. Second person singular prefix in verbal forms and nouns ("thy-").

34. Ket *ūk* "thy", *k-, ku-(*2nd PERS. SG. prefix), Yug *k-, ku-(*2nd PERS. SG. prefix), etc.

35. Kabardian *wa,* Khinalug *wɨ,* Tsakhur *wu,* Archi *un,* Avar *mun,* etc. "thou"; Chechen *vaj* "we" (inclusive)', etc.

36. In all dialects *un* (absolutive second person singular pronoun). Cf. Archi and Udi *un* id.

37. Ket *ū,* Yug *u,* Kott, Arin *au,* Pumpokol *ue* "thou".

the second person singular likewise had two different stems, **(ʾ)u)Gwu vs. *(ʾa)wu*). Remnants of these stem alternations persist to this day, for example (Table 2):

Table 2. Suppletive personal pronouns in modern Dene-Caucasian languages

	Stem 1	Stem 2
Caucasian	Khinalug ***wɨ*** "thou" (direct)	***oχ*** "thee" (dative)
	Tsakhur ***wu*** "thou"[38] (direct)	***jiʁ-*** "thee, thy" (genitive)
Burushaski	Hunza ***un*** "thou" (direct)	***góo*** "thee, thy" (genitive)
Yeniseian	Ket ***ū*** "thou"	***ūk*** "thy" (possessive)

As is totally natural subsequent simplifications and reorganizations took place in the DC daughter languages. For example, in Basque the first stem of each pair was selected for the direct form: **ni* "I/me" and **hi* "thou/thee". (The East Caucasian language Dargwa made the same selection: *nu* and *ħu,* respectively.) West Caucasian, on the other hand, generalized the second stem of each pair (e.g., Kabardian *sa* "I", *wa* "thou"). Sino-Tibetan generally retains **(ʾa)ŋV > *ŋa* for "I",[39] but **(ʾ)uKu > *Kwa* "thou" is restricted to Tibetan, Burmese, and Gurung, and another second-person stem **na-(ŋ)* (of uncertain relationship to PDC **(ʾa)wu-n-*) is more general. In Na-Dene Athabaskan has first-person **š-*, possibly related to PDC **(ʾa)dzu* (devoiced as in West Caucasian and Nakh). The apparent reflexes of PDC **(ʾ)uKu,* Athabaskan *χw-* and Tlingit *γi (yi)* are second-person plural. The usual Na-Dene second-person singular stem, Athabaskan **ŋən* and Haida *daŋ,* may be related to Sino-Tibetan **na-(ŋ).*

In sum, it is clear that the Dene-Caucasian first and second person singular pronouns are quite distinct in form from those of Eurasiatic.

3.2.3 *Interrogative stems*

When we compare the interrogative stems of the two macrofamilies we again find clear differences. The Nostraticists propose two primary interrogative stems for

---

38. Tsakhur has an alternate direct form *Gu,* apparently a merger or contamination of both stems. Possibly Ket *ū* "thou" and *ūk* "thy" also represent at least a partial merger.

39. Peter Norquest (p.c.) mentions the proposals of Sagart (1999), who argues that "the Old Chinese 1st person singular pronoun should be reconstructed as **la*, as it occurs almost exclusively during the Shang and Western Zhou periods. The pronoun **ŋa* does not appear until Eastern Zhou, and he argues that it is based on analogy with the early 1st person collective pronoun **ŋaj* (corresponding to 2nd singular **na* and 2nd collective **naj*). Moreover, he argues that Tibeto-Burman occurrences of 1st singular **ŋa* are borrowings from Chinese, and that original Tibeto-Burman 1st singular was actually **ka* (meaning, if you agree with his arguments, that an original Sino-Tibetan **ŋa* would be completely spurious and anachronistic)."

Eurasiatic, one of which, **k'V* (Bomhard's **k^{wh}a-/*k^{wh}ə-*, Greenberg's *k-*) may have originally been the animate interrogative. Illich-Svitych distinguished **ko* "who" from **mi* "what" (Dybo 2004: 118). Greenberg (2000: 217) agreed, finding that *"k-* is essentially personal, unlike *j-* … and *m-* … which are practically always non-personal." Some of the reflexes include Latin *qui-s* "who", *qui-d* "what", Hungarian *ki,* Komi *kin*, Tundra Yukaghir *kin,* Turkish *kim,* Mongolian *ken,* Aleut *kin* (the last six all "who"). Some of the reflexes of **mi* "what" include Finnish *mi-kä,* Hungarian *mi,* Old Korean *mï-* "what", Turkish *–mi* (interrogative particle), Chukchi *mi-kin* "who", etc. (Attestation in Indo-European is spotty, e.g., Hittite *mahhan* "how", Tocharian B *mä-k-su* "which, who".)[40]

In Dene-Caucasian there are some scattered traces of the interrogatives **kwi*[41] and **mV,*[42] but far more common are the stems we can characterize as **sV* and **nV.*

First, for **sV* we find in Basque the interrogative stem **se-,* in common Basque *ze*-r "what", *ze*-in "which" *ze*-la(n) "how", etc. In Caucasian the stem is very widespread, e.g., Ingush *se* "what", Akhwakh *s:un-* "who, what" (oblique), Dargwa *se* "what" Ubykh *sa* "what", etc. Burushaski the particle **-sa-/*-se* appears in interrogatives, e.g., Hunza bé-*sa*-n "what, which", bé-*se* "why".[43] In Sino-Tibetan **su* is widespread for "who" (Tibetan, Kiranti, Gurung *su,* Burmese *əsu,* Lepcha *šu*), and Lepcha has an interrogative particle *să-,* as in *să*-re "which, what", *să*-ba "where", *să*-lo "how", etc. Yeniseian **'as- /*sV-* is found in Ket *aśeś* "what", *aś-ka* "when", Kott *ašix* "what, how", *ši-na* "what", etc. In Na-Dene the **sV* stem is found in Haida gúú-*su* "what", gii-*s* "where", *us* (yes-or-no question marker), Tlingit dàà-*sá* "what", daa-*sá*-yu "what is that?", wàà-*sá* "how", gùù-*sá* "where", *sá* (interrogative and indefinite particle), Chipewyan *–sã´* (interrogative suffix), Navajo *–š, -šã'* (interrogative suffix), díí-*š* "this one?",[44] etc.

---

40. A third common interrogative stem is Illich-Svitych's **ja* = Bomhard's **'ay-, *'ya-* = Greenberg's *j-*. Starostin preferred to reconstruct **ŋiV* (Turkish *ne* "what", Korean *nu-gu* "who", Japanese *nani* "what", IE **yo-, *i-* > Sanskrit *ayám* "he", Greek ὁς ἥ ὅ "he, she, this," etc., Lithuanian *jìs* "he", German *er* "he", etc.).

41. **kwi* is only East Caucasian, and sparse even there: Avar *ki-b* "where", *ki-n* "how", Archi *kwi* "who", Khinalug *k-la* "who". (? Cf. Haida *gúu-s* "what", *gíi-s-duu* "who, which", *gi-nu* "where", etc.)

42. Chechen *mi-la* "who", Karata *heme,* Burushaski **men, *ámen* "who", Ket *bi-śśe* "who" (masc.), Kott *bi-li* "where", etc.

43. Per Bertil Tikkanen (p.c.) these *s*-elements in Burushaski are better viewed as original (inanimate) class markers.

44. The shushed *š* in Navajo (and Chipewyan *s* < **š*) recalls the *š* in some Caucasian words: Hunzib *šijo* "what", Udi *šu* "who", etc.; in Sino-Tibetan, Lepcha *šu* "what, which, who".

For the **nV* stem Basque has **no-*, as in common Basque *no*-r "who", *no*-n "where" (cf. Tabasaran *na'an* id.), *no*-iz "when", *no*-la "how". In Caucasian we find words such as Hunzib *hi-na* "how?", Tsez *neti* "when", *na* "where", Lezgi *ni* "who" (erg.), Tabasaran *na'an* "where", etc. In Burushaski the stem is restricted to the Yasin dialect: *áne* "where, wherever", *ánum* "whence", *ána* "whither, where (to)". Sino-Tibetan has Old Chinese **n̲āj* "how, what", Burmese *naṅh* (marker of special question), Jingpo gə-*naŋ,* gə-*niŋ* "where", etc. In Yeniseian the etyma are Ket *anet, ana* "who", Yug *anet* "who". (As far as I know, the **nV* interrogative is absent from Na-Dene.)

The following table (Table 3) summarizes some of the differences between the proto-languages of Eurasiatic and Dene-Caucasian:

Table 3. Differences between Eurasiatic and Dene-Caucasian

	Feature type	Eurasiatic	Dene-Caucasian
phonology	laterals	lateral resonants only	lateral resonants, fricatives, and affricates
	velar/postvelar	velar only	velar and uvular
morphology	noun class/gender	animate/inanimate	multiple noun classes
	pronouns (first and second person singular)	**1. *mi** **2. *ti (*si)**	**1. *(ʾa)ŋV / *(ʾa)dzu** **2. *(ʾu)Gwu / *(ʾa)wu**
	interrogative stems	***KV/*mi**	***sV/*nV**

4. Fruitfulness

Edward Sapir, in a 1921 letter to Alfred Kroeber, remarked on the fruitfulness of his "Sino-Dene" hypothesis in words that convey his excitement of discovery: "If the morphological and lexical accord which I find on every hand between Na-Dene and Indo-Chinese is "accidental", then every analogy on God's earth is an accident. It is all so powerfully cumulative and integrated that when you tumble to one point a lot of others fall into line."[45] This has been my experience with the DC hypothesis as well. Some examples of this fruitfulness are summarized below.

In the course of my investigation of a DC phonological feature, the lateral affricates, I have discovered a number of fascinating facts about the DC lexical system, one of which is the semantic relationship between "bottom" and "left (hand)", originally with a lateral affricate in both words, probably the glottal lateral affricate **tl'*. Note the Caucasian and Na-Dene examples below.

45. Golla (1984: 374). This quote was also cited by Ruhlen (1994b: 25).

4.1 **"bottom": Caucasian:** Bezhta *õtl'o,* Tindi *hinlu, etc.* (< PNC **Htl'ŏnŭ* ~ **Hnŏtl'ŭ*) **Na-Dene:** Haida (S) *s-tl'əŋ* ("bottom"), *s-tl'ə-χul* ("anus", i.e., "bottom" + "hole"); Navajo *–tl'ááh* ("bottom"), *-tl'àà'* ("buttocks"), Mattole, Hupa *–tl'a'* "buttocks, bottom", Chipewyan *–tl'á* "buttocks, tail of fish, hind part"

4.2 "left" (hand): Caucasian: Bezhta *õtl'a*-tl'as, Tindi *antl:i*-tl:a-s:a (< PEC **'VñtlV* + locative) **Na-Dene:** Haida (S) *s-tl'aanə* "left" (place noun); Navajo *–tl'àh* "left-handed", *tl'àá* "left-handed one", Mattole *tl'ehe'* "left-handed"

We know of semantic associations with "right" or "left" (hand) in other language families. (Cf. Sanskrit *dakṣiṇa-* "right, south", Hebrew ימין "right, south".)

Another semantic cluster, also involving lateral affricates, is seen in the following words involving the putting on, and taking off, of footwear. We can begin with a lexical comparison made many years ago by Hermann Berger, the late Burushaski specialist. Berger, who had studied under the noted Basque scholar Koldo Mitxelena (a.k.a. Luis Michelena) noticed that Basque and Burushaski had the same word for "barefoot". In the Bizkaian or farthest western dialect of Basque the word is *ortoz,* in Gipuzkoan it is *ortuz,* in the "French" Basque dialects of Lapurdi and Basse Navarre "barefoot" is *orthuts,* and in the northeastern Zuberoan dialect there are the verbs *erthǘst* "to have bare feet", *orthǘst* "to take off one's shoes", as well as the adjective *orthǘts* "barefoot".

Berger (1959, p. 27, note 35, and p. 33, note 57) compared the Basque words with a word found in the Yasin dialect of Burushaski (but not, apparently, in the other two dialects of Hunza and Nager): *holtáṣ* or *hultáṣ* "barefoot" (where *ṣ* represents a retroflex sibilant).[46] This adjective appears to be connected with a verb found in all dialects of Burushaski and can be reconstructed as **-ltá-* "to put on shoes or stockings",[47] as in *matúmiśo muzámuts u-**ltá**-m* "dressed in long black boots".[48] I have connected Burushaski **-ltá-,* **-lta-* with a number of words in other DC languages relating to footwear:

Caucasian: Proto-East Caucasian * *=ōmd̲lV* "to put on (trousers, shoes)" > Chamalal *=itl':n* "to put on (shoes, footwear, trousers)", Akhwakh *itl':e-l* "stocking", Avar (Andalal) *huntl':* "sock, stocking", Archi *=ubtla-s* "to put on (trousers)", etc.

46. Berger (1959) lists the words as *holtaṣ, hultaṣ.* Berger (1974) has only the form *hultáṣ,* with final stress.

47. In verb forms with a prefix the stem is realized as *–ltá-,* without a prefix as *tá-(*Berger 1974, 1998). There is a corresponding transitive verb **–lta-* "to put shoes or stockings on someone".

48. Berger (1998, vol. 2, p. 270).

> Sino-Tibetan: Proto-Sino-Tibetan **luamH* "a kind of shoe" > Old Chinese **loŋ'* "shoes for criminals with lopped toes", Tibetan *lham* "boot shoe"[49]
>
> Na-Dene: Navajo *–tlé, -tlèè'* "socks, stockings, leggings"

The very presence of a verb specifically for "putting on footwear" (as opposed to a combination of verb and noun) is notable, and furthermore the existence of a cognate with that meaning in several branches of DC seems to constitute another piece of powerful evidence for the reality of the DC macrofamily. The striking parallel between Basque **oɾ̄tútś* and Burushaski **hultáṣ* "barefoot" suggests that the suffix (Bsq *tś* = Bur *ṣ*) marks a "reversive" of the DC verb "to put on footwear". As an example of Sapir's "when you tumble to one point a lot of others fall into line" (mentioned above), I will point out that there is a recurrent phonological correspondence of Basque **r̄t* = Burushaski *–lt-* = Caucasian *–tl(')-* = Na-Dene *–tl(')-* in other etymologies. (See Bengtson 2004 for more examples.)

5. Inference to the best explanation

Finally we come to the last question enumerated in Section 2: the question of the best explanation for language diversification (and thus ethnogenesis) in northern Eurasia: are the superior hypotheses Eurasiatic and Dene-Caucasian, or are other hypotheses preferable, or even possible?

Here we can discuss the very nature of hypothesis and its role in science. The Honoree of this volume has this to say about the importance of hypothesis to historical linguistics:

> Our greatest strength, us long rangers, is hypothesis breeding or creation. Next to that is our better understanding of what science is all about. Historical linguistics is not a branch of mathematics or formal logic. As Greenberg [1995] has argued brilliantly … the concept of proof is misused by linguists – from a scientific standpoint. … Charles Darwin's theory has never been proven, has it? It just gets less and less likely to be false or more and more likely to be true. "Proof" is for algebra or courts of justice (Fleming 1994).

As shown by Peirce scientific hypotheses are created by abductive reasoning, the process of reasoning to the best explanation for a set of facts. Though abductive reasoning, like all modes of reasoning, is fallible, it remains the only logical process that

49. PST reconstruction modified to correspond to the latest findings (see http://ehl.santafe.edu or http://starling.rinet.ru).

can create new knowledge. Once the hypothesis has been formed, predictions may be inferred by deduction and tested by induction (Peirce 1878; Anttila 1972: 196–197). This is the scientific process by which, in Fleming's words, Darwin's theory "gets less and less likely to be false or more and more likely to be true."

As outlined briefly in Section 1, there has been a succession of hypotheses about the early linguistic diversification of northern Eurasia. The earliest hypotheses were rather vague and amorphous,[50] and as they evolved linguists began to realize the distinctiveness ("otherness" in the Whitehouse sense) of Basque and Caucasian and separated them from the remaining families. As the Eurasiatic hypothesis (at first called "Nostratic") continued to evolve linguists extended its bounds farther to the East, taking in obscure families of unwritten languages such as Yukaghir, Chukchi-Kamchatkan, and Eskimo-Aleut, based on the agreement of their morphology and lexicon with the diagnostic features of the already recognized Eurasiatic languages. Some linguists continued to include in their Nostratic the ancient Afro-Asiatic family, prestigious because of its ancient literature and cultural importance, while others, cognizant of the differences between relationship and taxonomy (Ruhlen 2001, 2005) increasingly viewed Afro-Asiatic as an ancient sister macrofamily coordinate with Eurasiatic.

It was not until the twentieth century that the beginnings of the Dene-Caucasian hypothesis began to take shape, probably because it involved mainly obscure, isolated, non-prestigious and non-literary languages (except for Chinese, Burmese, and Tibetan). The idea of Dene-Caucasian, with its widely separated and diverse membership, was even more audacious than the concept of Nostratic/Eurasiatic (Bengtson 1998), but a few linguists, driven by what they saw as diagnostic grammatical and lexical morphemes, persisted in developing the hypothesis, often in the face of withering criticism (Bengtson 1994, Ruhlen 2001). Though there were a few forerunners, we can date the beginnings of the Dene-Caucasian hypothesis to Trombetti and Sapir in the 1920's (Ruhlen 2001; Bengtson 1994). Now, some eight decades later, we think we have a clearer picture of both the membership of DC and its supporting evidence (morphology, lexicon, phonology).

Are the Eurasiatic and Dene-Caucasian hypotheses the best explanation for language diversification in northern Eurasia? As shown above (final table of Section 3) we can now describe many of the diagnostic differences between the two macrofamilies. Once we understand that the proper basis of genetic classification is taxonomy

50. Even a relatively recent hypothesis, the Vasco-Dene or Basque-Dennean of Morris Swadesh (Swadesh 1971; Fleming 1987, 1990b), was rather amorphous. In fairness to Swadesh, his life ended prematurely before he could fully refine the hypothesis, one that anticipated the "mega-super-phylum" (Fleming 1987: 214) that Fleming (1991) and others call "Borean" or "Boreal" (= northern).

(Ruhlen 2001, 2005), we can realize that the languages of northern Eurasia can be classified into two major groups based on the "specific, and distinct, grammatical and lexical morphemes that characterize each" (Ruhlen 2005: 348), in the same way that, for example, mammals can be classified into different families and genera on the basis of relatively few diagnostic traits.[51]

> Once the historical linguist has identified a diagnostic core of traits that cannot plausibly be coincidental, are too extensive to be explained by known language contact phenomena, and clearly do not exist together in any other languages, the hypothesis of genetic relatedness becomes the best explanation … Once this begins to happen, the hypothesis could be considered "proven," at least in practical terms (Vajda 1999: 88–90).

After many decades of abduction, deduction, and induction I believe we are now at this "proven in practical terms" stage with the Eurasiatic and Dene-Caucasian hypotheses. The essential diagnostic traits of each are fairly well understood (see Section 3), and there has been enough work done on both hypotheses to show that they are fruitful (Section 4).

Another indication of the high probability of a linguistic hypothesis is the presence of *shared irregularities* or suppletive paradigms, such as the Germanic pattern represented by English *good/better,* German *gut/besser,* Swedish *god/bättre,* etc. According to Greenberg (1987: 30) this type of paradigm "is obviously of enormous probative value," and Newman (2000: 264) agrees, going so far as to state that suppletions such as these "would permit one to postulate relatedness between these languages even in the absence of other evidence." Indeed, this kind of shared irregularity can be found in both Eurasiatic and Dene-Caucasian. Both Dolgopolsky (1984) and Greenberg (2000) agree that the first-person-singular suppletion characteristic of Indo-European **egHom* "I"/**me* "me" is found in far away Chukchi Kamchatkan (†South Kamchadal *kim* "I"/*ma* "me"), and both represent a pattern inherited from Proto-Eurasiatic. Shared irregularities in Dene-Caucasian first and second person pronouns have been discussed above (Section 3.2: Khinalug ***wï*** "thou"/***oχ*** "to thee", etc.).

For some perspective, it must be admitted that the evidence for Eurasiatic and Dene-Caucasian is comparable to that assembled so far for African (macro-)families (or phyla) such as Niger-Congo. Although "no comprehensive reconstruction has yet been done for the [Niger-Congo] phylum as a whole" its genetic unity is accepted by Africanists mainly on the basis of "noun class systems, verbal extensions and basic

---

51. Whitehouse (p.c.) cautions against an overextension of biological analogies, since "biology is physically constrained in a way that language is not, and therefore much more clear-cut. After all, a reptile species cannot "borrow" a placenta from its mammal neighbours in the way that word order or phonology can be modified by areal contact."

lexicon" (Williamson & Blench 2000: 11–12). And in spite of the fact that in "some families or branches [of Niger-Congo] the [noun class] system has been remodeled, or indeed lost with virtually no trace" (*Ibid.*, p. 12; cf. the analogy to Dene-Caucasian, described in Section 3.2), genetic unity is still accepted on the basis of the other two criteria. "Given the extensive similarity of sound and meaning in particular grammatical elements as well as in basic vocabulary, and the complex changes they have undergone, it is inconceivable that this could be due to chance and extremely unlikely that it is due to borrowing" (*Ibid.*, p. 13). Though Greenberg's methodology has been severely criticized (wrongly, in my opinion),[52] his four African phyla remain and are the basis of African historical linguistics as currently practiced (Fleming 2000–2001; Heine & Nurse 2000).

I propose in conclusion that the best explanation offered so far for linguistic diversification in northern Eurasia is the postulated existence of two old language families (macro families): Eurasiatic and Dene-Caucasian. No one has proposed a third possibility, at least for the northern half of this landmass,[53] and the nearest macro families, in southern Eurasia, appear to be Afro-Asiatic (to the southwest) and Austric (to the southeast). The intriguing possibility of yet another macro family, Indo-Pacific, on the Eurasian landmass has been posited by Whitehouse, et al. (2004).[54]

But the work is far from finished. All the linguists who work on these hypotheses would agree that extensive testing of etymologies (lexical hypotheses) through phonology, separating loanwords from inherited words, and refinement of protolanguage morphology must continue and will continue. Another thing that needs to be done is to define more precisely the membership of each macrofamily. It is embarrassing that each paleolinguist seems to have a different list of members of Eurasiatic and/or Dene-Caucasian, a fact that prevents some linguists from taking either hypothesis seriously. The most probable taxonomic status of, for example,

---

52. As Fleming (1987, 2000–2001) and Ruhlen (1987, 1994a, 2001, 2005) have repeatedly pointed out, Greenberg used the same taxonomic methods as were used by the discoverers of all accepted language families.

53. Linguists have assumed, for example, that the Saami (Lapps) once had their own language and adopted their present Uralic language, though what their original language was is not known. It is likewise assumed that other early Europeans had languages that were submerged by Indo-European. Nevertheless, there is no direct evidence of these submerged languages, or whether or not they were distinct from Eurasiatic and Dene-Caucasian.

54. It is highly probable that Indo-Pacific (and Australian) began as a Eurasian language family whose speakers migrated to the islands off the southern coast of Asia, and Australia-Tasmania. The IP languages that remained in southern Asia were eventually submerged by Austric, et al. except Kusunda (according to Whitehouse, et al.).

Etruscan, Sumerian, Gilyak (Nivkh), Kusunda, and Ainu should be more definitively determined.

In the spirit of the Honoree of this volume,

> One must agree with Michalove [1997] … that bold ventures are worth making because sometimes, when given a full hearing, they carry new truth, teach us something, show new paths. … Dolgopolsky and his students and Carleton T. Hodge inspired me and others to join in the hot pursuit of deep human prehistory. Ducking the stones thrown by the pettifoggers,[55] we find the whole exercise quite exciting and wonderfully fruitful. Is that not what science is supposed to be? (Fleming 1999: 425)

Acknowledgments

I am grateful for the helpful comments of colleagues, especially Václav Blažek, Allan Bomhard, Alice Beck Kehoe, Peter Norquest, and Paul Whitehouse. I am also thankful for the support of the Evolution of Human Languages Project through the Santa Fe Institute, and for guidance from the late Sergei A. Starostin.

And, of course, I am deeply beholden to Harold C. ("Hal") Fleming for his constant encouragement, support, and friendship over the past two decades.

Abbreviations

A	Alaskan (Haida dialect)
AN	Alto Navarro (Basque dialect) = High Navarrese
B	Bizkaia (Basque dialect)
BN	Bas Navarrais (Basque dialect) = Low Navarrese
C	common Basque
DC	Dene-Caucasian
EUR	Eurasiatic
G	Gipuzkoa (Basque dialect)
H	Hunza (Burushaski dialect)
L	Lapurdi (Basque dialect) = Labourdin
M	Masset (Haida dialect)
N	Nager (Burushaski dialect)

55. Pettifogger: "A lawyer whose methods are petty, underhanded, or disreputable: SHYSTER; one given to quibbling over trifles" (Webster).

R Roncalés (Basque dialect)
S Skidegate (Haida dialect)
Y Yasin (Burushaski dialect) = Werchikwar
Z Zuberoan (Basque dialect) = Souletin

References

Anttila, Raimo. 1972. *An Introduction to Historical and Comparative Linguistics.* New York: Macmillan.

Bengtson, John D. 1994. "Edward Sapir and the 'Sino-Dene' Hypothesis." *Anthropological Science* 102: 3. 207–230.

Bengtson, John D. 1997a. "The Riddle of Sumerian: A Dene-Caucasic Language?" *Mother Tongue* (Journal) 3. 63–74.

Bengtson, John D. 1997b. "Ein Vergleich von Buruschaski und Nordkaukasisch." *Georgica* 20. 88–94.

Bengtson, John D. 1998. "Caucasian and Sino-Tibetan: A Hypothesis of S.A. Starostin." *General Linguistics* 36: 1/2. 33–49.

Bengtson, John D. 2004. "Some features of Dene-Caucasian phonology (with special reference to Basque)." *Cahiers de l'Institut de Linguistique de Louvain* 30: 4.33–54.

Bengtson, John D. 2008. "Materials for a Comparative Grammar of the Dene-Caucasian (Sino-Caucasian) Languages." *Aspects of Comparative Linguistics*, v. 3., 45–118. Moscow: RSUH Publishers.

Bengtson, John D. & Merritt Ruhlen. 1994. "Global Etymologies." Ruhlen 1994a. 277–336.

Berger, Hermann. 1959. "Burushaski Lehnwörter in der Zigeunersprache." *Indo-Iranian Journal* 3: 1. 17–43.

Berger, Hermann. 1974. *Das Yasin-Burushaski (Werchikwar).* Wiesbaden: Harrassowitz.

Berger, Hermann. 1998. *Die Burushaski-Sprache von Hunza und Nager.* 3 vols. Wiesbaden: Harrassowitz.

Blažek, Václav. 1995. "Towards the Position of Basque: A reply to Trask's critique of the Dene-Caucasian Hypothesis." *Mother Tongue* (Journal) 1. 104–110.

Blažek, Václav. 1997. Review of Ruhlen (1987). *Mother Tongue* (Journal) 3. 159–183.

Blažek, Václav & John D. Bengtson. 1995. "Lexica Dene-Caucasica." *Central Asiatic Journal* 39: 1. 11–50; 39: 2. 161–164.

Bomhard, Allan R. 1996. *Indo-European and the Nostratic Hypothesis.* Charleston, SC: Signum.

Bomhard, Allan R. 2004. "Preliminary Thoughts on Nostratic Morphology." Hegedűs & Sidwell 2004. 27–50.

Dolgopolsky, A[ha]ron B. 1984. "On Personal Pronouns in the Nostratic Languages." *Linguistica et Philologica. Gedenkschrift für Björn Collinder,* Ed. by Otto Gschwantler, et al., 65–112. Vienna: Wilhelm Braumüller.

Dolgopolsky, A[ha]ron B. 1998. *The Nostratic Macrofamily and Linguistic Palaeontology.* Cambridge, U.K.: The McDonald Institute for Archaeological Research.

Dybo, Vladimir A. 2004. "On Illič-Svityč's Study 'Basic Features of the Protolanguage of the Nostratic Language Family.'" Hegedűs & Sidwell 2004. 115–119.

Fleming, Harold C. 1987. "Toward a Definitive Classification of the World's Languages" (Review

of Ruhlen 1987). *Diachronica* 4: 1/2. 159–223.
Fleming, Harold C. 1990a. "Reconstruction and Classification: Differences Decrease." *Mother Tongue* (Newsletter) 12. 5–6
Fleming, Harold C. 1990b. "Dene-Caucasic, Nostratic and Eurasiatic or Vasco-Dene?" *Mother Tongue* (Newsletter) 12. 6–12.
Fleming, Harold C. 1991. "A New Taxonomic Hypothesis Borean/Boralean." *Mother Tongue* (Newsletter) 14 (no pagination).
Fleming, Harold C. 1994. "A Mild Rejoinder to Lyle Campbell." *Mother Tongue* (Newsletter) 23. 70–72.
Fleming, Harold C. 1999. Review of Dolgopolsky 1998 & Hegedűs, et al. 1997. *Anthropological Linguistics* 41: 3. 417–426.
Fleming, Harold C. 2000–2001. "Joseph H. Greenberg: A Tribute and an Appraisal." *Mother Tongue* (Journal) 6. 9–27.
Golla, Victor, Ed. 1984. *The Sapir-Kroeber Correspondence: Letters Between Edward Sapir and A.L. Kroeber 1905–1925.* Berkeley: University of California Press.
Greenberg, Joseph H. 1987. *Language in the Americas.* Stanford, Calif.: Stanford University Press.
Greenberg, Joseph H. 1995. "The Concept of Proof in Genetic Linguistics." *Mother Tongue* (Journal) 1. 207–216.
Greenberg, Joseph H. 2000. *Indo-European and Its Closest Relatives: The Eurasiatic Language Family. Vol. 1 Grammar.* Stanford, Calif.: Stanford University Press.
Hegedűs, Irén, Peter A. Michalove & Alexis Manaster Ramer, eds. 1997. *Indo-European, Nostratic, and Beyond: Festschrift for Vitalij V. Shevoroshkin.* (Journal of Indo-European Studies, Monograph 22.) Washington, DC.: Institute for the Study of Man.
Hegedűs, Irén & Paul Sidwell, eds. 2004. *Nostratic Centennial Conference: the Pécs Papers.* Pécs: Lingua Franca Group.
Heine, Bernd & Derek Nurse. 2000. *African Languages: An Introduction.* Cambridge, UK: Cambridge University Press.
Krauss, Michael E. 1968. "Noun-Classification Systems in Athapaskan, Eyak, Tlingit and Haida Verbs." *International Journal of American Linguistics* 34. 194–203.
Michalove, Peter A. 1997. "Altaic Evidence for Clusters in Nostratic." Hegedűs, et al. 1997. 243–256.
Newman, Paul. 2000. "Comparative Linguistics." Heine & Nurse 2000. 259–271.
Nikolayev, Sergei L. & Sergei A. Starostin. 1994. *A North Caucasian Etymological Dictionary.* Moscow: Asterisk.
Peirce, Charles S.S. 1878. "Deduction, Induction, and Hypothesis." *Popular Science Monthly* 13.470–482. [Reprinted in *The Essential Peirce: Selected Philosophical Writings,* Ed. by N. Houser & C. Kloesel, vol. 1, 186–199. Bloomington/Indianapolis: Indiana University Press.]
Peiros, I. & Starostin, S.A. 1996. *A Comparative Vocabulary of Five Sino-Tibetan Languages.* 6 vols. Melbourne: University of Melbourne.
Renfrew, Colin & Daniel Nettle, eds. 1999. *Nostratic: Examining a Linguistic Macrofamily.* Cambridge, UK: McDonald Institute for Archaeological Research.
Ruhlen, Merritt. 1987. *A Guide to the World's Languages. Vol. 1: Classification.* Stanford, Calif.: Stanford University Press.
Ruhlen, Merritt. 1994a. *On the Origin of Languages*. Stanford, Calif.: Stanford University Press.
Ruhlen, Merritt. 1994b. "An Overview of Genetic Classification." Ruhlen 1994a. 9–38.
Ruhlen, Merritt. 1994c. "Amerind MALIQ'A 'Swallow, Throat' and Its Origin in the Old World." Ruhlen 1994a. 242–251.

Ruhlen, Merritt. 2001. "Taxonomic controversies in the twentieth century." *New Essays on the Origin of Language,* Ed. by J. Trabant and S. Ward, 197–214. Berlin/New York: Mouton de Gruyter.

Ruhlen, Merritt. 2005. "Taxonomy, Typology and Historical Linguistics." *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics,* Ed. by J.W. Minett & W. S-Y. Wang, 341–368. Hong Kong: City University of Hong Kong.

Sagart, Laurent. 1999. *The Roots of Old Chinese.* Amsterdam: John Benjamins.

Shevoroshkin, Vitaly, Ed. 1991. *Dene-Sino-Caucasian Languages.* Bochum: Brockmeyer.

Shevoroshkin, Vitaly. 1999. "Nostratic Languages: internal and external relationship." Renfrew & Nettle 1999. 75–91.

Shevoroshkin, Vitaly. 2003. "Salishan and North Caucasian." *Mother Tongue* (Journal) 8. 41–64.

Shevoroshkin, Vitaly. 2004. "Proto-Salishan and Proto-North-Caucasian Consonants: a Few Cognate Sets." Hegedűs & Sidwell 2004. 181-191.

Starostin, Sergei A. 1989. "Nostratic and Sino-Caucasian." *Explorations in Language Macrofamilies,* Ed. by. V. Shevoroshkin, 42–66. Bochum: Brockmeyer.

Starostin, Sergei A. 1999. "Subgrouping of Nostratic: comments on Aharon Dolgopolsky's *The Nostratic Macrofamily and Linguistic Palaeontology.*" Renfrew & Nettle 1999. 137–156.

Swadesh, Morris. 1971. *The Origin and Diversification of Language,* Ed. by J. Sherzer. Chicago/New York: Aldine. Atherton.

Vajda, Edward J. 1999. "What would be required to prove a genetic link between Basque and other Eurasian language families." *Mother Tongue* (Journal) 5. 87–91.

Wescott, Roger W. 1999. "Introduction to *Mother Tongue* V." *Mother Tongue* (Journal) 5.1.

Whitehouse, Paul. 2003. "Inclusion Versus Exclusion: The Problem of Negative Evidence." *Mother Tongue* (Journal) 8. 3–9.

Whitehouse, Paul, Timothy Usher, Merritt Ruhlen & William S.-Y. Wang. 2004. "Kusunda: An Indo-Pacific language in Nepal." *Proceedings of the National Academy of Sciences* 101: 15. 5692-5695.

Williamson, Kay & Roger Blench. 2000. "Niger-Congo." Heine & Nurse 2000. 11–42.

Slaying the dragon across Eurasia

Michael Witzel
Harvard University and President,
Association for the Study of Language in Prehistory

Myths relating the slaying of a large reptile by a hero or trickster deity appear in many mythologies across Eurasia and beyond, in Polynesia and the Americas. They are an important part of the creation myths. The killing of the monster liberates the dammed up waters as to make the world fertile and inhabitable for humans. Related is the connection between summer solstice and the marriage of the dragon slayer (or a hunter) with a local virgin, ultimately, the marriage of sun and moon, as found from Old India via China and Japan to the Kekchi Mayas. Such myths are traced back to a reconstructed Late Paleolithic mythology, called "Laurasian" that incorporates myths from the beginning of the world to its final destruction. Historical Comparative Mythology is to be added as another approach in our quest to understand early humans.

Keywords: dragon slaying; Eurasia; Americas; Summer solstice; marriage of Sun and Moon

1. Introduction

Myths[1] all over the globe exhibit certain similarities that have attracted the attention of modern scholars for some hundred and fifty years. In any attempt to explain the early history of anatomically modern humans, comparative mythology would be expected to add important aspects, however, research along these lines has hardly been

1. A brief definition of myth may run like this: Myths are highly regarded, more or less standardized, non-secular tales dealing with questions of the origin, nature and ultimate destiny of the world and its human beings, including those of their societies, rituals and festivals. There are many other formulations; however, a very comprehensive and useful definition has recently been given by W. van Binsbergen, Paper at the Conference "Myth and the disciplines" at Leiden, Dec.12, 2003, see: http://www.shikanda.net/ancient_models/leopard/leopardwww.htm, http://www.shikanda.net/topicalities/kyoto/kyoto.htm.

carried out so far. This is not surprising as records of myths go back only some 5,000 years and most have been recorded in much more recent times.

Whenever myths have indeed been compared on a larger scale it was based on preconceived notions of interpretation.[2] To be sure, myths have been studied for a long time,[3] ever since the times of the ancient Greeks, Indians and Chinese, and in the modern Occident for at least some 330 years.[4] However, such studies, comparative or otherwise, have not yet yielded a cogent system of relationships.[5] There are several ways to explain such wide-ranging similarities.[6]

However, earlier types of explanations of myth proposed so far[7] fail to address the central, but generally *unnoticed* problem: the comparability of *whole systems of myths*; in other words, to use a linguistic simile, the comparison of *whole grammar*s, not just of a particular word, form, declension/conjugation or syntactical feature. When actually comparing whole systems of myths it can be noticed – though not explored so far – that local mythologies, such as the Vedic Indian, Japanese, Icelandic or Maya ones, not only have similar contents (individual myths with similar motifs), but that these items are also *arranged* in similar fashion. In exploring this feature I do not compare randomly from mythologies all over the world, but only from those that

---

2. Such as the two most current ones of diffusion (Frobenius, Baumann, etc.), common underlying features of the human mind (Jung, Campbell; also, though more limited in geographical space, Lévi-Strauss's binary systems), etc. Diffusion entails that the similarities in widely distributed myths are due to a gradual dispersion of such motifs from a known or reconstructed center. The other current theory, of common universal traits of the human psyche, is based on C.G. Jung's psychology (followed by J. Campbell and others): certain motifs, or their composite parts (archetypes) are *universal* and can appear in dreams, visions and myths, and can re-emerge at any time.

3. See Doty 1986, 2000.

4. See Feldman & Richardson 1972; and cf. Matsumura 2005.

5. See however, my recent proposal of the Laurasian and Gondwana mythological regions in some previous papers: *Zinbun (*1990), *MT* VI, 2001, 45–62; EJVS 12 (2005); Out of Africa: The journey of the oldest tales of humankind. Conference Presentation at the *Conference on Generalized Sciences: Peaceful World and Enriching Lives*, Tokyo March 17, 2005; *ESCA* 7: 2006: 284–318; a longer version of the present paper, including a detailed analysis of aspects of the summer solstice, was given at Nanhua University, Taiwan, Oct. 2005. The proposals will be detailed my forthcoming book, *Origins* (working title).

6. There is a long list of interpretations of myth, from the Classical 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.

7. See W. Doty 1986, 2000.

follow *a certain narrational scheme.* Indeed, a fairly large number of them exhibit a *common story line,* that I call Laurasian, after the geological name of the early northern supercontinent.

This narrative scheme encompasses, in succession, the ultimate origins of the universe and the world, the subsequent generations of the gods, an age of semi-divine heroes, the emergence of humans, and the origins of (noble) 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.

The new approach, of historical comparative mythology, has been proposed earlier (*Mother Tongue* VI, 2001: 45–62).[8] It has recently, though unwittingly, been called "essentially romantic"[9] as it looks for, and points toward, a *common source*, that certainly "may no longer exist," as William Jones put it in 1786 with regard to the Proto-Indo-European parent language. Indeed, as pointed out in *MT* VI, this new approach, and the steps taken, are similar to the well tested methods of historical (and long range) linguistics.

In the present paper, a certain type of myth, the slaying of the dragon, will be explored in some detail. It belongs to the important series of creation myths that has been discussed elsewhere.[10]

---

8. See further Witzel 2005; Witzel, M. Out of Africa: The journey of the oldest tales of humankind, given at the *Conference on Generalized Sciences: Peaceful World and enriching Lives*, Tokyo March 17, 2005; forthcoming book, Origins (working title).

9. As W. Doniger chose to call it in *The New York Times Book Review* (July 14, 1991: 3, 26): "Given cultural convergences the theoretically possible explanations are: (a) diffusion, (b) derivation from a common source (c) derivation from structural characteristics of the human mind. [Ginsburg] rejects the idea of a common source because he rejects a model which is Romantic even before it is positivist: that of the genealogical tree." However it is precisely this model that has been successfully used by comparative historical linguistics, palaeontology, and – visible in popular accounts since the Fall of 1990 – in the very influential genetic studies (cf. Witzel 2001). Incidentally, in her review, Doniger had many of the facts in hand that would have allowed her to observe the opposition between Eurasian (Laurasian) and sub-Saharan African (Gondwana) mythology, but due to the engrained "path dependencies" of the psychological interpretation, from Freud onwards, she failed to draw the obvious conclusions discussed in this and earlier papers. Recent advances in human genetics lend additional support to this scenario; such results (especially the early, Paleolithic emigration from Africa along the coasts of the Indian/W. Pacific Oceans) will be dealt with separately (Witzel, *Origins:* working title) .

10. Witzel 2006.

2. The dragon

The dragon in the form of a giant lizard-like creature or as a giant snake enjoys a worldwide spread (S. Thompson, 1932–6: Motif B11).[11] This spread has usually been explained by diffusion (Smith 1919) or by archetypes (C.G. Jung), see above. However, Blust (2000: 520) provides a survey of its typical traits. He believes that dragon ideas "arose through processes of reasoning which do not differ essentially from those underlying modern scientific explanations." He traces this back, ultimately, to the observation of rainbows.[12] Though this natural phenomenon has been interpreted as a giant snake by the peoples of many areas (Sub-Saharan Africa, South America, Australia), it will be seen that the reptilian[13] form that (mostly) possesses legs and that is found in large areas of Eurasia, is of a different nature. In the perspective of Laurasian mythology, its appearance in "mythological history" is tied to a particular stage in creation myths.

After the initial creation of the universe, of the earth, and of light and sunshine,[14] the new earth is not yet ready for living beings. It has to receive moisture, whether (sweet) water or the blood of a primordial creature. In many traditions, it is the latter. Only after the earth has been fertilized by the Dragon's blood it can support life.

---

11. In some detail: B 11: data from Europe, India, Korea; B11.1. Origin of the dragon (e.g., from worm); B11.2. Form of dragon; B11.2.1. Dragon as compound animal (Europe, China, Egypt).; B11.2.1.1. Dragon as modified serpent (Amerindian, Japanese, Indian) or fish (China); B11.2.3. Many-headed dragon (Europe, Iran, Indian; Japan, S. America: Araucanian; Africa: Fulah; B11.2.11. Fire-breathing dragon; B11.6. Deeds of dragons; B11.7. Dragon as rain-spirit (China) or bringing water (India, Japan); B16.6. Giant devastating serpent, cf. R111.1.3. Rescue of princess (maiden) from dragon; B11.11. Fight with dragon (Europe, Egypt, Iran, China, Japan; Africa (Fang); (cf. A531. Culture hero overcomes monsters); cf. also B15.1.2.2.- B11.2.3.1/B15.1.2.8.1. Three-to Nine-headed dragons. B11.7. Dragon as rain-spirit; A1111. Impounded water: Water is kept by monster so that mankind cannot use it. A hero defeats the monster and releases the water (worldwide). Also: A876. Midgard Serpent; B91.1. Naga. Serpent demon; B91.2. Plumed serpent; A139.3. Dragon god; A132.1. Snake-god. – It has to be noted that S. Thompson's data are confined, by and large, to Eurasia and the Americas.

12. Barber & Barber (2004: 232–244), too, explain the dragon, somewhat simplistically, as derived from certain features observed in nature, however, mostly those of northwestern Europe. They do not include a worldwide survey such as done by Blust 2000.

13. It has scales in all areas surveyed by Blust 2000: 520 (Europe, Near East and Egypt, India, Far East, Mesoamerica, North America); however, apparently not in S. America (but note a many-headed dragon with the Araucanians, Thompson 1932–6: Motif B11.2.3.1); see further below, at the end.

14. See M. Witzel 2005: *Vala and Iwato*.

Frequently, (Father) Heaven and (Mother) Earth are the primordial deities. Their children are, e.g., the Greek Titans, Indian Asuras or Jpn. *Kuni.no Kami* ("mundane gods"). The latter's younger and victorious cousins are the Olympian gods, the Indian Devas or the Jpn. *Ama.no Kami* ("heavenly gods"); their older cousins are regarded as enemies or monsters who have to be slain or at least be subdued temporarily.

Most prominent among such fights is the slaying of the primordial Dragon by the Great Hero, a descendant of Father Heaven. In the Vedic texts of early India, it is the great god Indra who kills the three-headed reptile, just like his Iranian counterpart Thraētaona (Avesta texts) kills the three-headed dragon, or as their distant equivalent in old Japan, the god Susa.no Wo (Kojiki, Nihon Shoki), kills the "eight-forked" dragon, *Yamata.no Orochi*.[15]

The same is echoed at the other end of Eurasia. In England, it is Beowulf, in the Icelandic Edda it is Sigurd (the Siegfried of Wagner's opera and of the medieval Nibelungen Epic) who perform the heroic feat of slaying the "worm."[16] We may also compare Herakles' killing of the Hydra of Lerna. Herakles is the mortal son of the king of the Olympian gods, Zeus. In his famous 12 deeds, Herakles not only kills various monsters but he also finds the cows, or dawns (§ 5): in other words, he acts just like the Vedic Indra.

Closely related to the latter is the Slavic myth of the hero's fight with *Veles*. His name reflects the Avestan *Vara*, Vedic *Vala,* both of which are terms for an underground fortress or cave that contains the "cows" (dawns), the sun and moon as well as the goods desired by humans. (In the Nuristani myths of NE Afghanistan, it is called "the house near heaven"). The dichotomy between Slav. *Veles* (Lithuanian *Vêlinas, Vélnias;* Latvian *Vél*s) and *Perun*' (Lith. *Perkűnas,* Puhvel 1987: 226 sq) is still seen in place names, even in such relatively late Slavicized areas as Dalmatia (Katičić 2001).

The Indo-European myths have recently been studied by C. Watkins (1995)[17] and to some extent by Barber and Barber (2004: 232–244).[18]

---

15. In Japan (and to some extent in Polynesia), 8 is the preferred number in enumerative systems, while it is 9 in northern (shamanic) Eurasia and China, and 7 in the Greater Near East: interesting regional features that need further exploration.

16. As the dragon is called in Germanic languages. We may also compare Thor's and Tyr's killing of the giants.

17. Many Indo-European comparisons (especially Slavic ones) have been made by Ivanov and Toporov 1970, 1974.

18. They find simple explanations, based on the experience of nature. However, they mostly deal with northwestern Indo-European and some Near Eastern dragons but they neglect Indian, Chinese, Japanese and older pre-Occidental ones, such as the well-known chain of transmission of mythemes from Persia > Rome/Byzantium > Germanic tribes (Bächtold-Stäubli 1987: 364–367. *s.v. Drache*). The only exception is their study of some astronomical features in their

Further afield, in ancient Egyptian myth, the victorious sun (Re) each night slays the dragon of the deep (Apophis, "with a knife on his head"), when it passes underground back towards the east as to rise again. Even Apophis' bones are destroyed; there is total destruction, and not even his shadow is left. In ritual, too, Apophis is burnt daily *in effigie* at dawn and dusk. In Mesopotamia, Marduk's killing of Tiamat is a related theme. The earliest Chinese mythology has the "black dragon" killed; the dragon was not yet regarded then as a beneficial being, as later on.

There are even echoes as distant as in Hawai'i (*mo'o* monsters), while the myth as such seems absent, *prima facie,* in the Americas. On closer look, however, there are a number of myths that speak of killing various types of ogres, even outside the Na-Dene speaking Athapaskan tribes (Bierhorst 1986: 68) that are more closely related to Siberia. Examples are found among the southern California Chumash (Bierhorst 1986: 94), and as far as the most distant South American tribes, such as those of the Gran Chaco (Wilbert & Simoneau 1987: 703, 729)[19] and of Tierra del Fuego (Campbell 1988: I 2: 256; Wilbert 1975: 39–43; Gusinde 1931: 593–595, 597–599).[20]

3. Japan, India, and Iran

To facilitate a closer comparison, individual mythologies are investigated, to begin with, the oldest Japanese texts, Kojiki and Nihon Shoki (712/720 CE). According to them, the dragon *Yamata.no orochi* lives on the river *Hi* in Izumo,[21] the land assigned to Susa.no Wo, originally the lord of the Ocean. He is the son of the primordial parent deities Izanagi and Izanami. Nihon Shoki 1.51 (Aston 1972) says that the dragon in the land of Izumo, on the Hi river, "had an eight-forked head and

chapter 16. (Barber & Barber 2004: 177–217): The Dragon (*Draco*), the constellation between the two Bears (*Ursa*) was killed or driven off, and center of the world and of the night time sky (north pole) shifted (close) to our present Pole Star, *Polaris*. Cf. Witzel 1984.

19. Monsters: Motif A53: Wilbert & Simoneau no. s 97, 169, 170, killing the vulture, which has many other correspondences, e.g., to Motif B11.7.1+, B16.5.1.2, etc. (Wilbert & Simoneau 1987: 703); cf. also s.v. *snake* (Wilbert & Simoneau 1987: 729, B242+); see also Sullivan (1988: 75): primordial monsters killed (Desana), fish monster (Warao) p. 122, etc.; see index *s.v. monsters*.

20. Note also the seven-headed dragon with the Araucanians, Thompson 1932–6: Motif B11.2.3.

21. Or at Nihon Shoki 1.54: on the upper river Ye in Aki (Hiroshima Prefecture). Note that at Nihon Shoki 1.56, Susa.no Wo with his son Iso-takeru ("fifty courageous") went down to Shiragi (Korea), at Soshimori (*mori* = Kor. *moi* "mountain", Aston); he says "I will not dwell in this land", takes a clay (!) boat and crosses over to Mt. Tori-kamu.no Take at the upper Hi river. Then Susa.no Wo dwelt on Mt. Kuma-nari (Mt. Kumano in Izumo? near Suga) and finally went to Hades. (Kumanari: I p. 59, 232; Kōkuri = old name *Koma*, 367).

eight-forked tail; his eyes were red like the winter cherry; and on his back firs and cypresses were growing.[22] As it crawled it extended over a space of eight hills and eight valleys,"[23] with the typical Jpn. stress of the number eight. Susa.no Wo gets the dragon drunk with Sake, and cuts off one head after another.[24] Tearing him apart, he finds a sword (*kusa-nagi.no tsurugi*) in the dragon's tail which is to become important later on in Japanese myth (and as the sword of the Emperor). The dragon's spilled blood makes earth fertile. Susa.no Wo[25] marries the virgin Kushi-nada Hime whom he had rescued from the dragon, and finally enters the Netherworld (*toko-yo. no kuni*) via Cape Kumano (Nihon Shoki 1.60)

The old Iranian and Vedic Indian myth of slaying the dragon is of Indo-European origin, but it has undergone some local influences, especially in the shared Central Asian homelands of both peoples (Witzel 2000, 2003, 2004). The dragon is the primordial guardian of productive forces or of riches, and the divine Vedic hero Indra or the Iranian heroes Thraētaona (Avesta: Yašt 5. 33–35, Yasna 9. 7–8) or Kərəsāspa (Yt 19. 38–40, Y 9. 11) are his slayers.

It is one of Indra's main deeds to overcome Vṛtra, which originally meant just "resistance" (Benveniste-Renou 1934). He was imagined in IIr. tradition as a dragon or as a giant snake, lying on the primordial mountain or in the ocean. Its dragon form is found in IIr. as **aj'hi*, Old Iranian *aži,* Vedic *ahi* "dragon", a three-headed (*tri-śīrṣan/ θri-kamərəδa*) reptile monster. In Vedic he also appears as the three-headed Viśvarūpa, son of a primordial deity, Tvaṣṭṛ,[26] who is the adoptive father of Indra. When Indra kills the dragon Viśvarūpa, he thus kills his "cousin" (or due to "adoption" by Tvaṣṭṛ, even his step-brother), which clearly reflects the common IIr dichotomy between the *Deva* and *Asura* deities discussed above. However, there is archaeological evidence for the dragon from Southern Central Asia, an area where the speakers of pre-Vedic and and pre-Avestan have passed through (Witzel 2003, 2004).

There are many representations (Francfort 1994) of the dragon in the Bactria-Margiana Archeological Complex (BMAC), an early South-Central Asian Bronze age culture (2400–1600 BCE), an area that the ancestors of the speakers of Old Iranian and Vedic have passed through or even stayed for a while, and where they were deeply

---

22. Similar description of the 3-headed monster is found in the Avesta.

23. Aston 1972: 55sqq.

24. Cf. myth of Perseus and his killing of the Gorgon Medusa, see Graves 1955, vol. I: 238.

25. His child is Oho-na-muchi.no Kami, Oho-na-muji, Oho-na-mochi, etc (cf. Kojiki 1.10, Philippi 1968: 67), whose beneficial acts include that he "established this sub-celestial world, and created medicines … and controlled calamities of birds, beasts creeping things …" (Nihon Shoki 1. 59).

26. Cf. Avesta, Yašt 19.18 *θwōrəštar* as "creator" of Ahura Mazda's creation, cf. Yasna 29.6, Oberlies 2000: 370.

influenced in mythology and ritual (Lubotsky 2001; Witzel 2003, 2004, 2006) (Interaction between the BMAC and steppe peoples is now clearly visible: the BMAC has certain steppe influences, in pottery etc., and the opposite direction of influence is sometimes assumed for the Arkhaim/Sintashta culture[27] in the Urals area.) In the BMAC, the dragon mainly appears as an ugly, scaled, human-headed man standing with a water vessel in one arm (Francfort 1994). In most IIr. descriptions, however, the dragon is seen not in human form but as a giant reptile, killed by the heroes Thraētaona (Yašt 5. 33–35, Yasna 9. 7–8) or Kərəsāspa (Yašt 19. 38–40, Yasna 9. 11), who was resting and cooking on the beast. (cf. Oberlies 2000: 371 sq.). Slight differences of IIr. myth in the Avesta and the Veda must seen within the context of the Avesta that represents the local successor of the BMAC culture. We would then have, in IIr., these epithets of an old Dragon Slayer god (Witzel 2004):

Indo-Iranian		
**indra vṛtraghan-*	::	**aj ʰi (*yaz) vṛtra*
"strong slayer of resistance"	::	"dragon, the resistance"
Vedic /Old IA		
Indra	::	Vṛtra, ahi, (*tri-śīrṣan Viśvarūpa RV 10.8.9; 2.11.19, = *tri-śīrṣan trikakúd krimi* AV 5.23.9) Śuṣṇa, Cumuri (local)
Avestan		
(Indra), Vərəθragna	::	aži, (Y 9.8 *ažīm dahākəm θri-kamərəδam xšašašīm*)
(Kərəsāspa,[28]	::	yellowish monster, exuding yellow poison;
cooking a meal in metal pot at noon, Yt 19.38–41)	::	Gaṇdarəβa with yellow heel
(Ātar, son of Ahura Mazdā Yt 19.47)	::	aži Dahāka, θri-kamərəδa
(Tištriia Yt 8.13–23: in human, cattle, horse form)	::	Daēuua Apaoša, ka-mərəδa, black, bald horse

27. For the archeological background of contact between the steppe cultures (such as that of the Indo-Iranians) and the BMAC see Hiebert, Shishlina & Hiebert 1998 and Witzel 2003: 48sqq.

28. Note that his name "having emaciated horses" (Ved. Kṛṣāśva, cf. *kṛśagu, kṛśapaśu*) reflects the situation before the release of the waters; the name would fit Tištriia better. – Note also the stress in Zoroastrian tradition on the miserable situation (cf. Y 51.12) of Zarathustra ("having old camels"?) before he succeeded in gaining some followers.

However, the reptile also appears, with local Indian and Hindukush adaptations, as a giant cobra (*vyaṃsa,* Schmidt 1963) and it is in this area that an overlap with the snake form (*Nāga*) emerges. Even then, these northwestern Nāgas (found in Dardic and Nuristani speaking populations) are, to this day,[29] guardians of water in the form of ice and snow, unlike their tropical Indian forms that are linked to the monsoon rains (Witzel 2004, and forthc.)

In the BMAC, however, the dragon appears as scaled *anthropomorphic* demon of draught who fights the eagle faced hero (Francfort 1994).[30] The Eurasian motifs have evolved into a typical, local variety which has representations in its art of the motifs of the primordial dragon guarding and inhibiting the waters, the dragon-slaying hero, and a divine eagle.

Oxus religion :: *Indo-Iranian religons*	
Goddess	anthropomorphic DRAGON
fertility	of drought;
vegetation	then releases waters
Anāhitā?/ Sarasvatī/Rasā	*Aži/Ahi / "*Vərəθra"/Vṛtra*
Aditi, Dezālik	*Apaoša* (Forssman 1968)
	3-headed (IIr.) : θri-kamərəδa/
	tri-śīrṣan Viśvarūpa
Vṛtra > cobra snake in India : Vyaṃsa	
	His children:
combined forms, anthropomorphic	
lion/snake ~ *Sēnmurv?* (Schmidt 1980)	
	FIGHTS WITH THE HERO:
	eagle faced[31]
	(*Circaetus Gallicus* "snake eagle", Avest.
	saēna?),Eagle flies in Winter over the
	Hindukush (*upairi saēna, upari śyena*);
	catches and eats snakes;
	Hero in human form:*Vərəθragan/*
	Vṛtrahan, Iṇdara/Indra

29. The main river of Nuristan, Lu Nang, is derived from *Deva Nāga; Kashmirian Nāgas are so described in the local texts Nīlamata Purāṇa (8th cent. CE?) and Rājataraṅginī (1151 CE).

30. Various Old Indo-Aryan and Old Iranian data have been added (in italics) for the sake of comparison.

31. For a detailed example, see the reproduction in *Afghanistan* 2002: 204, of the eagle-faced hero found on a bronze axe, from Daulatab near Balkh, of c. 2000 BCE.

The old Indo-European myth of dragon slaying has been adjusted in the Avesta under the influence of the BMAC and its successor cultures. There is both the killing of the dragon but also Tištriia's fight with the demon of drought, Apaoša, and the generation of clouds and rain, reflecting what Francfort has reconstructed for the BMAC belief system. Even in the Ṛgveda, Indra is not just the dragon slayer but is also closely connected with the release of the waters. The Ṛgvedic giant cobra, *vyaṃsa,* surrounds the waters and must be killed (at least temporarily) to let them flow. This is more of an Afghanistan and Indus myth (Falk 1997) than a monsoon myth (Vajracharya 1997). In Central Asia, Afghanistan and the Panjab, the penned up waters, encapsulated by (the **Nāgas* of) snow and ice, are released by the snow melt, resulting in the late spring/summer floods so prominent in the Avestan and Ṛgvedic texts (Falk 1997).

The Indo-Iranian myth, however, lacks the episode of freeing a young woman from the clutches of the dragon, a motif that is found in later Iranian texts and that has spread from there to Armenia (myths of Mher),[32] the Caucasus and Europe, mostly in the form of the medieval Christian legend of St. George.[33] The relationships between the dragon and the heroes can be summarized for the Germanic, Indo-Iranian, and Japanese areas as follows.

Sigurd/Siegfried Beowulf	**Indra**	**Susa.no Wo**
dragon/ "worm"/ midgard snake	dragon *ahi/aži/* snake/Nāga is slain > releases water	dragon *yamata.no orochi* is slain > fertile land
Met, etc. going berserk	Soma invigorates Indra	Sake is given to dragon, gets drunk, is killed

Just as in the Vala/Iwato myth (Witzel 2005), the dragon myth shows a close relationship between old Japanese and early Central Asian myth (c. 2000 BCE), represented by the Vedic and Old Iranian traditions. Such links are in need of further exploration. Attention so far has been focused on the relationship between Japanese and Scythian/Greek myths.

32. See the early medieval Armenian myth/epic of *David of Sassoon*.

33. For a detailed background of this tale, see *Handbuch* (Bächtold-Stäubli 1930sqq. 1987) vol. 3: 647, and cf. Barber and Barber 2004. The earlier Byzantine tale of St. George killing the Dragon is attested in Germany and in England only since the 12th century.

At the other end of Eurasia, in ancient Greece, the motif is first found in the Homeric hymn 3.179 ff., where the sun deity, Phoibos Apollo, kills a female dragon at his temple of Pytho, at Crisa, below the Parnassos mountain.

> … Apollo … with his strong bow, the son of Zeus killed the bloated, great-she-dragon, … cruel Typhaon, … a plague among men … until the lord Apollo, who deals death from afar, shot a strong arrow at her. Then she, rent with bitter pangs, lay drawing great gasps of breath and rolling about that place… and so she left her life, breathing forth in blood. The Phoebus Apollo boasted over her: "Now rot here upon the soil that feeds men!" … and darkness covered her eyes.
> (Evelyn-White 1914: 351)

In this version of the myth, however, nothing is said about fertilizing the earth or providing water for it. (As in several other Greek myths, early Near Eastern influence may be seen; cf. §4). We can also compare the myth of Kadmos and the dragon.

> Kadmos founded castle of Kadmeia, the later Thebes. He killed a dragon, descended from Ares, with stones. He broke off the teeth of the monster and sowed them into the earth. Immediately, fully armed men arose from it, the ancestors of the Theban nobility. After an eight-year penance for having killed the dragon, Kadmos was married to Harmonia, daughter of Ares and Aphrodite; all gods attended. (In his old age both emigrated to Illyria where they turned into snakes and finally were admitted to Elysion). (see Graves 1995: 195–6; cf. 198–200).

4. Eurasia

Still older is the version preserved in the Mesopotamian text *Enuma Elish*, (tablet IV) which was recited at New Year. The deities elect the god Marduk as their leader and tell him:

> "Go, and cut off the life of Tiamat!"
> He fashioned a bow, designated it as his weapon,
> Feathered the arrow, set in the string.
> He lifted up a mace and carried it in his right hand,
> Slung the bow and quiver at his side,
> Put lightening in front of him,
> His body was filled with an ever-blazing flame.
> He made a net to encircle Tiamat with it, marshaled the four winds
> so that no part of her could escape …
> And set his face towards Tiamat who raged out of control.
> In his lips he gripped a spell,
> In his hand he grasped a herb to counter poison …
> The lord spread his net and encircled her …
> He shot an arrow which pierced her belly,
> split her down the middle and slit her heart,

vanquished her and extinguished her life.
He threw down the corpse and stood on top of her…
The Lord trampled the lower part of Tiamat.
With his unsparing mace smashed her skull,
Severed the arteries of her blood,
And made the North wind carry it off as good news. (Dalley 1989: 249 sqq.)

The story continues, in the fashion of the Ymir-Puruṣa-Pangu myth[34] to explain how the world was fashioned out of her bones.

In China, a dragon myth belongs to the oldest strata of local mythology. Nüwa (Nügua),[35] the second of the primordial "emperors," accomplishes the work of dragon slaying: in the beginning, the earth was still in chaos, some heroes must put it in order.

> The 4 extremes and the nine provinces were dislocated … The heaven did not cover earth completely … Fire transgressed everywhere without being mastered, water accumulated without being dispersed. Beasts devoured men, rapacious birds took away the old and weak. Nügua purified the fire of the stones of all colors, killed the black dragon, … accumulated the ashes of reeds to stop the overflowing waters … She cut the feet of the grand tortoise in order to fix the 4 extremes (quarters of the sky) … Then men could live on earth. (*Huainan zi*)[36]

Here the topic of establishing the oikumene is most clearly expressed and killing the dragon is one of its requirements.[37] Another version has, for the first time, a peaceful, beneficial dragon as habitually found in later Chinese myth. (Other texts see it as an uprising, Liezi/Liehtzu, ch.5: poem Tianwen/T'ang-wen).[38] Then, there is the myth of Kung Kung,[39]

---

34. Dissection of the primordial giant in Iceland, India, China (Yang & An 2005: 176sqq); variations are found elsewhere.

35. On her, see Yang & An 2005: 170–176; Mathieu 1989: 40 and especially p. 73 sq. In S.E. Chin. myth she escaped the great flood in a calabash. Nüwa is one of the three sovereigns of primordial age (Yang & An 2005: 170sqq), usually feminine, and associated with Fuxi, her brother, later her husband, in medieval, Tang texts. In Han time, she has a human head and a serpent body surrounding that of Fuxi. She created the human race, invented the flute.

36. Yang and An 2005: 172, Birell 1993: 69–72, 97–98, 146. For the tortoise, cf. other Eurasian and N. American myths; the black dragon = excess water; see Huainan zi, ch. 1 p. 3b.

37. Other texts see it as an uprising, Liehtzu, ch. 5: T'ang-wen.

38. E.W. Lai 1995; Yang & An 2005: 94–95.

39. Yang & An 2005: 124–126.

> Gong Gong [Kung Kung, Kanghui] extended the flood for 22 years… His son Yu emerged in the form of a horned dragon. Gun's body also transformed into a dragon at that time and thenceforth lived quietly in the deeps… Yu led other gods to drive away Gong Gong, distributed the Growing Soil to remove most of the flood, and led the people to fashion rivers from Ying's tracks and thus channel the remaining floodwaters to the sea. (Huainanzi, ch. 6: Lanming)

As to other early Chinese dragon-slayer myths, Lai (1995) focuses on the legendary Xia/Hsia mid-dynastic anthropogonic figure of Emperor K'ung-chia. He sees the Archer Yi of the East Coastal Region as belonging to the historic Shang, whose totem is the sun-bird. The prehistoric/legendary Hsia is in the Center and its totem is the snake-fish Dragon complex.

Southern China is home to a large number of Austric peoples. In one of their myths, coming from Sichuan (Szechwan), the ancient land Ba (Pa).[40]

> The Pa serpent is said to have a black body and a green head. It is so gigantic and greedy that it could swallow an elephant whole. Downstream east lay the Grotto Court Lake, and the Pa serpent also lurked in the waters there and did harm to many fishermen.
>
> Archer I, the hero of the I people in the east, killed this Pa serpent in a big battle. There is a small hill by the side of Lake Grotto Court that is called the Pa Mound. It is located at the southwest of Yueh-yang, Hunan province. It is where the bones of this gigantic Pa serpent were supposed to have been piled up after Archer I had killed the monster.[41]

In a late, Muslim version from Yunnan, the Drought-chaser turns into a dragon:

> During a long drought, the Drought-Chaser went to the sea and arrived at a crystal palace of the Sea Dragon, unlocked a thorny gate with a dragon-patterned board and entered. He found the Sea Dragon asleep, holding his rain pearl in his mouth. The Drought-Chaser grabbed the pearl, put in his own mouth, and ran as fast as he could retrace his footsteps. At that juncture, the Sea Dragon woke up. Discovering the pearl gone, he immediately gave chase. The Drought-Chaser swallowed the rain pearl and used the dragon-patterned board as a weapon. He struck the Sea Dragon's head with it and brought this monster down. Suddenly though, the Drought-Chaser himself turned into a dragon. He flew out of the crystal palace on wings, and brought down a torrential downpour that helped to end the drought.

---

40. The sign for Pa is a picture of a snake.

41. E.W. Lai, personal communication (Feb. 29, 2004).

The dragon here has two forms: asleep under the sea (winter) it withholds the subterranean water; awake to take to the air (spring) it brings down the rain from the clouds. The dragon board is the Son of Heaven's insignia ordering the spirits of nature about.[42]

Finally, in Polynesia where we do not expect any dragons (Hawai'i has no snakes or Komodo dragons), we still hear of them in the form of large lizard gods (Haw. *mo'o*, Maori *mokomoko, mokoroa,* Tregear 1891: 249), who also appear in many other, smaller shapes. They are prominent in the creation story, which seems superficially influenced by Christian motives (Fornander 1969). However, the very similar Maori version (Tregear 1891: 57) has old verse lines mentioning them. In this myth, we find a "fallen chief," there is the lying lizard Ilioha at the tree with the forbidden fruit of Kane (Maori *Tane*). In Fornander's Hawai'ian version, the first man (Haw. *Kumu-honua*, Maori *Ko-honua*) is formed out of earth. The Gods give him a garden in "the land that moved off," with pig, dog, *mo'o* of many sorts, and a *tapu* (taboo) tree.[43]

Comparison: The Eurasian dragon fight

Egypt	Mesopot.	Greece	India	Japan	China
Seth	**Marduk**	**Python**	**Indra**	**Susa.no Wo**	**Nügua**
god of thunder attacks dragon of the deep, kills and dismembers him each night (Apophis)	⟨new year myth!⟩ attacks Tiamat & monsters: dismembered	Kadmos & the dragon	dragon *ahi/ aži* slain, dismembered	Yamata.no orochi slain, dismembered	Black Dragon killed
gets drunk by red beer			Soma invigorates Indra	Sake is given to dragon drunk and killed	
			[Iran/Rome: St. George/ virgin saved]	virgin Kushinada Hime saved	

42. Interpretation by E.W. Lai, pers. comm. (February 29, 2004).

43. Cf. also the African version of the myth, with the Bassari in Togo (L. Frobenius 1924: 75–76, see Campbell 1988: 1. 1: 14).

5. Dragon and summer solstice

The appearance of the virgin or princess to be rescued from the Dragon by the Hero opens a wider vista that cannot be pursued here at length and will be treated in a future paper. This mytheme is the link to two widespread sets of myths.

(1) The myths of the slaying of the dragon e.g., Susa.no Wo in Japanese myth, as detailed above. This is sometimes connected with Summer Solstice and with marrying a local virgin. Examples include that of Indra slaying the dragon (Vedic India), the Thraētaona's slaying the Avestan three-headed dragon in summer, the BMAC/Nuristani dragon, as Nāga, melting ice and snow in early summer Witzel 2004). The Japanese version involved the (temporary) marriage of the dragon slayer to a local virgin; it is a mirror image of the myth of the release of the Sun woman from the cave (Witzel 2005).[44]

(2) The myths of a (temporary) marriage with a divine nymph or "Weaver Woman" by a celestial or early human ancestor, Hunter or Cowherd. It takes place around the time of summer solstice. Examples include the Vedic Indian Purūravas and Urvaśī (Ṛgveda 10.95; Witzel 1987), the Cowherd and the Weaver woman in China (July 7). Ultimately, it seems to go back to the foundational myth of the marriage of Sun and Moon, as is seen among the Kekchi Maya in Guatemala (Witzel 2005).

One could stop here and regard the dragon stories as old myths that deal with the doing away of the monsters populating the newly emerged earth. They need to be overcome as to allow life on earth (frequently, even before humans emerge). However, as indicated, a closer look at these myths reveals that they are part of a grander scheme. We begin in Greece, where we find a myth that is close to what we can detect in Vedic India in connection with another primordial myth, the creation of light (see Witzel EJVS 12–1, 2005). In this Vedic myth, the dawn is symbolized by reddish cows that are released from a cave through the onslaught of the great heroic god Indra, or Tajikara in the Iwato myth of the Jpn. Kojiki. In early Greece, the cows of Geryoneus are rescued by Herakles.

> Geryoneus, a giant with three bodies, is the son of Chryasor, son of Gorgo Medusa. He owns a great heard of cows on the island of Erytheia "redland", which is situated in the extreme West. The great hero Herakles, son of the sky God Zeus, in his 10th work, travels there, puts up his two columns at Gibraltar/Atlas, crosses

44. Note that the northwestern land of Izumo, viewed from the early center in Yamato (Nara area east of Osaka), is an exact mirror image of Yamato: the land and its mythology (Kojiki 1. 19–31) represents the dark, evening/night aspect of Japan (including the entrance to the netherworld). It is opposed to that of southeastern Yamato, where the Sun Goddess rules at Ise, and where her descendant, the emperor, entered from another "sunfacing" (*himuka* = Hyūga) country in Kyushu, and then resided at Asuka/Nara.

> the Okeanos in a golden beaker of Apollo (which he otherwise uses every night to cross the ocean). Herakles kills the herdsmen, the giant Eurytion, the two-headed dog Orth(r)os, and finally, Geryoneus with the shot of an arrow.[45] He then drives back the cows to Greece (where Eurystheus offers them to Zeus' wife, Hera).[46]

Obviously, the cows in the west are the opposite of the Dawn cows in the Vala cave of the East (Witzel 2005). Just as in the Veda (Jaiminīya Brāhmaṇa 2. 440–2), the island of the cows is situated in the extreme west, inside the ocean (*Okeanos*, Ved. *Rasā*), or rather on the borderline of subterranean waters, where one can get only in the same way as the sun crosses the ocean of night (as in Egyptian myth). The subterranean situation is also stressed by the appearance of the two-headed dog that reminds, in India, of the two "four-eyed" dogs of Yama, at the gate of the netherworld. Erytheia has long been understood as the "other world" (Jpn. *Tokoyo*), as the other world or the Savaiki (Hawai'i) paradise in Polynesian myth: both lie in the western direction.[47]

Therefore, the gaining of the cows by Herakles – who seems to be like a Greek Indra – looks like a summer solstice/evening myth, a mirror image of the Vala/Iwato myth, that is a winter solstice/morning myth (Witzel EJVS 12-1, 2005). Why should Herakles bring back the cows in the evening or at summer solstice? They are likely to disappear below the western horizon "forever". The effect thus is the same as in the Vala/Iwato myth: restoring the cows/dawn to humankind.

In Vedic myth, the slaying of the dragon is expressively correlated with the midday pressing of the sacred drink Soma and with Summer (hot season/later, = onset of monsoon). In earlier, Ṛgvedic times, when people speaking Old Indio-Aryan dialects were still residing in Afghanistan, Gandhara and the Panjab, this coincided with the snow melt in the high mountains of the Hindukush, Pamirs, and Himalaya and flooding of the rivers (Falk 1997). In the BMAC, Nuristani, and northwestern Indian myths, this is linked to the slaying of the great dragon/snake/Nāga (Witzel 2004, and forthc.) which cannot be explored here in detail.

However, a brief look will be taken at the other myth that links the Sun and Moon with Summer solstice. It comes from ancient China, and is known in Japan under the name of Tanabata, where it was introduced during the early Heian period, at

---

45. Cf. also the various myths of archers/bow shooter (Apollo, Marduk, etc.) and those in Avesta, Veda, China: 10 suns, nine shot down (note also the "10 little Indians" song?); further the Meso-American shooter myths: 3–4 suns are shot down, and the usurper, the fake sun (red Seven Macaw bird) falls from the tree, etc.; see Popol Vuh, II (Tedlock 1985: 90 sqq).

46. Note also the apples of paradise (China, Japan, Polynesia, Greece, etc.); for the Maori myth, see Tregear 1891: 56 sqq, s.v. *Hawaiki*.

47. The original blissful home of the Polynesians, *Savaiki > Savai, Hawaiki, Hawai'i, see Tregear 1891: 56 sqq.

c. 800 A.D. The cow-herd (*K'ien-niu, K'ien-niu-lang, Niu-Lang*) and the weaver girl (*Chih-nü),* are first mentioned already in the song *Ta-tung* "great east" of the Shijing/ Shih-king of the Zhou/Chou period (1027 BC.-771 BCE), and until c. 5th cent. BCE (cf. Yang & An 2005: 221).

> The cow herd lives in the asterism Eagle (Altair), and the weaver girl in Vega and Lyre (formed by a triangle of a larger and two smaller stars in Vega and Lyre). The weaver girl, the daughter of Heaven (Yu Yi), weaves heavenly clothes on the eastern bank of the heavenly river, the Milky Way. Heaven allows her to marry the cowherd. But as she neglects her weaving, she is banned to the eastern shore. (Alternatively, her mother uses her hair pin to scratch a dividing line, the Milky Way). Only once per year, in the night of the 7th of the seventh month she is allowed to visit her husband (see also King-Ch'u suei-shih-ki). This, she can do across a bridge (*Feng-su ki*), made by the wings of a magpie, a symbol of conjugal fidelity.

As the Milky way is a revolving river, which becomes clear in the myth of the Taoist Jun Ping, a famous Taoist of the 1st c., who lived at Chengdu in Sichuan: a certain person entered the Milky Way on a float and returned after one year.[48] The weaver girl and the cowherd live on opposite banks of the Milky Way. It is part of the ocean or broad river at the end of the world and it flows around (the top of) heaven and around the earth in the course of a year (Witzel 1984); entering on it, one returns to earth in due course. The once-per-year crossing must thus carry a special meaning.

Taking a closer look at the old Chinese situation, we notice for the old center of Chinese culture at Xian in Northern China, at 1000 BC: Altair (*Alpha Aquilae*) rises at 8:45 p.m, and sets at 9:41 a.m. on June 21, 1000 BC; while Vega rises at 5:42 p.m. and sets at 10:17 a.m.[49] Thus, Vega rises at sunset on the NE horizon, and Altair a few

---

48. There was a man who lived on an island near the coast. Each year, in the 8th month, inevitably a float passed by. The man took the strange decision to observe from that float. … He took hold there and departed. For about 10 days he observed the stars, the moon, the sun, and the whole firmament. Then, in the immeasurable darkness, he could no longer distinguish day and night. At the end of ten days (a decade), he suddenly arrived at a place resembling a village with walls. The dwellings appeared very austere. From far, he saw many weaver girls in the palace and also a man giving drink to cattle near a small island. The cowherd asked him, astonished: how did you come here? The man explained, and asked him about his origins. The cowherd told him: "when you will come to the district of Shu and visit Junping, you will learn it." He returned following the rhythm (of the stellar currant). In Shu, Junping told him: "on such and such a day, month, and year, a "visitor star" (a comet or shooting star) broke into the house of the cowherd." Making the calculation of the years and the months, the man noted that it was exactly the time when he had entered the Milky Way (on the float). (Bowu zhi, ch. 10).

49. Vega is visible at 6h morning c. Nov.–August, or at 6h evening c. May–March. In evening Vega and Altair are visible ca. from June–Dec.; in the morning c. from Dec.–July. This corresponds well with the summer and winter solstices.

hours later. Both are separated by the two branches of the Milky Way. The "wings" of the magpie (*Cygnus*) are spread between the two, a little off northwards, all across the Milky Way.

Weaver women are often identified with the sun: The Japanese sun goddess Amaterasu is or has weaver girl(s); in Vedic India day (masc.) and night (fem.) are frequently described as weaving a cloth, in a complicated pattern of day and night, and Night herself is also described as cloth. Even in far off Guatemala, there is a Kekchi Maya myth which speaks of a hunter (of deer) and a weaver woman, a "king's" daughter, who is locked into a room and emerges from there just as Amaterasu or Uṣas from the cave (Witzel 2005).

If the Chinese weaver girl represents the Sun goddess (dawn) at night, and as her representation, Vega, becomes visible in the evening around summer solstice, then the Cowherd[50] can well be the Chinese version of Indra/Susa.no Wo. Only, around Summer solstice, he does not break open the cave of the Sun goddess, but instead it is the sun goddess who comes to visit him. In all cultures involved she is the daughter of heaven: whether as Uṣas, the daughter of Dyaus (Vedic India), or as the Weaver woman, the daughter of Heaven (China), or the weaving daughter of a "king"(Kekchi Maya).

The Kekchi of Guatemala[51] tell a long story about the courtship of Sun and Moon. The future moon, a weaver woman and daughter of a "king" was shut up in room, and released by a hunter (the sun) in form of a colibri. Both escaped through the key hole, and the woman was killed by volcano fire but then reborn. The tale revolves around the marriage of the Hunter and the Weaver girl. In the end, Hunter and Weaver girl are again separated as Sun and Moon.

Winter	**only India**	**only Japan**	**Kekchi Maya**
sun (dawn) married or has sexual relations with her brother	marriage of Sūryā, RV10.86	Amaterasu and Susa.no Wo are siblings;	[1st part of myth] [Hunter and Weaver girl probabaly are original siblings]
(Weaver woman & cow herd in China)	Uṣas attacked by brothers (and father as antelope)	Susa.no Wo interferes with A.'s realm, attacks her (and servants) sexually; violent ascent to heaven	Hunter approaches Weaver girl; unites with her, entering her room through key hole, at night, as colibri/human, woman gets pregnant
Hunter and Weaver (Kekchi)			

(*Continued*)

50. Or the historically older, Meso-/Neolithic Kekchi figure of a *deer hunter.*

51. Bierhorst 1990: 112.

(Continued)

Summer	Indian myths	China	Kekchi
Sun woman (dawn) is married to her brother, the violent god (Susa.no Wo, Indra, cf. Cow Herd) who is assigned the lower world (or the moon)	Nymph Urvaśī married to human descendant of gods, Purūravas	Weaver girl is married to cow herd	[2nd part of myth] Weaver girl is "married" to deer Hunter
Sun woman (Sūryā, Urvaśī) has many previous lovers	Urvaśī is promiscuous, like Uṣas with poets		(meets with Hunter in colibri form, at night)
Separated from S. by Milky Way N.1.36;	separated one year; meeting again at lake, in company of nymphs in form of birds	separated by Milky Way	(separated by water/ ocean)
Deities are generally jealous of any sexual relation with a demi-god or human	Urvaśīs's heavenly (sexual) partners succeed in finally separating the two by lightening	she neglects her weaving; Heaven separates the two	father is "jealous", separates the two by volcanic fire
"oldest love story":	Purūravas roams around, madly	the two cannot meet	Hunter roams around for 2 weeks (half moon!)
(Bringing of Fire from demi-gods)	in desperation succeeds to meet her once per year.	meet once per year	meets her after "rebirth" from waters/bottles
meeting on/near ocean/water/ Milky Way	Gandharva fire (children)/ by going to heaven (RV, 10.95, st. 18); but around summer solstice, on Viṣuvat day of Gavām Ayana ("cows' walk") one rests on an island at the top of the sky/Milky Way near the North pole; cf. facilitating role of bird-nymphs	on July 7, crossing the Milky Way via a ford (the wings of magpie) they unite for one night only	(crossing of waters)
regular yearly meeting (full moon in Summer?)	Purūravas and Urvaśī meet once per year only (BŚS)		they had united for one night only; now separate as Sun and Moon

In other words, around summer solstice, the "cave opener" (Indra, Susa.no Wo, Cowherd, Kekchi hunter, etc.) and the goddess Dawn/Sun (Uṣas, Amaterasu, Weaver woman, Kekchi weaver girl) meet for sexual contact, which they *do not do* at winter solstice (as is hinted at in India and Japan, Witzel 2005).

In winter, the cave is magically opened through the performance of sexual rites *outside* the cave, in order to entice the dawn (Uṣas, Amaterasu) to come out. In summer, the opposite takes place. The female sun moves across the heavenly river (or water) to visit Indra (as Purūravas), the Cowherd, the Deer Hunter.[52]

The appearance of the myth of a temporary marriage between figures symbolizing the sun and the moon, both in Meso-America as well as in Eurasia, point to a date of the myth well before the first immigration into the Americas, by (at least) 15,000–11.5000 BCE.

In sum, the myth is a Laurasian one of the Late Paleolithic period.

It helps to shed light on the frame of mind of Stone Age humans.[53] In combination with other myths (Witzel 2004, 2005), the Laurasian mythology of the Late Paleolithic period is gradually emerging, and beyond that, the still earlier myths of anatomically modern humans, before they left Africa (Witzel 2005, 2006; van Binsbergen 2006; Berezkin 2002).

In sum, the approach of comparative mythology to prehistory is another pillar, next to the study of language, paleontology/genetics and archeology, in our quest to understand early humans. Collaboration in this long-ranging, complicated quest

---

52. Can this also be the situation of the Vedic Vṛṣākapi hymn RV 10.85 (Witzel 2005):

India: Indra + Indrāṇī :: Vṛṣākapi + Vṛṣākapāyī
Japan: Ta-jikara + Amaterasu :: Susa.no Wo + Uzume.

Note also the role of the magpie as bridge and the "eagle" (*Aquila*) asterism and cf. that of the Vedic *ariklava* messenger bird in the Jaiminīya Brāhmaṇa myth, and the role of the "gate" and the islands at this location of the Milky Way (Witzel 1984) .

53. The appearance of the Rainbow Snake in Sub-Saharan Africa, Australia and South America seems to represent an earlier stage in the evolution of human mythology, that of the Gondwana type mythologies. However, its appearance in S. America, that seems to contradict the existence of the dragon motif in early Laurasian myth, may well be a remnant of earlier Gondwana traits (Witzel 2005, 2006), such as the motif of the origins of humans from trees, that is found in Africa, Australia, but also in Laurasian Taiwan/Japan and Iceland. Further, the history of South American mythologies still is, by and large, untraced (Bierhorst 1988: 14 sqq), similar to that of the spread of S. American language families (see W. Davey, *MT* X: 162–171). Note also the appearance of the *Ogre* in African mythology, van Binsbergen 2006: 336.

is eagerly sought by the aforementioned and an emerging international group of scholars.[54]

References

Afghanistan, une histoire millénaire. Paris: Editions de la Réunion des Musées nationaux 2002.

Aston, W.G. 1972 [1896]. *Nihongi. Chronicles of Japan from the Earliest Times to A.D. 697*. Rutland: Tuttle.

Bächtold-Stäubli, H. & Hoffmann-Krayer E. 1987 [1930]. *Handbuch des deutschen Aberglaubens.* Vol. 2, 364–409. Berlin: de Gruyter.

Barber, E.W. & Barber, P.T. 2004. *When They Severed Earth from Sky: How the Human Mind Shapes Myth*. Princeton: Princeton University Press.

Beckwith, C.I. 2004. *Koguryo, The Language of Japan's Continental Relatives.* Leiden: Brill.

Beckwith, M. 1940. *Hawaiian Myth*. New Haven: Yale.

Benveniste, E. & Renou, L. 1934. *Vṛtra et Vṛthragna. Etude de mythologie indo-iranienne*. Paris.

Berezkin, Y. *Dwarfs and Cranes. Baltic Finnish Mythologies in Eurasian and American Perspective. 70 years after V. Toivonen*. Paper presented at a conference at Tartu 2002.

Binsbergen, W. van. 2003. "Rupture and fusion in the approach to myth. Situating myth analysis between philosophy, poetics, and long-range historical reconstruction, with an application to the ancient and world-wide mythical complex of leopard-skin symbolism. Paper at the Conference 'Myth and the disciplines' at Leiden, Dec. 12, 2003 http://www.shikanda.net/ancient_models/myth%20mineke%20defdefdef.pdf

Binsbergen, W. van. 2006. "Mythological Archaeology. Situating sub-Saharan African cosmogonic myths within a long-range intercontinental comparative perspective." *Proceedings of the Pre-Symposium of RHIN and 7th ESCA Harvard-Kyoto Round Table,* Ed. by T. Osada, 319–349. Published by the Research Institute for Humanity and Nature (RHIN), Kyoto, Japan. [cf. http://www.shikanda.net/topicalities/kyoto/kyoto.htm]

Bierhorst, J. 1986. *The Mythology of North America.* New York: William Morrow.

Bierhorst, J. 1990. *The mythology of Mexico and Central America*. New York: William Morrow.

Bierhorst, J. 1988. *The Mythology of South America*. New York: William Morrow.

Birrell, A. 1993. *Chinese Mythology. An Introduction*. Baltimore: Johns Hopkins University Press.

Blust, R. 2000. "Origins of Dragons." *Anthropos* 95. 519–536.

Campbell, Joseph. 1988. *Historical Atlas of World Mythology. Vol. I. The way of the animal powers.* Part 1: *Mythologies of the primitive hunter and gatherers*. New York: Harper & Row.

Colarusso, J., Ed. 2002. *Nart Sagas from the Caucasus*. Princeton: Princeton University Press.

Dalley, S. 1989. *Myths from Mesopotamia. Creation, The Flood, Gilgamesh, and Others.* Oxford: Oxford University Press.

---

54. An Association for Comparative Mythology (http://groups.yahoo.com/group/compmyth) was founded during the recent Beijing Conference on Comp. Myth., sponsored by Harvard and Peking Universities (http://www.people.fas.harvard.edu/~witzel/BeijingProgram.pdf); cf. also the three year Project of the Harvard Asia Center (2003–6) http://www.fas.harvard.edu/%7Esanskrit/myth.html; see now: http://compmyth.

Doty, W.G. 1986. *Mythography.* Tuscaloosa : University of Alabama Press.
Doty, W.G. 2000. *Mythography : the study of myths and rituals* (2nd Ed.). Tuscaloosa: University of Alabama Press 2000.
Evelyn-White, H.G. 1977 [1914]. *Hesiod. The Homeric Hymns and Homerica*. Cambridge, Mass.: Harvard University Press.
Falk, H. 1997. "The purpose of Ṛgvedic ritual." *Inside the texts, Beyond the Texts. New Approaches to the Study of the Vedas,* Ed. by M. Witzel, 67–88. Cambridge, Mass.: Harvard Oriental Series, Opera Minora.
Feldman, Burton & Robert D. Richardson. 1972. *The Rise of modern Mythology 1680–1860*. Indiana University Press.
Fontenrose, J. 1980 [1959]. *Python*. Berkeley: University of California Press.
Fornander, A. 1879–1885. *An Account of the Polynesian Race; Its origins and Migrations*, *and the ancient history of the Hawaiian people to the times of Kamehameha I.,* 3 vols. London: Trubner. [Repr. Rutland: Tuttle 1969].
Forssman, B. 1968. "Apaoša, der Gegner des Tištriia." *Zeitschrift für Vergleichende Sprachforschung (KZ)* 82. 37–61.
Francfort, P.-H. 1994. "The Central Asian dimension of the symbolic system in Bactria and Margiana". *Antiquity* 68. 406–418.
Frobenius, L. 1924. *Volksdichtungen aus Oberguinea*, *Vol.* 1, *Fabuleien dreier Völker.* Jena: Diederichs.
Graves, R. 1955. *The Greek Myths*, Harmondsworth: Penguin.
Gusinde, M. 1931. *Die Feuerland-Indianer: Ergebnisse meiner 4 Forschungsreisen in den Jahren 1918 bis 1924, unternommen im Auftrag des Ministerio de Instruccion Publica de Chile*. 3 Vols. [Vol. 1. *Die Selk'Nam*] Mödling: Verlag Sankt Gabriel.
Handbuch des Deutschen Aberglaubens, see: Bächtold-Stäubli 1987 [1930].
Ivanov, V.V. & Toporov, V.N. 1970. "Le mythe indo-européen du dieu de l'orage pursuivant le serpent. Réconstruction du schéma." *Echanges et Communications. Mélanges C. Lévi-Strauss*. Paris: Mouton.
Ivanov, V.V. & Toporov, V.N. 1974. *Issledovaniie v oblasti slavianskikh drevnostei*. Moskva: Nauka.
Katičić, R. 2001. *Auf den Spuren sakraler Dichtung des slawischen und des baltischen Heidentums.* Wiesbaden: Westdeutscher Verlag.
Lai, E.W. 1995. "Unmasking the Filial Sage-King Shun. Oedipus at Anyang." *History of Religions* 35. 163–185.
Lubotsky, A. 2001. "The Indo-Iranian Substratum." *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations,* Ed. by Chr. Carpelan, A. Parpola, P. Koskikallio, 301–317. Helsinki, Suomalais-Ugrilainen Seura.
Mathieu, R. 1989. *Anthologie des mythes et légendes de la Chine ancienne.* Paris.
Matsumura, K. 2005. "Myth Theories and War." Presentation at the XIXth World Congress of the International Association for the History of Religions (IAHR), Tokyo, March 2005.
Oberlies, Th. 2000. "Pūṣans Zahnlücken und Hermes' Vorliebe für Backwerk: Ererbte Strukturen des Pantheons der ṛgvedischen Religion." *Indoarisch, Iranisch, und die Indogermanistik, Arbeitstagung der Indogermanischen Gesellschaft 1997 in Erlangen,* Ed. by B. Forssman & R. Plath, 369–391. Wiesbaden: L. Reichert.
Philippi, D.L. 1968. *Kojiki. Translated with an Introduction and Notes.* Tokyo: University of Tokyo Press.

Puhvel, J. 1987. *Comparative Mythology*. Baltimore/London: John Hopkins Press.
Schmidt, Hanns-Peter. 1963. "Die Kobra im Ṛgveda". *Zeitschrift für Vergleichende Sprachforschung* (*KZ)* 78. 296–304.
Schmidt, Hanns-Peter. 1980. "On Birds and Dogs and Bats." *Persica* IX. 1–85 and plates I–XI.
Shishlina, N. & Hiebert, F.T. 1998. "The Steppe and the Sown: Interaction between Bronze Age Eurasian Nomads and Agriculturalists. *The Bronze Age And Early Iron Age Peoples of Eastern Central Asia,* Ed. by V. Mair, 222–237. JIES Monograph 26, Washington/ Philadelphia: The Institute for the Study of Man/The University of Pennsylvania Museum Publications.
Smith, G.E. 1919. *The Evolution of the Dragon*. London.
Sullivan, L.E. 1988. *Icanchu's Drum. An Orientation in Meaning in South American Religions*. New York: Macmillan.
Tedlock, D. 1985. *Popol Vuh. The Mayan Book of the Dawn of Life*. New York: Simon and Schuster.
Thompson, S. 1932–36. *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, Ind.: Indiana University Press. [2nd Ed. 1966; computer file, Blomington: Indiana University Press/Clayton: InteLex corp. 1993].
Watkins, Calvert. 1995. *How to kill a dragon: aspects of Indo-European poetics.* New York: Oxford University Press.
Wilbert, J. 1975. *Folk Literature of the Selknam Indians*. Berkeley: University of California Press.
Wilbert, J. & Simoneau, K. 1987. *Folk Literature of the Nivaklé Indians*. Berkeley: University of California Press.
Witzel, M. 1984. "Sur le chemin du ciel." *Bulletin des Etudes indiennes* 2.213–279. http://www.people.fas.harvard.edu/~witzel/CheminDuCiel.pdf
Witzel, M. 1990. "Kumano kara Woruga made." *Zinbun* 36.4–5. http://www.people.fas.harvard.edu/%7Ewitzel/Zinbun.htm
Witzel, M. 1997. "Saramā and the Paṇis. Origins of Prosimetric Exchange in Archaic India." *Prosimetrum*: *Crosscultural Perspectives on Narrative in Prose and Verse*, Ed. by J. Harris & K. Reichl, 397–409. Cambridge: D.S. Brewer.
Witzel, M. 2000. "The Home of the Aryans." *Anusantatyai. Fs. für Johanna Narten zum 70. Geburtstag* (Münchener Studien zur Sprachwissenschaft, Beihefte NF 19), Ed. by A. Hintze & E. Tichy, 283–338. Dettelbach: J.H. Röll. http://www.people.fas.harvard.edu/~witzel/AryanHome.pdf
Witzel, M. 2001. "Comparison and Reconstruction: Language and Mythology." *Mother Tongue* 6. 45–62. http://www.people.fas.harvard.edu/%7Ewitzel/Comp_Myth.pdf
Witzel, M. 2003. *Linguistic Evidence for Cultural Exchange in Prehistoric Western Central Asia*. Philadelphia: Sino-Platonic Papers 129. 1–70.
Witzel, M. 2004. "The Ṛgvedic Religious System and its Central Asian and Hindukush Antecedents." *The Vedas: Texts, Language and Ritual,* Ed. by A. Griffiths & J.E.M. Houben, 581–636. Groningen: Forsten.
Witzel, M. 2005a. "Vala and Iwato. The Myth of the Hidden Sun in India, Japan and beyond." *Electronic Journal of Vedic Studies (EJVS)* 12: 1.1–69 http://www.courses.fas.harvard.edu:80/~indst206/Background_Information/Vala_Iwato.pdf
Witzel, M. 2005b. "Out of Africa: The journey of the oldest tales of humankind." *Generalized Sciences of Humanity services, Vol I.* Tokyo, research Institute for languages and Cultures of Asia and Africa 2006:21–26.

Witzel, M. 2005c. "Slaying the Dragon in Eurasia." [Longer version of the present paper, including aspects of the summer solstices]. 2nd Indology conference at Nanhua University, Dalin, Chiyi, Taiwan, Oct. 2005.

Witzel, M. 2006. "Creation myths." *Proceedings of the Pre-Symposium of RHIN and 7th ESCA Harvard-Kyoto Round Table,* Ed. by T. Osada, 284–318. Kyoto: Research Institute for Humanity and Nature (RHIN).

Witzel, M. (forthcoming). *Origins and Dispersal of our First Mythologies*. [Book in preparation] (working title)

Yang, L., D. An, with J. Turner. 2005. *Handbook of Chinese Mythology*. Handbooks of World Mythology. Santa Barbara, Calif.: ABC Clio.

Trombetti

The forefather of Indo-Pacific

Jonathan Morris
São Paulo, Brazil

The work of the Italian linguists Alfredo Trombetti and Riccardo Gatti on their hypothesis of genetic relationship between the languages of the Andaman Islands, Papua New Guinea, Australia, Tasmania, and the Dravidian languages is discussed in detail. It is shown that Trombetti and Gatti had formulated a coherent precursor of the "Indo-Pacific" hypothesis (Greenberg 1971) by 1906.

I would like to begin by expressing a few words of thanks to Hal Fleming for his generous support. In our current era of academic hyperspecialisation, creating and sustaining an open forum such as Mother Tongue over two decades is a rare achievement in itself, but is made even rarer by Hal's enthusiasm for discussing and developing ideas, my work being a case in point, since he was the one who insisted on my developing a casual observation into an article on Trombetti's views on Indo-Pacific. Since only half-a-dozen of the 800-odd pages of his major work Glottologia *(Trombetti 1923) deal with the subject (see Morris 2006), this is no easy task. These few pages nevertheless contain many references to another earlier work by Riccardo Gatti (1906–1909) with two introductory essays by Trombetti himself, showing that the latter figure was closely involved with the project. Gatti merely trawled through the existing vocabularies of the day (E.M. Curr for Australia, Ray for British New Guinea, Schmidt for German New Guinea and Portman for the Andaman Islands) looking for cognates, but found so many that only a small fraction can be presented here. I have thus chosen to concentrate on his Andamanese data, both on grounds of relative completeness and because it demonstrates that Trombetti's belief in an intimate genetic relationship between the languages of the Andaman Islands, Papua New Guinea and Australia/Tasmania and the Dravidian languages was solidly data driven. Trombetti's comments also show that he had formulated a coherent precursor of the "Indo-Pacific" hypothesis by 1906.*

In my review of Trombetti (Morris 2006), I attempted to dispel the popular caricature that he had fallen into disgrace by espousing a single origin of language, as well as to introduce a more nuanced portrayal of him as an outstandingly gifted natural linguist who rose from the most modest circumstances to hold a chair at the university of Bologna, receiving official recognition until his death in 1929.

I also sought to show that his point of departure was a rigorous classical training in Indo-European and Semitic linguistics and a conviction that conclusions had to arise from his linguistic data. This put him very much at odds with the largely German linguists of the late 19th century such as Müller, who in Trombetti's view had acquired an ideological baggage that caused them to subordinate ideas about language to notions of race based mainly on physical anthropology. Such views, inspired by the evolutionary ideas of Ernst Haeckel, held that humanity had been divided into irreconcilable racial categories and hence that language, which arose at a later stage, must necessarily be polygenetic in origin. Trombetti saw himself as crusading for a return to the inclusive view of humanity of the Enlightenment, albeit with a difference.

In Trombetti's view, figures such as Franz Bopp had perceived that families such as Indo-European could have genetic links to other language families (notably the Polynesian family) but lacked the theory to elucidate the nature of these links. In the same way as Greenberg built up his theory of Eurasiatic by showing that Indo-European, Eskimo, Uralic, etc. shared a wealth of morphological elements, Trombetti believed that a similar theory was possible at an even greater level of generality, and that by categorising these, he would be able to explain many features e.g., of Indo-European, which had appeared baffling and arbitrary to Bopp. Trombetti suggested, for example, that what appeared (and still appear to us) to be indivisible elements, e.g., pronouns such as mi/ni/ki/ik actually had a fine structure in which the vowel was the semantic element and the consonant more of an auxiliary structuring element, or that verbal roots could actually be fusions of two separate verbal roots, $CV_1 + C_2(V_2)$.

Since, unlike his contemporary, Saussure, Trombetti regarded languages as inherently conservative, he believed that surviving languages could still provide insights into the origins of language itself, although he considered that the process which generated a fully functional and sophisticated agglutinative language had long since been completed.

Trombetti was writing without the benefit of our knowledge of hominid evolution or modern genetics, at a time when "primitive" peoples such as the Khoisan were widely regarded as a missing link between apes and humans. His view that they were highly intelligent humans adapted to a hostile environment was thus strikingly modern. At the same time, he did not share Rousseau's romanticised view of the "noble savage". If the Andamanese tribes had an extremely primitive lithic technology, it was, in Trombetti's view, because their skils in the area had generated, albeit by occupying the most remote and marginal habitats that were not subject to the pressures of civilisation, they had succeeded in maintaining their hunter-gatherer lifestyle until modern times. At the same time, the fact that his data showed the long-range conservation of morphological elements between "primitive" languages (e.g., numerals, pronouns, prefixes and suffixes) led him to equate cultural conservatism with linguistic conservatism.

Furthermore, Trombetti located the cradle of language in India, on the grounds that it was the area containing the largest number of major language phyla and had an apparent relative abundance of fossil evidence. From an Oceanic perspective, India was not such a bad guess, since, on the one hand, the modern "Out-of-Africa" initial rapid "beach-hopping" model actually does posit India as a secondary origin for such languages, and on the other, an "unrooted tree" approach to examining long-range linguistic relationships between Africa and Oceania is arguably more objective than one rooted in Africa.

Weaving these two strands of thought together, it should be clear that it was entirely natural for Trombetti both to expect the languages of a series of stone age peoples such as the inhabitants of the Andaman Islands, Papua New Guinea and Australia to preserve traces of their original unity in a South Asian homeland, regardless of how remote they had become from each other in geographical terms.

What is far more remarkable is that Trombetti appears to have changed his views radically some time between late 1905 and mid-1906, where he realised that his Andamanese-Papuan-Australian[1] family was distantly related to but discrete from Austro-Asiatic.

This is a significant change from the position of Schnorr von Carolsfeld (1890), who had proposed that the languages of Oceania belonged to a superfamily. In his publication dated July 1905, Trombetti appears to accept the former linguist's claim at face value:

> Indeed, the reciprocal affinity between all the languages of Oceania was affirmed and almost sufficiently demonstrated by Schnorr von Carolsfeld; (it is a pity that the author has not yet given us his promised work on the languages of Tasmania). The Malayo-Polynesian group was certainly connected with the Mon-Khmer languages, while the languages of our Andamanese-Papuan-Australian are particularly close to the Kolh languages, which in turn were connected to the Mon-Khmer group of E. Kuhn and others. If we then also remember the relations with the languages of Africa, the resulting image is one of a cycle or network. [Trombetti 1905: 16]

His preface to Gatti published a year later adopts a very different tone.

> As [Gatti] advises, the work is independent of Schnorr von Carolsfeld, despite the merits of the latter, who proposed to show the connection of **all** the Oceanic languages and thus extended his comparisons to Malayo-Polynesian. The need to distinguish two groups is nevertheless evident for lexical and above all grammatical reasons. I shall now indicate what seem to be the principal characteristics of the Andamanese-Papuan-Australian group.

1. Trombetti always assumed that Tasmanian was subsumed into Australian.

1. The phonetic system is simple and without spirant sounds. The Papuan spirants are of secondary origin.
2. Words are formed through prefixes and suffixes. Many adjectives are formed by duplication, those with a negative sense often derive from corresponding words with a positive sense.
3. In several languages of this group, there is grammatical gender. In Andamanese, we may note a kind of classification through prefixes (of which traces remain in Papuan and Australian) in names of parts of the body and kinship.
4. The declension takes the form of suffixes or postpositions.
5. An ergative case in use – i.e. of the operant subject. The verb that refers to the same often seems to be conceived as a passive. An extremely common ergative suffix is –da, -du, which is often omitted for personal pronouns.
6. The construction is inverse B-A.
7. The pronoun is declined like the noun. The dual and plural are formed in different ways from that observed in Malayo-Polynesian. In the first person of the dual and plural, there is often a distinction between inclusive and exclusive forms.
8. Possessive pronouns are formed in a different way from Malayo-Polynesian. Generally, they result from compounds of the personal pronouns followed by a particle (contrary to Melanesian and Polynesian languages).
9. Conjugation is complex. A final form may be distinguished that is a kind of supine like Latin ire dormitum.
10. Numbers are low, often merely binary.

In almost all these points, the characteristics of Malayo-Polynesian are opposite or different.

[Gatti 1906–1909, Vol. 1, Introduction viii–ix]

In *Glottologia* (Trombetti 1923), he merely embellished the above 10 points, suggesting that he had done little additional work on Andamanese since the work with Gatti[2] was completed before 1910.[3] While even in his late works, Trombetti could be irritatingly vague,[4] the fact nevertheless remains that between 1905 and 1906, he shifted from the

2. Apart from the work on Andamanese-Papuan-Australian, published in three volumes in 1906, 1908 and 1909, he appears to have published a small pamphlet on gender in Indo-European in 1912. There is also a small vocabulary of the local dialect of the town of Jesi (near Ancona) in 1920 by one Riccardo Gatti, but it is not clear whether the author is the same.

3. I have found references to an article which Trombetti wrote in 1921 on Papuan-African connections (*Festschrift Meinhof*) and another in 1926 on Tasmanian (*Acts of the 22nd International Conference of Americanists*), but have not succeeded in locating the original papers.

4. E.g., with his bad (by modern standards) habit of describing languages as "intermediate" between language families, such as Basque as "intermediate" between Kartvelian and Hamito-Semitic.

jumble of languages proposed by Schnorr von Carolsfeld to a clearly defined language family which subsumed Indo-Pacific.

His extension of this family to include Dravidian appears to have taken place by the time he had published his study of pronouns (Trombetti 1908: 1 I pronomi personali), largely as a result of his work on pronouns. The Narinyerri[5] language in particular intrigued him:

> The general and extremely close agreement between the Dravidian and Australian forms appears from the pronouns …
> Tamil: engal- (we, exclusive) = Aus: ngali, ngadli; Tulu: yenkulu (gentiv e), yenkule (id.); Aus: ngule, ngadli; Tulu: yenkulenu; Aus: ngulina
> Drav: nān (I), nām (we) = Narrinyerri: nān (me); nām we = Dabu: (Papua) nana (I); Drav: nam (we) = Narrinyerri: nām (we); Drav: num (we) = Narrinyerri: nōm (we)

Given that Narinyerri also bore similarities to Andamanese:

> Narrinyerri: ngu-rra, ngu-rre (you sing.) = General Andamanese: ngo-lla, ngu-le [also found in Aus 88, 205 ngooro, 84 ngurra, ngurru, 85 nooroo, 207E ngoro]
> Narrinyerri: ki-tje (he) = ki-te (Kede), li-le (Juwoi)
> Narrinyerri: mei-ke, mey-a-k (North) (who?, what?) = me-če (who?), mi-a-k, me-a-k (what?)
> Narrinyerri: ninka-ienk, ninga-u (South) (two) = ninaga (Önge)

While my primary aim is an accurate portrayal of Trombetti's ideas and data rather than an assessment of their merits, we may note in passing that the revision of his hypothesis to include Dravidian has the anomaly of showing better matches with Australian than with Andamanese or Papuan. See Appendix II.

Furthermore, Trombetti had noted that Dravidian showed far greater affinities to Hamito-Semitic (including a link to Nilotic through Elamite) than to Bantu-Sudanese,[6] while the converse was true of Munda-Polynesian. At the same time, in his analysis of prefixes, Trombetti believed that Bantu and Andamanese had cognate prefixes, **Bantu**: aka-mwa (mouth) and **Béa:** aka-bang-da, **Bálé:** aka-boang, etc. (mouth), where aka was originally a prefix indicating body parts, and **Bantu:** ele-(one of a pair)/**Béa**, **Bálé** I-dal, **Kédé:** er-tol (eye); **Kédé:** ir-pol (two).

Viewed from a modern perspective, I wonder whether Trombetti's link to Dravidian is not demonstrating something else, i.e. evidence of a much later migration from India to Australia during the Mesolithic/Neolithic and possibly associated

---

5. Modern Ngarindjerri – located on the lower Murray river and the Fleurieu peninsula in South Australia.

6. I use the term Bantu-Sudanese with Trombetti's meaning. It is clear from his work that the term is equivalent to the modern "Niger-Congo". See Morris (2006: 92) for discussion.

with flooding of ancestral habitats at the end of the Ice Age, as is suggested by recent genetic work (Redd et al. 2002). If so, then the implications of Trombetti's work are that the South Indian populations which migrated to Australia were Dravidian speakers, even if he was probably wrong to include this family in his Andamanese-Papuan-Australian phylum.

1. Andamanese

In *Glottologia*, Trombetti states his belief that languages related to those of the Andaman Islands had once been spoken on the mainland of South East Asia (from where the Andaman Islands themselves had originally been populated), and along the Indonesian archipelago all the way to New Guinea, being displaced by Malayo-Polynesian speakers at a much later date. He thus believed that aboriginal populations of Malaysia such as the Semang were related to the Andamanese, but that only a few traces of the latter's original language remained as a substrate in an otherwise Mon-Khmer language. He nevertheless cited some examples of these drawn from Skeat and Blagden (e.g., **Semang Snake** jĕ-kob, i-kob, see **Snake-2** below).

Returning to Gatti, the following is a list of all Andamanese words cited in his work, with respective cognates in Australian and Papuan. Gatti selected a series of around 100 lexical items, giving a set of polymorphisms for each item. His most copious raw material was the Australian material drawn from Curr. If, say, Curr described 10 Australian word families for snake, and the fifth word family had an Andamanese cognate, then this will be described below as **Snake-5**. It goes without saying that Trombetti did not have a comprehensive set of linguistic data to work with. Indeed, Curr himself presents a great deal of information on languages in S/SE Australia and hardly anything on the languages of Arnhem Land, which is evidently a serious shortcoming of his work.

At the same time, I take issue with the only modern citation of Gatti's work that I have found (Usher 2002: 67).[7] In my view, Usher's comments are invalid for three reasons. Firstly, it should be clear from the above discussion that Trombetti was perfectly well aware of the sharp morphological distinction between Andamanese-Papuan-Australian, as well as of the territorial distinction between what he called

7. "Both Trombetti and Gatti believed Tasmanian to be a member of a vast family, including the languages of New Guinea and Australia, along with Andamanese, and proposed many hundreds of etymologies of widely varying credibility, to support their contention. However, very little data from Indo-Pacific New Guinea was available to them, and their "Papuan" data is mostly Austronesian."

Melanesian languages spoken on the coast of New Guinea and the Papuan languages spoken in the interior (Trombetti 1923: 68). Secondly, I note that Trombetti's lexical items drawn from Ray's list of 22 Papuan languages and Schmidt's list of a further 15 languages substantially agree with the word lists/reconstructions given by Foley for the Lower Sepik, Gorokan, Kainatu and Proto-Highlands families (Foley 1986: 215–16, 246–8, 253–4, 257–8), which presumably pass the smell test as genuine Papuan data. Thirdly, despite little overlap between lexical items, there are some excellent matches between Trombetti's data and Whitehouse and Usher's own Kusunda cognates, with the data of the former extending the analysis of the latter to Australian.[8] It is true that Gatti includes a list of Malayo-Polynesian cognates in Vol. II, but it is abundantly clear from his and Trombetti's work that they regarded this relationship as a distant one.

2. Andamanese-Papuan-Australian Cognates

Father-5: **And Bálé** mámá, **Béa** mám-ola (father-in-law); **Aus** 10/190/213 mama, 9/12/25/27/28 mamma, 23/31 mam, 34/40 mumma, 199, māmoo, 201 maamo, 202/203 mamai, 204 maame, 206/207A, 207E maam, 209B maama, 207B/207F/207G/207H/208E mami, 21/22 mama-n, 16/17/19/40 mamma-n, 18/20/24/26/30 mammo-n, 208G maami-n, 209D mamo-n, 29 mama-tha; **Bng Koiari/Eikiri/Koita** mame, **Kupela** mama; **Gng Kai** mama, **Bongu** mem

Mother 1: **And Béa** chána-da**. Aus** 25/32/33 kun, 201 konoo, 190 gunnee, gunnie, koo-nee, koonea, goonee, gunnee, goonee, koonee, gooni

Mother-2: **And Kol** aute-tu-nen; **Aus** 159 too

Mother-3: **And Kédé** mémí, **Cháriár** ta-mémí (cf **Milk-2** and **Father-5**)

Mother-8: **And Júwoi** nāū-lekile [see wife][NB this suffix very common in **Júwoi**]; **Aus** 188 naae, 55 noa, 46 nooa (wife); **Bng Kupela**, **Meroka** neia, **Motumotu** nou; **Gng Valmau** nue, **Hatzfeldhafen** nana, **Wenke** nyan

Woman-2: **And Bálé** áb-cháu-pal; **Aus** 102 kooberrro (negress);

Woman-3: **And Béa** áb-cháu-da [see mother]; **Aus** 162 keen,

Woman-5: **And Kédé** eb-buku, 36 lāō-buku; **Aus** 43 boku, 45 bookoo; **Bng Kauralaig** i-pikai;

---

8. 1. **Kusunda Father** məm and **Father-5** above; 2. **Kusunda Short** potoə, **Gatti Small** Aus 16 poton, 19 bottene, 25 botine; **Gng** 1 pitine; **3. Kusunda Daylight** jina ïkya, **Gatti Sun** Aus 74 ooko, And Önge 9 éké; **4. Kusunda Breast** ambu = **Gatti Breast** Aus 69A ama, 72 amma, 145 ammooa, 155/164/167 amoo, 171 ema, 190 amoo, 47/48/74/76/79/107/172/173 umma, 77 ummi, 170 ummoo; **5. Kusunda Egg** gwá/góä poss. link to **Gatti EGG** Aus 107 kookurry, kokarri; **Bng Kauralaig, Saibai** kakur, **Gng Poom** kókoile-madeine; **6. Kusunda Dog** agai poss. link to **Gatti Dog** Aus 113 kia, 115 kāyā, 110 kai-a, 116 gyai; Gng Hatzfeldhafen ké.

Woman-6: And Béa áb-pāil-da, **Bálé** áb-pail [see **Wife-3**]; **Aus** 207E pulle-pulle, 197 balla-n

Woman-7: And Púchíkwar áb-ob-da, **Júwoi** á-ú-lekile **Kol** e-op-che [see **Wife-4**]; **Aus** 164 abo-n (mother), 170 oopar (wife);

Man-2: And Púchíkwar kāūro-da; **Aus** 185 koari, 186 korry, 187 korri, 188 kurri, 206/207G/208H koole, 207H kooli, 208J koday; **Bng Saibai** gara (boy), **Toaripi/Motumotu** karu (man); **Gng Valmau** kôl

Man-3: And Béa áb-chābil (having wives); **Aus** 136 kabulla, 179 gibberra, 164 gibere

Man-4: And Kédé é-tairu, **Cháriár** e-táru; **Aus** 60 thura, 61/62 tura

Man-5: And Béa búla-da, **Bálé** búla; **Aus** 181/190 boori; **Bng Elema** bira

Man-8: And Béa áb-wára-da (bachelor); **Aus** 176 woori-n

Child-1: And Béa áka-kádaka-da, **Bálé** áka-koádoko, **Púchíkwar** ó-kádaká-da, **Júwoi** oko-kádaká, **Kol** ó-kadaká-che; **Aus** 100 kuttukka

Child-2: And Cháriár chóté; **Aus** 53 kidtha, 196 goodtha, 182 kootha-ra

Child-3: And Kol á-t're-che, **Bójigiáb** ab-tiré, **Kol** e-tirá, **Önge** é-tire; **Aus** tuli, 186 talli, 210 tally-leet, 180 talli-waku (children)

Brother-1: And Púchíkwar ar-chúlutú-da, **Júwoi** rá-chúlútu **Kol** áke-chúlutú; **Aus** 28 koorda, 36 quertea

Brother-2: And Béa ár-dóáti-da, **Bálé** ar-doto; **Aus** 159 dutha, 167 dooda, 181 daidi, theady, tiade 102 theti, 194 tatha, 205 date (younger brother), 103 tita, 105 titi

Sister-1: And Béa ár-dóati-dar, ár-doti-t; **Aus** 205 date, 207P tati

Sister-2: And Púchíkwar á-chuletú-da, ra-chûletu, á-chúletú-n; **Aus** 24 quaret chukán, 33 quarrutchook

Old Man-1: And Púchíkwar áb-kára-da; **Aus** 69A karoo, 50, 51 kurroo, 141 kyerra, 52 kooroo, 155 kiara, 181 kure, 140 kaera

Old Man-2: And Béa áb-búla-da, **Bálé** áb-búla**: Aus** 97 boolgin-boola, 59 bool-ka, 60/61/62 pool-ka

Old Man-3: And Béa, **Bálé** áb-chāūraga-da [NB: the āū represents a long ō]; **Aus** 162 goorki, girki-l

Commander/Chief: And Béa áb-maiola; **Aus** 89 moolia, 101 moola

God: And Béa púlúga-da, **Kédé** bilke, **Cháriár** bilek-che, **Púchíkwar** bilik-da, **Júwoi** bilak-, **Kol** bilak-che; **Aus** 89 pargi-gi

Blood-3: **And Béa** téi-da, **Kédé** te-yi, **Cháriár** é-té, **Bálé** té, **Bójigiáb** te-wa, **Púchíkwar** teura-da, **Júwoi** tewa-, **Kol** tewa-che, **Gng?** so; **Aus** 163/165 dee, 163 du

Stomach-2: And Púchíkwar chúta-da, **Júwoi** chute, **Kol** chúte-che; **Aus** 190 gida-u, 77 koonto

Stomach 3: And Bójigiáb é-chulu; **Aus** 124 keela, 5 gooro

Eye-1: And Béa i-dal-da, **Bálé** i-dal, **Kédé** er-tol; **Aus** 114 dily, 9/10 toola [NB 50 languages]

Eye-4: And Chériar er-ulu; **Aus** 92 ale

Foot-8: And 9 mugé; **Bng Dabu** mak; **Aus** 1 macka, 96/99/130 mago, 100 mukko, 174 moko;

Foot-14: And Bójigiáb óng-ta; **Aus** 38/39 inga, 37 inka/inniga

Bone-4: And Béa ta-da, **Bálé** tóá, **Púchíkwar** tāū-da, **Júwoi** tau-, **Kol** tau-che, **Kédé** e-tu-we, **Chárіár** étoi-i; **Gng Kai** sie; **Aus** 140 toa, 159 deea, 169 dea

Ear-4: And Béa il-póko-da, **Bálé** id-poku, **Púchíkwar** ir-bó-da, **Júwoi** ré-bāūkāū, **Kol** er-bóke-che, **Kédé** ér-bu, **Chárіár** ér-buáh; **Aus** 37 i-bagi-ta, il-poki-ta, 38 il-poca-rta, ill-pocke-rta, 39 il-pucki-ta; **Bng Eikiri** i-piko, **Koiari** i-fiko, **Koita**, **Favele** i-hiko, etc.

Hair-6: And Béa pich-da, **Bálé** pich-da, **Púchíkwar** pāīch-da, **Kol** pāīch-che, **Kédé** paitch, **Chárіár** paitch; **Aus** 180 bookia, 214A pōk-kan.

Hand-2: And Béa ón-kāūro-da, **Bálé** óng-kāūro, **Púchíkwar** óng-kāūra-da, **Júwoi** āūn-kúrāū-, **Kol** ón-kāūre-che, **Kédé** ám-koro; **Aus** 1 queear-warra, 162 gillee

Head-2: And Bálé ót-chektá; **Aus** 66 kockerti

Head-5: And Béa ót-cheta-da; **Gng Bongu** gate, **Manikam** kadi, **Bogadjim** kate; **Aus** 18/19/20/30/123/133/142 katta, 22/141/147 kutta, 24 kata, 16 cata, 31 kaat, 122 kida, 123 kudda, 131 kutha, 132 katha, 144 kada, 158 kuddo, 61 a-kartee.

Head-6: And Kédé erchu; **Aus** 152 ulkey

Mouth-3: And Béa ákan-tewi, **Bálé** ákan-tewi, **Púchíkwar** óm-tewe, **Júwoi** āūkom-tewi, **Kol** óm-tewe; **Aus** 114 thowa, 138/140 towa, 129 dthowa

Nose-1: And Béa chāūronga-da; **Bálé** chaurnga**; Aus** 207B karnook

Nose-2: And Púchíkwar kāūté-da, **Júwoi** kāūte-, **Kol** kāūte-che, **Bójigiáb** mir-katto; **Aus** 117 kooda, 114 kootha

Skin-1: And ót-kāūpó, ót-kobau; **Aus** 161 kooba, 164 koba-ra, 166 kubari

Tongue-3: And Béa áka-tegi-da, **Bálé** áka-tegi, **Púchíkwar** ó-teke-da, **Júwoi** āūkāū-teke-, **Kol** āū-téki-che, **Önge** áka-teku, **Chárіár** áka-teku; **Gng Augustafluß** tega-l; **Aus** 207E thage

Tongue-4: And Béa áka-étal-da, **Bálé** áka-atal, **Púchíkwar** ó-tatal-da, āūkāū-tatal-, **Kol** tatal-che; **Aus** 16 tallan, 19 dallan, 158 dalli, 174 talle (v, widespread); **Tas Jorgensen** tullana

Teeth-4: And Béa tóg-da, **Bálé** tóg-; **Aus** 53 tiga, 167 ding, 169 deang

Teeth-11: And Puchikwar péla-da; **Aus** 152 pirra

Teeth-12: And Önge mákúe; **Aus 8** meeku-rro

Breast-1: And Béa kám-da, **Bálé** koam, **Bójigiáb** ir-kam-da; **Aus** 158 kammoo, 178 ngania

Breast-2: And Chárіár ot-char; **Aus** 109 coyar, 207H koroo-m, 207G kor-m

Fly-4: And Béa búmila-da, **Bálé** búmulá; **Aus** 183 boomal

Fly-5: And Chárіar pulimu, **Aus** 190 booreema, 197 borema-n

Mosquito-2: And Béa teil-da, **Bálé** tel**, Púchikwar** tél-da, **Júwoi** tel-, **Kol** tel-che, **Kédé** téil, **Chárіar** téil, **Bójigiáb** téil; **Aus** 191 doo-ra

Snake-2: And Púchíkwar chúpe-da, **Júwoi** chúpe-, **Kol** chupē-che

Prawn-2: And Béa wáka-da, **Púchikwar** waka-da, **Júwoi** wáká-, **Kol** wákāī-che; **Aus** 76 wegi-ga

Duck-1: And Béa kúlāla-da, **Kédé** kúlāl-da, **Púchikwar** kúlāl-da, **Júwoi** kúlāl-, **Kol** kúlāl-che; **Aus** 8 kooleyalli

Light-1: And Béa ar-chál-da, **Bálé** ár-chal, **Púchíkwar** ar-chol-da, **Júwoi** rá-chol-, **Kol** tu-chol-de (sunshine); **Aus** 176 gilli, 140 garra, 163 kirree, 205 karo

Light-2: And Kédé diö; **Aus** 175 dei, dooegi; **Bng Mowat** duo

Earth/Land-2: And Béa gara-da, **Bálé** goárada; **Aus** 179 garra, 82 karra

Earth/Land-4: And Önge tutanó; **Aus** 83 tuni, 181 toon, 181 towon/thoun/down/thone

Earth/Land-5: And Júwoi pákar-, **Kol** peakar-che; **Aus** 16 bugiara; 24 booga

Earth/Land-7: And Púchikwar pér-da; **Aus** 115 pōōră, 188 parri, 106 pulo, 113 boora, 186 bur-ray, 187 burrai, 169 burri

Fire-2: And Béa chápa-da, **Bálé** choapo; **Aus** 193 kanbi, 196 gunbey

Fire-3: And Önge tuké; **Aus** 7 toko

Hill, Mountain-3: And Béa bāūrōīn-da, **Bálé** bāūrōīn-da, **Púchíkwar** búrin-da, **Kol** búrin-che, **Kédé** burin, **Cháriár** burain; **Aus** 114 burry, 128 barrie, 153 birrie, 142 byrre, 41 purri, 102 poori, parroo, 147 paree, 150 parri

Moon-6: And Kédé chirké, **Önge** chilemé; **Aus** 136 karka, 144 karkurra

Moon-7: And Béa ógar-da, **Bálé** ógar-da; **Aus** 143 oggera

Moon-14: And Púchikwar púki-da, **Júwoi** púkui-, **Kol** púki-che, **Bójigiáb** puki; **Aus** 68 piki

Rain-7: **And Béa** yum-da, **Bálé** yum; **Bng Manukolu** ieme, **Aus** 154 ammoo, 155 amoo (next to k-amo, very widespread)

Water-8: And 9 ingé, **Béa** ina-da, **Bálé** ina, **Bójigiáb** ena, **Kédé** ine, **Cháriár** inó; **Aus** 171 yong

Water-10: And Púchíkwar léké-da, **Júwoi** léke-, **Kol** léke-che; **Aus** 5 lucka; **Tas Lhotsky** luga-na

Night-4: And Béa gúrug-da, **Bálé** gurug; **Aus** 76 kailka; 131 gōōrungā

Night-6: And Kol pāūti-che; **Aus** 177 pitta; 207 boroin (v. widespread); **Bng** 40 faita, faita buru

Star-4: And Bálé chalámi; **Aus** 120 karomi-n/karrome-n; 166 kal'bar

Star-5: **And Béa** cháto-da, **Kédé** kátáin, **Cháriár** kátáin; **Aus** 159 goonda, 26 ginda, 31 chindi

Stone-4: And Béa tāili-da, **Bálé** tāili; **Aus** 179 tarro, 98 diur, 178 tharo

Stone-5: And Önge taiyi; **Aus** 159 taye, 167 teya/doee

Stone-7: And Júwoi māka-, **Kol** méaka-che; **Aus** 88 maak, 87 mok, 203 maaki, 199/202 mukki

Sun-2: And Önge éké; **Aus** 74 ooko

Wind-3: And Bálé poát-nga, **Púchíkwar** pāūte-, **Kol** pátē-che; **Aus** 129 poodtha, 176 padoo-na;

Wind-6: And Béa wul-nga-da, wulanga-da; **Aus** 204, 207A willa, 202/203 willa-ngi

Wood/Tree-4: And Púchíkwar tákal-da, **Júwoi** tāūkal-, **Kol** tāūkal-che, **Bójigiáb** tákel; **Aus** 101 tooker

Wood/Tree-7: And Béa aka-tang-da, **Bálé** aka-toáng, **Púchikwar** ó-tong-da, **Júwoi** āūkāū-tang-, **Kol** tāū-táng-le; **Aus** 101 tooker, 110/113 toko

Wood/Tree-10: And Béa pútú-da, **Bálé** pútu; **Aus** 128 budda, 153 boodi
Day-1: And Bálé koarlo; **Aus** 159 geurlo
Grass-1: And Cháriár chálu-taich; **Aus** 148 kalla, 50 koola, 17 gila, 15 goola
Shine/Light/Sun-1: And Béa ker, **Bálé** kar, **Púchíkwar** ker, **Júwoi** ker, **Kol** ker; **Aus** 140 karra, 176 gilli, 163 girree, 205 karo
Shine/Light/Sun-2: And Béa bétel, **Bálé** bétel, **Púchíkwar** bātel, **Júwoi** bétal; **Aus** 69A peri
Thunder-1: And Béa púlúga-la, gāūrawa-che, **Bálé** púlúga-le, kúradá-ke, **Púchíkwar** bilak-le gāūrawa-ke, **Júwoi** bilak-le t'rémē-che, **Kol** bilak-ke, pārak-le; **Aus** 86 poorache, 85 poorook, 176 booringa, 149 booroon-ga, 152 baringa, 162 baroonji, 179 boorongi
Egg-7: And Cháriár jo-péro; **Bng Toaripi** fare; 155 parroo
Egg-9: And Béa ar-māūlo-da, **Púchíkwar** ár-múle-da, **Júwoi** rá-múle, **Kol** ta-mule-che, **Kédé** mulo; **Aus** 169 mor; **Bng Motumotu** mere, **Domara** muru
Egg-10: And Bálé māūlāīch; **Aus** 199 mīrkoo, 19 morgoo, 206/207A/207B/207D/207G/207J/207K/208C/208G/208H/208I mirk, 207H merk
Food-2: And Kédé te-jo, **Cháriár** áka-jeo; **Aus** 159 ja
Food-3: And Bójigiáb tama-da (see **Eat-7**); **Aus** 158 toomoo; 164 thumun, 205 tooman
Food-4: And Béa yát-da; **Aus** 91 yaddii, 177 yude, yuddy
Canoe/Ship-3: And Önge dángé; **Aus** 155 tangi-n, 182 toongoo-n
Bark-1: And Kédé ot-kápo, **Cháriár** ot-kaba; **Aus** 164 kumba, kumbar, 166 kombar, 99 cimbi-n
Milk-1: And Béa kám-rāīs-da, **Bálé** koám-yúrúch, **Púchíkwar** kom-rāīch-dá, **Júwoi** kāū-mērāīs- , **Kol** kom-rāīch-che; **Aus** 155 kammoo-n
Milk-2: And Kédé ir-mama-ti-óné; **Cháriár** ir-mamát-tí jöne (cf **Mother-3**); **Aus** 28 mimee, 163 maam, 197 mimi-n
Good-2: And Béa béringa-da; **Aus** 90 booraga, 94 poorooga
Bad-4: And Béa chāūru, **Bálé** chóáro **Púchíkwar** chárōō, **Júwoi** cherāō, **Kol** ch'rāō-wan; **Aus** 27 koorie, 1 guarra, 2 kawarra, 210 kiario; **Bng Kiwai** karakaroi
Hot-3: And Bálé úwia; **Aus** 213 oueba, 107 yowee
Cold-2: And Béa choki-da; **Aus** 103 kitcha
Cold-3: And Púchíkwar térem-da, **Júwoi** t'rem-che; **Aus** 197 thurrun
Large-4: And Júwoi cháki, **Bálé** kócku; **Aus** 124 kuka, 178 kuku-ne
Large-5: And Béa doga-da; **Aus** 187 tooka-l, 186 tucka-l
Large-11: And Béa bódia-da; **Aus** 9 boota, 213 budda
Small-4: And Béa ketia-da, **Púchíkwar** kátia-da, **Júwoi** chóté, **Kol** kātawa-le; **Aus** 180 kuthier, 99 kuddah
Small-5: And Cháriár jo-tāā-u; **Aus** 102 tjie-tjie, thieu, 104 tii
Dead-1: And Púchíkwar om-pil-nga, **Júwoi** am-pil-chikan, **Kol** om-pil-en, **Kédé** em-pil; **Aus** 181 bolloo, balu-n, 71 pooree, 190 ballu-n, ballo, boo-loo, pallo-nee, 61 palli-no
Dead-2: And Önge béchá-mémé; **Aus** 74/76/77 booka, 75 bōkkă, 80 pukka, 82 bocca, 90 boo-kia-ba

Hungry-1: And Béa wéraliké; **Aus** 74 wilka-wilka, 7 willu-ka, 82 willkaya, 79 wilka, 81 wilkoa

Hungry-2: And Púchíkwar kélape, **Júwoi** k'lipa, **Kol** kalipi, **Cháriár** o-chérpi; **Aus** 155 karbe-rri

Thirsty-1: And Kol áka-pāī, **Kédé** ta-pai, **Cháriár** ta-pai; **Aus** 179 boii

Very, Many, Full-1: And Kédé ir-kure; **Aus** 133/189 koora, 8 koolya

Very, Many, Full-2: And Bálé kóchu; **Aus** 199 koko, 80 koga

Very, Many, Full-3: And Júwoi á-chápar; **Aus** 123 curbara, kulburra

Very, Many, Full-4: And Béa doga-da; **Aus** 118 ducki-n

Very, Many, Full-5: And Bálé ár-púlia-da; **Aus** 18/19/21/22/24/25/27/31, etc. boola, 23 poola

Very, Many, Full-6: And Önge li-wángé; **Aus** 164 wingo-re

Sweet-1: And ? tálang-da, **Júwoi** tāūlang-, **Kol** tálang-le

See-1: And Púchíkwar ir-tilu, **Júwoi** re-t'liu, **Kol** er-tīlu; **Aus** 140 tilli-kuuckela, 131 telli-nulla, 141 etc. tilli [Eye]

Go/Walk-2: And Púchíkwar chóle, **Júwoi** chólē, **Kol** chóle, **Cháriár** chóle, ōi-choló; **Aus** 21 gooley, 21 koola, 33 kulli-ng

Go/Walk-3: And Kédé óichó; **Aus** 63 ooki-ta, 64 ooku-tta

Eat-3: And Kédé tojo, **Cháriár** tojojo; **Aus** 83 takkin, 186 tackenay, 155 daka, 187 taki, 187 thakoo, 207C takk, 160 dagga, 148 dangain; **Tas Milligan** tuggana

Eat-8: And Púchíkwar táme, **Júwoi** támē [See **Food-3**]

Eat-10: And Béa mek; **Aus** 214B maichimiak

Eat-13: And Kol āūkāū-yéu; **Aus** 179 yooa

Drink-2: And Béa to-ku, **Chárіar** to-ku; **Aus** 15 howa

Sit-2: And Júwoi ré-deka; **Aus** 148 teeka, 109 tucai, 62, tecku-nda, 65 ticka, 67 dikka-nie, 66 teku-nny

Sit-3: And Béa / Bálé aka-doi; **Aus** 149 tee-in

Sleep-6: And Béa mámi, **Bálé** mámi; **Aus** 61 meya, 63 mia, 60 miya

Sleep-7: And Kol móli, **Púchíkwar** móli; **Aus** 59 meer

Sleep-8: And Kol o-mókábé; **Aus** 48 mookapri-ri

I-2 – And Júwoi te-kile, **Kol** tú-che, **Kédé** tui, **Cháriár** tio; **Aus** 190 thu, 67 ti

I-4 – And Önge mi; **Aus** 186 mee, 214B my

You-2 (s/pl)-2: And Béa ngal-la, **Bálé** ngol, **Púchíkwar** ngule, **Kol** ngu-lu; **Aus** 88/205 ngooro, 84 ngurra, ngurri, 85 nooroo, 207E ngoro

One-5: And Béa/Bálé úbatúl; **Aus** bather

Two-6: And Béa ik-pāur-da, **Bálé** īd-pāūro-tot, **Púchíkwar** ir-pāūr, **Júwoi** ré-pāūr, **Kol** er-pāūr; **Aus** 69/76/163/179/181, etc. boola (v. widespread); **Tas Milligan** pooalih, **Peron** bura, **Jorgensen** boula

No-3: And Béa yába-da, **Bálé** yábo; **Aus** 214D yabba-la, 208D yembá, 164 abay

No-4: And Púchíkwar póye-da, **Júwoi** póya-, **Kol** póyi-che, **Kédé** puiyo; **Aus** 120 bai

Immediately-2: And Bálé ár-kāūwer, **Púchíkwar** ar-kewar-da, **Júwoi** rá-káūwer, **Kol** áká-wer-chē; **Aus** 163 kurra, 40 karree, 165 gurra, 44 karra; 62 earie.

Today-1: And Bálé il-kaólót; **Aus** 28 kordey; 207E kirdoo, 208F ketowit, 207F kerdo

Today-2: And Púchíkwar ábe-chil, **Kol** itabi-chél-lákele; **Aus** 190 keeli
Today-3: And Júwoi éte-kéle; **Aus** 205 keto
Tomorrow-1: And Béa wāi nga(i), wai-len, **Bálé** wó-nga-len; **Aus** 23 wooloolan, 31 Woolelan, 42/43/45/47 wongara
Where-1: And Béa tekári-cha; **Aus** 190 dagara, tugera
Where-2: And Kédé téin; **Aus** 38 thina
Where-3: And Bójigiáb ilé; **Aus** 171 yella, 173/178 ille; 172 illy, 184 yilla

2.1 Additional Andamanese-Papuan cognates in Volume II

Arrow: Bng ? ta-bora; **Gng Hatzfeldhafen** ta-barak; **And Júwoi** pelak
Eye: Bng Miriam pone; **And Bálé** punu, **Púchikwár** ir-bein-da
Foot: Bng Saibai koko; **Gng Kai** kike; **And Béa, Bálé** ar-chag-da, **Púchikwár** ar-chok-da, **Júwoi** ra-chok, **Kol** a-chok-che
House: Bng Motumotu umi, **Kabana** ema**; And Púchikwár** emi-da, **Júwoi** ami-, **Kol** emi-che
Husband: Bng Kabana harea; **And Púchikwár** ab-kara-da
Fortified Village: Bng Kauralaig kaura; **Gng Manikam** guré, **Bogadjim** kure**; And Béa** gara-da, **Bálé** goara
Lip: Bng Domar bibi-ta, **Mairu** noga pipi-ta, **Mowat** ipu; **Gng** pipi; **And Béa** pe-da, **Bálé** pa, **Púchikwár** pa-da
Mother: Bng Toaripi, Elema lou; **And Púchikwár** auto-lu-da
Sea-1: Bng Miriam gur; **And Bálé** júru, **Púchikwár** chira-ada, **Júwoi** chire-, **Kol** chire-che
Sea-2: Gng Kai hawe, **Bongu** kiwe; **And Béa** tauko-kewa-da, **Bálé** tauka-kewa (seashore)
Yam: Bng Miriam ketai; **Gng Kai** kise, qaso; **And Béa** cháti-da
Drink: Bng Kupela e-bai; **And Púchikwár** pai, **Júwoi** poi, **Kol** pai
He: BNG Manukolu oi, **Maiari, Meroka** oe; **And Júwoi** o-le, **Kol** o-le
His: BNG Mairu eke-ero; **And Bálé** égé
Basket: Bng Evorra kapi; **And Púchikwár** chop-da, **Júwoi** chop-, **Kol** chop-che
Body: Bng Kabana kan; **And Béa** chao-da, **Bálé** chaoo
Bow: Bng Saibai gagai, **And Júwoi** kok-, **Kol** kok-che
Fruit: Bng Saibai kauda; **And Béa** chita-da
Root: Bng Kauralaig kwiku, 32 kwik; **And Béa** ar-chag-da, **Bálé** ar-chag, **Kol** ta-chok-le
Red-1: Bng Koita kerekare; **And Béa** cherema, **Bálé** cherema
Red-2: Bng Koiari korika; **And Júwoi** chetak
Go/Come-1: Bng Mowat guitogu; **And Béa** katik
Go/Come-2: Bng ? onaio; **And Béa** on, **Bálé** aun, **Púchikwár** úne, **Júwoi** óne, **Kol** une
Live: Bng Kabana asi; **And Béa** ig-ati
Dig/Spade: Bng Miriam daiwi, dakellu; **And Púchikwár** tive-tau, **Júwoi** tiwetokau
My: Gng Bogadjim dyo(te); **And Béa** dia-da

Your: Gng Kai gole; **And Kol** la-ngú-le
Sky: Gng Hatzfeldhafen lamalam; **And: Púchikwár** lémar-da, **Júwoi** lémar-, **Kol** lémar-che
Good: Gng Manikam boleng; **And Béa** beringa-da
Come!: Gng Bongu gira!; **And Bálé** kélé

2.2 Key to Curr's Australian Classification[9]

1 Port Darwin, 2 Adelaide River, 3 Port Essington, 4 Rallies Bay, 5 Caledon Bay, 6 Roper River, 7 Cape York, 8 De Grey River, 9 Shaw River, 10 Nickel Bay, 11 NW Cape to 30 miles South of Gascoyne River, 12 Shark's Bay, 13 Mouth of Murchison River, 14 Northampton Tribe, 15 Champion Bay, 17 Victoria Plains, 18 Newcastle (WA), 19 Perth Tribe, 20/21 York District, 22 Pinjarra, 23 Kojomp/Etiomp, 24 Banbury, etc., 25 Blackwood District, 26 Lower Blackwood, 27 Irwin & Murchison, 28 Upper Sandford, 29 200 miles NE of Newcastle, 30 Mount Stirling, 31 King George's Sound, 32 Kent District, 33 Coast from Doubtful Bay to Israelite Bay, 34 Eyre's Sand Patch, 35 Eucla, 36 Head of Great Australian Bight, 37 Alice Springs, 38 Charlotte Waters, 39 Macumba River, 40 Streaky Bay, 41 Port Lincoln, 42 Peake Telegraph Station, 43 NW of Lake Eyre, 44 North Shore of Lake Eyre, 45 W of Lake Eyre, 46/47 Warburton River, 48 Cooper's Creek to E of Northern Branch and Koongi Lake, 49 Cooper's Creek in neighbourhood where Burke & Willis died, 50 Cooper's Creek near the Booloo River, 51 Nockatoonga, Wilson River, 52 Thargominda, Bulloo River, 53 Lower Bulloo River, 54 E of Strzelecki's Creek, 55 From Mt. Freeling to Pirigundi Lake, 56 Kopperamana, 57 Strangway Springs, 58 Umbartana, 59 Mt. Serle, 60 Beltana, 61 Wonoka, 62 E Shore of Lake Torrens, 63 Gawler Range, 64 Marachowie, 65 Mt. Remarkable, 66 Port Pirie, 40 miles E of, 67 Yorke's Peninsula, S. Australia, 68 Adelaide & neighbourhood, 69 Evelyn Creek, 69A Near NW Corner of New South Wales, 70 Country NW of Barrier Range, 71 Country about 60 miles NW from a point on the Darling, midway between Menindie & Wilcannia, 72 Boolcoomatta, 73 Torrewetto, 74 Lower portions of the Paroo & Warrego Rivers, 75 Bourke, Darling River, 76 50 miles below Bourke on the Darling River, 77 Wilcannia, 78 Tintinaligi, 79 Weinteriga, 80 Menindie, Darling River, 81 Tolarno Station, 82 Junction of Darling & Murray Rivers, 83 From the Banks of the Murray river where it enters Lake Alexandrina to the *embouchure* of that river & Lacapede Bay, 84 From Wellington on the Murray River to the North West Bend, 85 NW Bend of the Murray River, 86 Ned's Corner, 87 From the Mallee Cliffs to Wentworth,

9. The author wishes to thank Professor David Shanks of Ucl for his assistance in sourcing the material from Curr.

88 From the junction of the Lachlan and Murray to the junction of the Darling & the Murray, 89 E of Nicholson River and between that river & the coast, 90 Burketown, 91 Mouth of the Leichardt River, 92 Mouth of the Norman River, 93 Middle Norma, 94 W Bank of the Leichardt River, near sea, 95 Leichardt River, 20 miles below Kamilarci Station, 96 Kamilaroi Station, Leichardt River, 97 Between the Gregory and Leichardt Rivers, 98 Seymour, Templeton & Cloncurry Rivers, 99 Cloncurry River, 100 Flinders & Cloncurry Rivers, 101 Burke River, 102 Hamilton River/Lower Georgina River/ Between Georgina & Burke Rivers, 103 Head of Hamilton River, 104 On Hamilton River near Boulla, 105 Junction of King's Creek & the Georgina River, 106 Lower Diamantina, 107 Junction of Thomson & Barcoo Rivers, also the Whitula Creek, 108 Princess Charlotte's Bay, N Queensland, 109 Endeavour River, 110 Weary Bay, 111 Akconkoon, Palmer River, 112 Lynd River, 113 Granite Range, close to Head of Mitchell River and E of Hodgkinson Goldfields, 114 Near the Head of the Walsh River, 115 Country about Thornborough Diggings, and near the Head of the Mitchell, 116 Granite Range at the head of the Walsh River, 117 Head of the Gilbert, 118 Hinchinbrook Island and adjacent Mainland, 119 Herbert River, 120 Halifax Bay, 121 Headwaters of the Burdekin River, 122 Clarke River, 123 Top of Range near Dalrymple, 124 Cleveland Bay, 125 Mt. Elliott, 126 Mouth of the Burdekin River, 127 Porter's Range, 128 Charters Towers, 129 Upper Flinders, Hughenden, Dutton River, etc., 130 Watershed & Upper Portion of the Cape River. 131 Natal Downs Station, 132 Ravenswood, Upper Burdekin, 133 Mt. Black, 134 Lower Burdekin, 135 Burdekin River, various tribes, 136 From Port Denison to Cape Gloucester, 137 Tower Hill and Cornish Creeks, 138 Upper Thomson, 139 Head of Diamantina, 140 Diamantina River, Middleton Creek, 141 Western River, 142 Main Range between Belyando & Cape Rivers, 143 Belyando, 144 Logan Creek, Lower Suttor and Lower Mistake Creek, 145 Fort Cooper, 146 Scrubby Creek, 147 Port Mackay, 148 Broad Sound, Yaamba, Maryborough, St. Lawrence, 149 Rockhampton and Gracemere, 150 Eastern slopes of Expedition Range, Lower Dawson, Upper Fitzroy, Mackenzie and Isaacs, 151 Peak and Logan Downs, 152 Alice River, 153 Barcoo, 40 miles W of Blackall, 154 Blackall, Barcoo, 155 Barcoo, Tambo, Mt. Enniskillen and Ravensbourne Creek, 156 Negou, 157 Head of the Comet, 158 Brown River, 159 Dawson & Burnett Rivers, 160 Kuppol Bay, Calliope, Curtis Island, 161 Boyne River, 162 Bustard Bay, Rodd's Bay, Many Peak Range, 163 Baffle Creek, 164 N. Side of Moreton Bay, Maryborough, between Brisbane and Gympie, Great Sandy or Fraser's Island, 165 Upper Burnett, Mt. Debatable, Gayndah, 166 Mary River and Bunya Bunya Country, 167 Upper Brisbane River, 168 Brisbane River, 169 Condamine and Charley's Creek, 170 Stradbroke and Moreton Is., 171 Between Albert and Tweed Rivers, 172 Nerang Creek, 173 Tweed River and Point Dangar, 174 Part of Maranoa, 175 Balonne, Baleandoon, Nerrun, Weir and Moonie Rivers,

176 Dumaresque or Upper Mackintyre, 177 Paroo, Warrego and Mangalella Creek, 178 Richmond River, 179 Tenterfield, Glen Innes, 180 Queenbulla, Ashford and Quininguillan, 181 Barwan and Gwyder Rivers, Moree, Nundie, Wee-waa, Barraba, Bogabrie and Meeké, 182 Culgoa, 183 Brewarrina and Borwan, 184 Clarence, 185 Lower Macleay, 186 Port Macquarie, 187 Manning River, 188 Hunter River, 189 Hawkesbury and Broken Bay, 190 Castlereach River/Warren/Dubbo/Wellington/Hill End/Bathurst/Bogan/Sources of Bogan/Carcoar/Forbes and Levels/Candoblin/Waljeers/Wagga Wagga/Yanko, Urana, Billebong, Jerritderie/Deniliquin/Howlong/Albury, 191 Port Jackson, 192 Botany Bay, 193 Wollogong, Illawarra and Shoalhaven, 194 Jervis Bay, Mt. Dromedary, 195 Queeanbeyan, 196 Yase, 197 Moneroo, 198 Twofold Bay, 199 Swan Hill and Tyntynder, 200 50 miles S. of Swan Hill, 201 Piangil, 202 Bumbang, 203 Kulkyne, 204 Tatlarra, 205 Mt. Gambler, 206 Moreton Plains, 207A Lake Hindmarsh, 207B Lake Wallace, 207C Upper Glenelig and Wannon, 207D Glenelg, above Woodford, 207E Woodford, 207F Dartmoor, 207G Hamilton, 207H Mt. Rouse, 207I Portland, Condah, Eumeralla, 207J Hopkins River, 207K Hopkins River, 208A Moulmein, 208B Lake Boga, 208C Lower Loddon, 208D Gonn Station, 208E Gunbower, 208F Mt. Hope, 208G Kerang, 208H Natti-yallook, 208I Mt. Emu, 208J Moorabool, 209A Goulburn, near Seymour, 209B Upper Yarra, 209C Lower Yarra, 209D Merediyallook, 210 Gippeland, 211 Omee, 212 Snowy River, 213 Upper Murray, 214A Junction of Murray and Goulburn Rivers, 214B Tocumwall, 214C Ulupna (or Yoolupna), 214D Yiilima.

3. Trombetti's Dravidian-Australian correspondences

In Volume 2, Trombetti (1923) reproduces a list of general Dravidian-Australian cognates and then two separate lists of Tamil-Australian and Kannada-Australian cognates, based on Curr's Australian data and Caldwell's/Vinson's Dravidian data. These are transcribed here for reference purposes. I have also added Andamanese/Papuan etc. cognates in italics where these occur in Gatti's data.

The most obvious point is that Dravidian shows far more numerous links with Australian than with Andamanese or Papuan. Furthermore, he details a link between Dravidian and Hamito-Semitic, but finds more Bantu links to Andamanese. Viewed from a modern perspective, I wonder whether Trombetti's link to Dravidian is not actually demonstrating something else, i.e., evidence of a much later migration from India to Australia during the Mesolithic/Neolithic and possibly associated with flooding of ancestral habitats at the end of the Ice Age, as is suggested by recent genetic work (Redd et al. 2002). If so, then the implication of Trombetti's work is that the South

Indian populations who migrated to Australia were Dravidian speakers, even if he was wrong about including Dravidian in his Andamanese-Papuan-Australian phylum.

3.1 General Dravidian cognates

One-1: Malto, **Telugu** oru; **Aus** 105 ooroo, 106 orroo
One-2: Gondi vanda; **Aus** 205 wondo, 207E, F woondo, wondo
Two: Malto iru, īr; **Aus** 167 yero
Three-1: Malto mumdru; **Aus** 136 mundula; ***Gng 11 Augustafluß*** *mongul*
Three-2: Kui mūdu, **Kannada** mūdu; **Aus** 168 muddan
Three-3: Telugu mūnnu; **Aus** 11 mun-gooraba
Three-3: Brahui mūru; **Aus** 11 murra
Four: Malto, **Telugu** nālu, **Malto**, **Kui**, **Gondi** nal; ***Aus 63 nulla; Tas Peron nina; Gng 11 Augustafluß*** nun
I-1: Kaikadi anu; **Aus** 207A aan
I-2: Malayalam nanna; **Aus** 35 ngana
I-3: Malto nāṇ; **Aus** 18 nange
You-1: Malto, **Kui**, **Gondi**, **Telugu**, **Kuruth** ni; **Aus** 13 nee-nee
You-2: Gondi, **Tamil**, **Korri** nin; **Aus** 166 nin
You-3: Brahui ninu; **Aus** 120 nino
You-4: Kaikadi inu; **Aus** 164 in, 161 innoo
Foot-1: Kui, **Gondi**, **Telugu**, **Malayalam** kāl, **Kannada** kalu; 208H kaar, 208C kar, 179 garra, 213 gerra; ***Gng Kelana Kai*** *kiese*
Foot-2: Brahui khed; **Aus** 179 gidda
Nose: Malto, 4, 10 mūkku, **Kui**, **Gondi** muka, **Brahui** migu, **Kaikadi** mungeli; **Aus** 107 mingo, mingoo,
Eye: Malto kaṇa, khaṇṇa, **Gondi** kaṇṇa, **Kaikadi** kannu, 9 kan; **Aus** 37 ul-gana, 38 al-kna
Mouth-1: Tamil bai; **Aus** 142 beea, 177 be
Mouth-2: Malayalam ṭuḍḍI; **Aus** 62 thied
Mouth-3: Kannada noru; **Aus** 8 narra
Teeth: Telugu, **Kannada** pellu, **Kui** pell, **Gondi** pella, **Malayalam** pal; **Aus** 152 pirra; ***And Puchikwár*** *péla-da*
Hair-1: Kui magara, **Gondi** magri; **Aus** 170 magool
Hair-2: Brahui kūdalu; **Aus** 152 kultar
Hair-3: Tamil, **Malayalam** chutti; **Aus** 131 kuthy; ***Tas Jorgensen*** *cetha-na,* ***Lhotsky*** *ziti-na;* ***Gng Poom*** *hōdo*
Hair-4: Korri talith; **Aus** 73 tarta-woolka, 74 turtoo-bulki
Head: Tamil kukk; **Aus** 50 kooka, 48 koka, 181 kaoga; ***Tas Peron*** *cuegi,* ***Bng Kauralaig*** *kwiku,* ***Saibai*** *kuikö*
Belly/Stomach-1: Kui varag, **Gondi** vārga; **Aus** 11 waelgo, 12 wilguo; ***Gng Wenke*** *gulegim*
Belly/Stomach-2: Kui vayara; **Aus** 96 wyyeer

Belly/Stomach-3: Tamil kūl; **Aus** 118 kool-ko, 124 keela, 5 gooro
Belly/Stomach-4: Korri purath; **Aus** 154 burte
Belly/Stomach-5: Kaikadi vayara; **Aus** 190 daddo
Father: Korri abba, ābā; **Aus** 124/133 aboo [Correspondence with Munda]; ***Bng Kiwai, Miriam*** *aba,* ***Domara*** *abai,* ***Mairu*** *apai;* ***Gng Manikam*** *abú,* ***Bogadjim*** *abu,* ***Kadda*** *abe*
Mother-1: **Telugu** amma; **Aus** 104 amme, 102 ama; ***Bng Kauralaig, Saibai*** *ama;* ***Miriam*** *amau;* ***Gng Bongu*** *am*
Mother-2: Gondi gāmma; **Aus** 167 kame
Brother: Kaikadi dada; **Aus** 194 tutha, 167 dooda [Correspondence with Munda]
Sister: Korri bāyith; **Aus** 196 boady, 175 boade, 190 pate
Man: Korri murse; **Aus** 177 murdie
Child-1: Malto pillei, 10 pilla; **Aus** 97 bile, billa, 101 biller-biller
Child-2: Kui gunt, **Gondi** gunti; **Aus** 131 gundoo, 152 candoo
Child-3: Kaikadi miḍā; **Aus** 52 mitha-burlu, 51 mootha
Son/Young Man: Malayalam marri, **Kuruth** mār; **Aus** 145 maura, 62 meroo
Sun-1: Gondi khukkyā, **Brahui** chikka; **Aus** 89 koagigi, 169 koke, 176 koogee
Sun-2: Tamil binkō; **Aus** 135 boong-jouelbee, 205 boogil, 134 bangala
Water-1: Kui, **Gondi** tenni; **Aus** 201 teeni
Water-2: Telugu vellam; **Aus** 214 A, 214C wolla, 181 wollee, wollum, 94 wadda
Water-3: Tamil amm; **Aus** 131/142/154/174 ammoo, 156 amu
Water-4: Kuruth dīr; **Aus** 6 dilli-dilli
Go/Walk-1: [see **Foot**] **Telugu** po(ga); **Aus** 165 bego
Go/Walk-2: Tamil, **Korri** kalā; **Aus** 19 koola, 21 gooley
Eat-1: Malto, **Kui**, **Brahui** tinnu, **Malayalam** tin, **Kannada** tinu; **Aus** 46 tina, 49 tyena, 120 diana, 211 thianang
Eat-2: Tamil mōkhī; **Aus** 40 mungee, 214B maichimiak
Come/Walk: Tamil, **Korri** bara, **Malayalam** vara, **Malto**, **Kui**, **Gondi**, **Telugu** va; **Aus** 137 a-ber, 73 para-poo, 189 warre, 8, 9 wa-thi
Yes-1: Gondi hāl; **Aus** 69A ka, 159 ha-ha
Yes-2: Kaikadi a; **Aus** 41 a
No-1: Tamil alla; **Aus** 155 alla
No-2: Tamil mal'ā, **Korri** mala; **Aus** 102 malloo, 104 mallo; ***Tas Miligan*** *mallya leah*
No-3: Kannada lēdu; **Aus** 2 leita, leiter

3.2 Tamil cognates

Pelican: T kūlei, kadā; **Aus** 181 goola, guleala, 175 kuliallu, 190 koolay, 190 karlie, 214A kati-n
Crow: T kākkei; **Aus** 28 koko, 11 karko
Egg: T muttei, **K** motte; **Aus** 120 meto
Fish: T mīn; **Aus** 194 mena, 214B mani, 214D munni, 17 miye
Fly: T oara; 27 booara; ***Bng Kauralaig/Saibai*** *buli*

Snake: T pāmbu; **Aus** 156 bumba
Sister-1: T akkāl (elder); **Aus** 48 karoo, 97 koolakalla (elder & younger)
Sister-2: T tha-makkei [tha-honorary prefix]; **Aus** 84 maiko, 85 meeka, 87 maik (elder)
Brother-1: T tha-meiyan; **Aus** 201 mia, 182 moen
Brother-2: T thambi; **Aus** 120 tambua (younger)
Head: T mandei; **Aus** 99 munda
Hair-1: T mudi; **Aus** 6 moder; ***And Önge*** *māūdé*
Hair-2: T kūnthak; **Aus** 171 condur, 152 kuttar
Grass-1: T pul, pullu; **Aus** 10 peela, peelan; 6 bolea, 63 burree
Grass-2: T krāy; **Aus** 149 karra, 40 korra, 50 koola, 17 gila, 51 goola; ***And Chériar*** *chálu-taich*
Tongue: T nāru; **Aus** 116 nabbie
Breast: T kongei; **Aus** 208D chongo, 208G chaang, 204/207A chang
Bone: T mul; **Aus** moolale, modlallie, 75 mudda
Skin: T thōl, tholi; **Aus** 48 dulla, 46 dalla, 55 dula, 114 thilly
Guts: T kudal; **Aus** 171 giddirra
Light: T velichham; **Aus** 210 werrook
Cold-1: T peleiya; **Aus** 40 pyala, 63 pialla
Cold-2: T ārina; **Aus** 90 woorine, 171 waring
Cold-3: T kulir; **Aus** 181 karil, 149 kirroo
Forest/Wood: T kādu; **Aus** 179 gate, 178 gallee[10]
Hill-1: T malei; **Aus** 23 moolan, 106 meri
Hill-2: T mēdu, modu; **Aus** 99 minde
Hill-3: T kundu; **Aus** 166 kunda, 16 kata, 18 katta
Stone: T kal, kallu; **Aus** 108 koola, 181 gulla, 176 gooroa; ***Bng Kauralaig, Saibai*** *kula,* ***Kauralaig*** *kōla;* ***Gng Kai*** *gala, gola*
Bark: T pattei; **Aus** 42 peeta, 24 booto, 147 bittey, 63 patta
Bad: T athatha; **Aus** 28 thata
Food: [see Water, Eat]: T thīni; **Aus** 213 tunnam 190 dinu-ng, dana
Sleep: T urangu; **Aus** 10 warungo, 190 werrigoo
Drink: T kudi; **Aus** 107 kootha-ngo, 126 kudge-ogoo
See-1: T pār; **Aus** 170 parrari, 190 pirroo
See-2: T their; **Aus** 138 telli-mulla, 140 tilly-knukela; ***And Púchikwar*** *ir-tilu,* ***Júwoi*** *re-t'liu,* ***Kol*** *er-tīlu*
Sit: T kunthu; **Aus** 106 kunda, 159 gundower
Where?: T enge; **Aus** 147 anga, 83 yangi, 174 ingia
Full-1: T niranda; **Aus** 55 narpoo
Full-2: T miuli; **Aus** 150 mulea, 177 mulla-mulla, 10 maroo
Large-1: T perum; **Aus** 44 pirma; 46 pirna; ***Gng 5*** *boro;* ***Bng 47,48*** *baru-ka*
Large-2: T periya; **Aus** 46 piarree, 140 piala

10. Trombetti demonstrates the phonetic law: l = d, l = t elsewhere.

Small: **T** podi; **Aus** boti-ne, 150 pit, 16 poto-n
Dead: **T** patta; **Aus** 58 baad-lookoo, 167 bootir, 185 boote, 145 boonda
Earth: **T** mannu; **Aus** 125 mannie; ***Aus*** *71 mūndĭe;* ***Gng 5*** *moṇdam, 6 manidam, 7 mandam*
Tree/Wood: **T** maram; **Aus** 197 murru, 102 mooroo, 126 moora
Smoke: **T** pohei; **Aus** 23 poohey, 32 booey
Crayfish: **T** kalliral; **Aus** 205 keler
Turkey: **T** vankoli; **Aus** 39 wongarra
Dark: **T** irul; **Aus** 92 arreal
Today-1: **T** iṯ'téi; **Aus** 60 yeth, 65 yatto, 38 il-ytta
Today-2: **T** iṉ'dí; **Aus** 47 untie
Day/Sun: **T** ellei; **Aus** 13 ely, 190 eery, 181 eri
Tomorrow: **T** nalei; A ? noolar
See: **T** nokku; **Aus** 207 C naako, 190 nagoo, naagi, 47 nakkoo
Canoe/Ship-1: **T** kappel; **Aus** 24 kibera; *Gng 5, 6, 8 kabuṇ, 7 χubun*
Canoe/Ship-2: **T** kalam; **Aus** 209A, 209B korom
Dark/Grow Dark: **T** karu; **Aus** 131, 10 koora
Light: **T** vilakku; **Aus** 210 werrook; **And** 4 ár-lid-wálaich
Increase, Big-1: **T** migu; **Aus** 124 mooga, 113 muchan
Increase, Big-2: **T** mettu; **Aus** 190 mootoo, 179 mulla-mulla

3.3 Kannada cognates

Mouth: **K** bāyi; **Aus** 177 be, 142 beea
Elder Brother: **K** anna; **Aus** 142 att-ana (att is honorific prefix)
Large: **K** dadda; **Aus** 29 dudar
See: **K** nōdu; **Aus** 88 nithe, 188 natan, 178 nad, 161 natha
Hear: **K** kēlu; **Aus** 107 kurra, 191 gorai, 194 koori [ear]
Stand/Sit: **K** nillu; **Aus** 196 nulli, 181 naree
Where?: **K** yelli; **Aus** 171 yella, 184 yilla, 172 illy, 173/178 ille
Three: **K** mūru; **Aus** 64 murra, 14 marronoo
Fire: **K** ur-i; **Aus** 37, 38 oora, 39 ooraa
Old: **K** kiro; **Aus** 14 kyerra, 155 kiara, 140 kaera
Ear: **K** kêl; **Aus** 35 goolaya, 193/194 koori, 191 gora
Many: **K** pal; **Aus** boola (v. widespread); ***And Bále ár-púlia-da***
Boy: **K** mag-an; **Aus** 170 mugee; 190 makkoo
Mother/Elder Sister: **K** akka (elder sister) = **Aus** 212 yakkan, 210 yackan (mother)
Wind: **K** karr-u; **Aus** 190 gera, girar, kerare, 215 karrie

3.4 Sources of data

Australia & Tasmania (Aus/T): Curr (1886–1887)
British New Guinea (Bng): Ray (1895)
German New Guinea (Gng): Schmidt (1900–1902)

Andaman Islands (And): Portman (1898) [for Béa, Balé, Puchikwár, Júwoi and Kol]; Portman (1887) [for Bójigiáb, Cháriár and Önge].
Dravidian: Caldwell (1856); Vinson (1903); Skeat & Blagden (1906)

References

Caldwell, R. 1856. *A comparative grammar of the Dravidian or South Indian family of languages.* London: Harrison.

Curr, Edward M. 1886–1887. *The Australian race: its origin, languages, customs, place of landing in Australia, and the routes by which it spread itself over that continent.* Melbourne: John Ferres, Government Printer.

Foley, William A. 1986. *The Papuan Languages of New Guinea.* Cambridge, UK: Cambridge University Press.

Gatti, Riccardo. 1906–1909. *Studi sul Gruppo Linguistico Andamanese– Papua-Australiano, con una introduzione del Prof. Alfredo Trombetti* [Studies on the Andamanese-Papuan-Australian Linguistic Group], Vols. 1–3. Bologna: Luigi Beltrami.

Greenberg, Joseph H. 1971. "The Indo-Pacific hypothesis." *Current Trends in Linguistics. vol. 8: Linguistics in Oceania*, Ed. by T.F. Sebeok, 807–871. The Hague: Mouton.

Morris, Jonathan. 2006. "A Trombetti Documentary." *Mother Tongue* 10. 86–109.

Portman, M.V. 1887. *A Manual of Andamanese Languages.* London: W.H. Allen.

Portman, M.V. 1898. *Notes on the Languages of the South Andaman Group of Tribes.* Calcutta: Office of the Superintendent of Government Printing.

Ray, Sidney H. 1895. *A comparative vocabulary of the dialects of British New Guinea.* London: Society for Promoting Christian Knowledge.

Redd, A.J., Roberts-Thomson, J., et al. 2002. "Gene Flow from the Indian Subcontinent to Australia: Evidence from the Y Chromosome." *Current Biology* 12. 673–77.

Schmidt, P.W. 1900–1902. "Die sprachliche Verhältnisse von Deutsch–Neuguinea." *Zeitschrift für afrikanische und oceanische Sprachen* 5. 354–384, 6. 1–99.

Schnorr von Carolsfeld, Hans. 1890. *Beiträge zur Sprachenkunde Ozeaniens: Das australische Festland.* Munich: Bayerische Akademie der Wissenschaften.

Skeat, W.W. & Blagden, C.O. 1906. *Pagan Races of the Malay Peninsula.* London: Macmillan.

Trombetti, Alfredo. 1905. *L'unità d'origine del linguaggio.* Bologna: L. Beltrami.

Trombetti, Alfredo. 1908. *Saggi di glottologia generale comparata.* Bologna: Royal Academy of Science of the Institute of Bologna.

Trombetti, Alfredo. 1923. *Elementi di glottologia.* 2 volumes. Bologna: Zanichelli.

Usher, Timothy. 2002. "The Origin of the Tasmanian Languages." *Mother Tongue* 7. 65–84.

Vinson, Julien. 1903. *Manuel de la Langue Tamoul (grammaire, textes, vocabulaire).* Paris: Imprimerie nationale.

Otomanguean loan words in Proto-Uto-Aztecan maize vocabulary?

Jane H. Hill
University of Arizona

A suite of words for the maize plant, its cultivation and cuisine can be reconstructed for Proto-Uto-Aztecan (PUA), suggesting that its speech community included cultivators. Evidence is presented that some of this maize vocabulary may have been borrowed from an early stage of Western Otomanguean, no later than the breakup of Proto-Oto-Chinantecan. If this episode of contact can be supported, PUA was probably located in the northwest quadrant of Mesoamerica, no further north than approximately Queretaro, at 5000–4500 years ago. The possibility that the PUA/Western Otomanguean resemblances are descended from some ancient common source is considered, but contact seems a better explanation.

1. Introduction: Uto-Aztecan origins

Until quite recently, nearly all students of the Uto-Aztecan (UA) languages of North America agreed on the following broad outlines of Uto-Aztecan culture history. The proto-language community consisted of hunter-gatherers living in the uplands of the Gila River drainage system in present-day Arizona and New Mexico and the Mexican states of Sonora and Chihuahua (Fowler 1983). Long after the domestication of maize and the emergence of communities of cultivators in Mesoamerica, some UA communities adopted maize cultivation. The ancestors of the Aztecan subgroup moved south, reaching Mesoamerica itself by the sixth or seventh century A.D., and other groups moved as well, forming a southern UA corridor in western Mexico. All of the northern UA communities except the Hopi and a few of their nearest Southern Numic neighbors remained hunters and gatherers until the historic period.

Bellwood (1997) raised the first major challenge to this view, pointing out that the size and geographical distribution of the UA language family matched his model of language spread and radiation in a "Neolithic expansion", the demographically-driven spread of communities of early cultivators at the expense of neighboring hunter-gatherers. At about the same time that Bellwood began to make this argument, archaeologists in Mexico and the Southwest were revising the

prehistory of maize cultivation in several respects. First, radiocarbon dates on maize in the general region showed that cultivation was much more ancient in northwestern Mexico and the U.S. Southwest than had been previously thought, dating back to 4000 years ago in the Tucson Basin and the upper Gila River basin (Huckell 2005). Second, some early cultivator communities in the U.S. Southwest were clearly migrants, especially the Western Basketmaker II, who were established on the Colorado Plateau in the Four Corners area by 3500 years ago (Berry & Berry 1986; Wills 1995). Full sedentism of cultivators in central Mexico – suggesting that cultivation formed the principal basis of subsistence – dates only to about 4500 years ago. The rapidity of the spread of maize from central Mexico to the U.S. Southwest is also consistent with Bellwood's hypothesis.

Additional evidence in favor of Bellwood's position comes from historical linguistics. The lexicon of Proto-Uto-Aztecan (PUA) includes a suite of twelve items of maize vocabulary, shown in Table 1 (all but (6), recently identified by Alexis Manaster Ramer (personal communication) are from Hill 2001). All of these forms have cognates in both southern and northern branches of the family. While most of the cognates with meanings within the semantic domain of maize (instead of the pine-nut domain, where some of the northern cognates are located) come from a single northern language, Hopi, others are found in the Numic languages. Cognate for (1), part of the word for "tortilla," appear in Gabrielino and cognate for (4) is found in Luiseno.

Table 1. Proto-Uto-Aztecan maize vocabulary

1.	***sunu* "corn ear, maize"
2.	***sono* "maize byproducts such as cobs, leaves, cane"
3.	***pa7tz-i/a* "green corn ear ("elote"), corn kernel, seed"
4.	***saki* "parched corn, pop corn"
4.	***wïra* "to shell corn"
5.	***o7ri/a* "ear of corn, cob"
6.	***sura* "embryo of corn, tender ear of green maize"
7.	***tïma* "tortilla, tamale"
8.	***komal* "cooking surface for maize cakes"
9.	***kuumi/u* "to nibble small pieces of food, especially corn on the cob or popcorn; corn"
10.	***wika* "digging stick"
11.	***ïca* "to plant"

In Hill (2001) I argued that the PUA community was probably located in the northwest quadrant of Mesoamerica. However, given the speed with which maize moved from Mesoamerica to the U.S. Southwest, a Uto-Aztecan proto-community located in the Upper Gila drainage, as proposed by Fowler (1983), could have acquired maize and a vocabulary for it at a very early date. The purpose of the

present essay is to review new evidence for a Mesoamerican homeland, in the form of possible loan words from Otomanguean languages in the Uto-Aztecan maize vocabulary.

2. Otomanguean prehistory

Otomanguean is probably the most ancient of the Mesoamerican language families; Kaufman (1990) suggests that it is about 6000 years old. While its geographic extent is not as large as that of UA, it includes a much larger number of languages. Kaufman divides Otomanguean into two major branches, Eastern and Western. Eastern Otomanguean, with two major subbranches, each with three subdivision, includes two of the largest subfamilies: Mixtecan, with at least 25 daughter languages, and Zapotecan, with at least 40. Our concern here is with the smaller Western Otomanguean branch, which includes languages located in the northwest quadrant of Mesoamerica, a possible homeland for PUA. Kaufman organizes Western Otomanguean as in Table 2. We will be concerned mainly with the more northerly subbranch, Proto-OtoPamean-Chinantecan.

Table 2. Western Otomanguean

Proto-OtoPamean-Chinantecan
OtoPamean
Northern OtoPamean
Chichimec
Pamean (3 languages)
Southern OtoPamean
Matlatzinca-Ocuitelco
Otomi (4 languages)-Mazahua
Chinantecan
Tlapanecan-Chiapanecan-Manguean
Tlapaneca-Subtiaba
Chiapanecan-Manguean

Otomanguean languages are found throughout the central Mexican highlands, including the Tehuacan Valley and the Upper Balsas drainage, regions in which the earliest domesticated maize (and the earliest domesticated squash) have been identified. In the Upper Balsas grow the varieties of teosinte, the wild plant ancestral to maize, that exhibit the closest genetic relationship to maize itself (Benz 2001). Kaufman (1990) and Hopkins (1984) associate Proto-Otomanguean with the earliest Mesoamerican cultivators, the so-called "Tehuacan Culture".

In Section 3, I present evidence that several words in the PUA maize vocabulary are loans from Western Otomanguean, from a stage no later than Proto-OtoPamean-Chinantecan.

3. Resemblances between Uto-Aztecan and Otomanguean maize vocabularies

UA and Otomanguean maize vocabularies exhibit (1) words that are similar in form and meaning; (2) a resemblance in form between words for "corn" and words for "green"; and (3) a relationship between words in the "corn" and "green" complexes in both languages where word-initial *s alternates with word-initial *kw (in Otomanguean) or *ku/k (in UA).

The resemblance in form and meaning has long been noted. Campbell (1979: 949) suggested that Uto-Aztecan **sunu* "corn" is "almost certainly a loan from Otomanguean". Campbell thought that **sunu* was attested only in the southern UA languages, so that the loan would have not been into PUA, but into a later stage. However, **sunu* is attested in the northern languages, as seen in (1).

(1) **sunu* "corn ear, maize": Gabrielino *$ong-áxey* "tortilla" (*áxey* "thing to eat, pinole"), Hopi *songowï* "sand grass (*Calamovilfa gigantea*, a large grass with maize-like canes and tassels)", Tepiman **hu:nu-i* "corn", Taracahitan **sunú-t* "corn", Nahua **sin-tli* "ear of corn"

Campbell had in mind as a source the Otomanguean form seen in (2), as reconstructed by Rensch (1976). Rensch gives no Proto-Otomanguean (PO) glosses; the glosses given here are drawn from the meanings that Rensch gives for cognates in the daughter languages. Kaufman's slightly different reconstruction for the same item is also given.

(2) a. ***se(n)* (Rensch 1976, #285) "ear of corn, corncob, masa, corn flour"
b. ***sa(a')ai(n)* "ear of corn" (Kaufman 1990)

If this were an isolated resemblance, it could be dismissed as a likely case of chance. However, this item is part of a larger system of interlocking similarities. The first of these is that in both Uto-Aztecan and Otomanguean, and only in these language families in Mesoamerica, there are resemblant words in the proto-languages for "corn" and "green." Examining Mesoamerican iconography from the first millennium B.C., Taube (2000) observed a set of symbolic connections linking maize to two of the most important Mesoamerican prestige goods, greenstone (jade, jadeite, or serpentine, often carved into "celts" that represent maize ears and leaves) and iridescent green quetzal feathers. An obvious metaphoric link between maize and these two precious materials is the color green. The Uto-Aztecan forms are seen in (3), and the Otomanguean

words in (4). Note that these resemblances are not exact and often involve only the first syllable. The range of meanings of the "green" words include "grass", "leaf," and non-deciduous broad-leafed trees such as "willow." However, nothing remotely similar appears in any other Mesoamerican language families, where both present-day and proto-language words for "corn" and "green/grass/leaf/etc." are entirely distinct.

(3) Proto-Uto-Aztecan words for "corn" and "green"
   a. ***sunu* "corn ear, maize"/ ***suu-* : Nahua *xiwi-* "green, jade, turquoise"
   b. ***sono* "maize byproducts such as cobs, leaves, cane"/***so-* : In "grass" throughout family; Nahuatl *soso:wik* "something green, raw"
   c. ***saki* "parched corn, popcorn" : ***sa-* :PNUA *sakwa* "green, glue, color of turquoise", PUA ***sawa* "leaf"; PUA***saka* "willow, grass" (note Nahua *saka-tl* "grass", yielding Spanish *zacate*).

(4) Proto-Otomanguean words for "corn" and "green"
   a. ***se(n)* (Rensch 1976, #285) "ear of corn, corncob, masa, corn flour"/ ***se(n)*$^{1/2}$ "grass, hay" (Rensch 1976, #272)
   b. ***sa(a')ai(n)* "ear of corn" (Kaufman 1990)/ ***sa(n*) "green, raw" (Rensch 1976, #303); ***(n)sa* "green" (Kaufman 1990) (Suárez (1980: 59) reconstructs ***Ysa*)
   c. ***sau (-wV)* "masa, tortilla" (Kaufman 1990)

A brief discussion of these examples is required before we proceed. Note that in the UA examples in (3) there are two obviously very closely-related words for aspects of "corn", ***sunu* and ***sono*, the latter always having to do with something other than the ripe ear of corn, such as the cob left after shelling, leaves, stalks, or stubble. The word ***saki parched corn, popcorn*" had a third initial. Just as ***sunu* and ***sono* are parallel to two different words for "green" in ***suu-* and ***so-*, ***saki* resembles a third "green" word, a PUA word for "willow," **saka*. In the Otomanguean words in (4) there are similar confusions. Rensch (1976) and Kaufman (1990) disagree on the correct reconstruction of the word for "ear of corn." However, given the words for "green," for which Rensch recognizes a doublet ***se(n)* $^{1/2}$ (the superscript numbers represent reconstructed tone) and ***sa(n)*, it seems likely that there is a vowel doublet for "ear of corn" as well.

This complex of similarities between several words for "corn" and several words for "green" is not the end of the matter. In both proto-languages, there are also words for "corn" and "green" that have, not initial ***s*, but initial **k*^{w} or **ku* (simply **k* in the Uto-Aztecan form for "green"). These are seen in (5), for Proto-Uto-Aztecan, and (6), for Proto-Otomanguean.

(5) Uto-Aztecan words for "corn" and "green" in initial ***ku, *k*:
   a. ***ku:mi/u* "to eat food that comes in little pieces, especially corn and popcorn; corn (in PNUA only)"/**kan-* "willow" (PNUA only): Numic **kahna-*, Hopi *qaha*, Tübatulabal *ha:-l*, Cahuilla *qa:nki-sh*.

(6) Otomanguean words for "corn" and "green" in initial **kʷ:
   a. **kʷe* "corn, ear of corn, masa, atole, corn kernels" (Rensch 1976, #177)/ **kʷe* "green, blue, raw" (Rensch 1976, #174)/ ***w(y)e7* "green, blue" (Kaufman 1990); note Proto-Popoloca **šę~ kwhę7*; Proto-Otopame **poe-th*)
   b. ***kʷau* "corn" (Kaufman 1990)

Kaufman (1990) gives no cognate sets in support of his reconstructions. His reconstruction for "green, blue" seen in (6a) does not include initial **k*, but has only the glide ***w*. I give the Proto-Popoloca form as well as the Proto-Otopamean form from Rensch (1976), both of which have initial stops, to show that there is support for Rensch's reconstruction of an obstruent. It seems more likely that /kw/ would weaken to /w/ than that a stop onset would accrue to a glide, so I believe that Rensch is probably right about the initial ***kʷ*.

4. An Otomanguean origin for the resemblances

Section 3 shows that Proto-Otomanguean and Proto-Uto-Aztecan share several interesting features in their maize vocabularies. Not only are the words for "corn" resemblant, but in both language families, uniquely in Mesoamerica, these words for "corn" resemble words for "green." Finally, both languages show doublets, in both words for "corn" and words for "green," with words beginning in ***s* and words beginning in **kʷ,* ***ku*, or ***k*. While the words are short, and the resemblances are not exact, this is a fairly intricate set of interlocking resemblances, suggesting that it is not due to chance.

If these resemblances are the result of contact, the initial consonant alternation indicates that they are of Otomanguean origin. In PUA, the pair ***sunu/**kuumi/u* "corn, nibble corn" and the possible pair ***saka/**kan* "willow" are the only cases of ***s/**k* alternations within a semantic domain. However, in Otomanguean such alternations among initial consonants are broadly productive; word-initial consonant alternations are very well attested throughout the lexicon. Bartholemew (1965), Gudschinsky (1959), Longacre (1962), Rensch (1967) and Suárez (1980) all concur in the presence of a system of consonant alternations for Proto-Otomanguean and for several of the sub-families. The alternations probably result from the incorporation of prefixes into the stressed initial syllable of the root. Rensch (1967: 31–34) proposed six sets of consonant alternations for Proto-Otomanguean; the one of interest for our present purposes is **kʷ* ~ **y* ~ **s*. The Otomanguean "corn/green" words do not have any examples of initial **y*, but they attest ***kw* ~ ***s* in pairs like **se(n)* ~**kʷe* "corn" and **sa(n)* "green" ~ **kʷa* "bush, leaf, forage, etc."

In summary, several words in the Proto-Uto-Aztecan maize vocabulary resemble Proto-Otomanguean words in both form and meaning. They display the same metaphoric connection between "corn" and "green." They display a consonant alternation well-attested in Otomanguean, but not attested in Uto-Aztecan outside this narrow lexical domain. Thus it is likely that at least this part of the PUA maize vocabulary is borrowed from Otomanguean. More specifically, given that the distribution of Uto-Aztecan languages suggest an origin in the northwest quadrant of Mesoamerica, the loans are probably from an early stage of Western Otomanguean, before the radiation of the Oto-Pamean group as a separate northwestern branch of Western Otomanguean. The reason for this chronological suggestion is that, among the various Otomanguean consonant alternations, the **s-* ~ **k^w-* alternation does not reconstruct for the ancestor of the Oto-Pamean languages. The other major branch of Oto-Chinanteco is Chinantecan, which does preserve the relevant etyma with reflexes of initial **k^w* (Kaufman 1990: 103) and shows the **s-* ~ **k^w-* alternation (Rensch 1976: 335 (n. 14).

5. Implications for locating the Uto-Aztecan homeland

Kaufman (1990) dates the breakup of Western Otomanguean to about 4500 years ago, and the breakup of OtoPamean-Chinantecan to about 4000 years ago. This is not an unreasonable time period for PUA; while Miller (1983) suggested that PUA was six thousand years old, in Hale's (1958) glottochronological study of Uto-Aztecan, the oldest separation date is 4733. Kaufman locates the various dialect groups of Proto-Oto-Pamean-Chinantecan in the highlands of west-central Mexico. He places the dialects that gave rise to Chinanteco in the state of Morelos (the language is spoken today in a remote region of western Oaxaca), with Matlatzincan ancestral dialects to the west around Toluca in the state of Mexico, and Otomian ancestral varieties to the north in Querétaro. This crescent from Morelos through the Toluca region to Querétaro is an excellent candidate region for the location of Proto-Uto-Aztecan; these are largely pine-forested highlands, with precisely the same flora and fauna, at the generic level, that Fowler (1983) reconstructed for PUA in her argument for the upper Gila River drainage as a likely Uto-Aztecan homeland.

If the maize vocabulary of PUA indeed comes in part from Otomanguean, we can probably rule out the upper Gila drainages as the homeland, and locate a PUA *Urheimat* on the northwestern boundaries of, or indeed, well within the boundaries of, Mesoamerica. The most northerly languages in Otomanguean are the Pamean languages, spoken in the state of San Luis Potosí. The Pameans historically were hunter-gatherers, not cultivators; this certainly represents another

case of agricultural regression, since it is clear that Proto-Otomanguean was spoken by a community of cultivators. While one should not discount the blurring of the geographical history of Otomangean during four thousand years – indeed, we know that in the late prehistoric period Otomangueans (specifically Otomí speakers) were retreating in the face of Aztecan pressure – it seems very unlikely that any Otomanguean language was ever spoken as far north as the Gila River headwaters.

6. Central Amerind or some other ancient "common source"?

Since this essay is submitted in honor of our colleague Hal Fleming, a notable long-ranger, I have a responsibility to consider whether these resemblances between Otomanguean and Uto-Aztecan maize/green vocabulary are "sprung from some common source," rather than due to contact.

Greenberg (1987) groups UA and Otomanguean, together with Kiowa-Tanoan, in his Central Amerind branch of Amerind. If there was a "Central Amerind," it could not have had words for "maize." The most ancient radiocarbon-dated maize, from Guila Naquitz, Oaxaca, is 6200 years old. This is much too late for anything like "Central Amerind." This is very close to the date proposed for the initial breakup of Otomanguean by Kaufman (1990). As noted above, the dates proposed for PUA range from 6,000 to less than 5,000 years ago. Hale and Harris (1979) believed that Proto-Kiowa-Tanoan dated to between 3,000 and 2,500 years ago.

Cultivated maize cannot reproduce without human intervention, so no language had a word for "maize' before the plant was domesticated. However, hunter-gatherer populations in central Mexico must have been exploiting its immediate ancestor, *teosinte* (*Zea mays*, var. *parviglumis*), as a precursor to domestication. Spanish *teosinte* is from Nahua *teo-sin-tli* "divine-maize-absolute.noun", which looks like a marking reversal from Nahua *sintli* "fully ripe ear of maize, mazorca". Might teosinte have been the original plant denoted by something like ****sVn*? In fact, Greenberg (1987: 127) gives the Otomanguean word for "grass", already discussed above in Section 3, in his Central Amerind set for FLOWER, adapted in (7):

(7) Central Amerind FLOWER: Tiwa *ɬi-* "grass", PO ***(H)se(n)*$^{1/2}$ "grass, hay" (Rensch 1976, #272), PUA **se*/**si* "flower".

I think this set of resemblances is better accounted for within the framework of language contact suggested here. First, as I have pointed out above, I believe that the Otomanguean word is part of a "corn/green/grass/leaf" complex borrowed into PUA. Second, the PUA word for "flower" (which Campbell & Langacker (1979) recontruct as ***sïyo-*) probably does not belong here, because there is a much

closer set of resemblances involving words for "grass," between Proto-Northern Uto-Aztecan (PNUA) and Proto-Kiowa-Tanoan (PKT). This set is part of a set of several items of maize vocabulary (the "grass" word comes from a PUA word for "maize by-products") that I believe were loaned from PNUA into PKT during a period of contact between 3500 and 2500 years ago in the Four Corners region of the Colorado Plateau (the archaeological context is the contact between the Western Basketmaker II and the Eastern Basketmaker II (Hill 2002, 2008). During the same period, PKT evidently loaned several words for wild plants of the Colorado Plateau to PNUA (among these are the items in Greenberg's Central Amerind sets for "fruit" (the "pine-nut" pair) and "onion"; in both cases the items cited are only in NUA). The NUA-KT resemblant sets for "grass" are seen in (9) and (10). They follow a pattern attested in other pairs where a UA nasal consonant appears as nasalization on the KT vowel.

(8) PNUA **songo* "grass, tinder/hay, corn cob" (< PUA***sono* "corn by-products (e.g., cobs, stalks, stubble)": Mono *sona* "hay"; Tηmpisa Shoshone *soni* "grass" (archaic); Owyhee Shoshone *soni-pï* "grass", *soni-pi* "rye grass"; Comanche *soni* "grass"; Southern Paiute *šoni* "tinder"; Hopi *sö:ngö* "corn cob"

(9) PKT **sV̨-*: Taos *łi-ne* "grass"; Jemez *tʸį́:* "grass"; Kiowa *sǫ-dɔ* "grass"; cf. also Kiowa *sǫ-n* "single grain of oats, grass seed".

Hale (1962: 4) observes that in this set the Jemez and Kiowa forms are good cognates, but the vowel in the Taos form is wrong (the cognate vowel would be *ǫ*). It is precisely this aberrant Tiwa form that appears in Greenberg's set in (7); it is taken from Whorf and Trager (1937), who did not accept Kiowa as a genetic relative of Tanoan.

A second concern that should be raised in connection with Greenberg's proposal is that UA shares at least as many resemblances as it does with Otomanguean with another Mesoamerican language family, Mixe-Zoquean, which Greenberg places in his Penutian branch of Amerind. Wichmann (2003: 326) notes 43 resemblant sets for PUA-Proto-Mixe-Zoquean, including four affixes. One of Wichmann's affix pairs, shown in (10), is of special interest.

(10) UA **-mï* "plural", MZ **-ta-m* "plural"

The reason that this pair is interesting is that we can reconstruct alongside the UA plural suffix **-mï* a second suffix, **-tï*. While many UA languages have both these plural suffixes, only in Aztecan do they appear together in the same word, as in Nahua-*ti-n* and Pipil *-me-t.*, resembling the MZ form **-ta-m*. But the presence in both Uto-Aztecan and Mixe-Zoquean of **t* and **m* in plural suffixes may constitute an example of shared anomaly attesting to descent from a common ancestor.

Pursuing the Mixe-Zoquean question a bit further, we find that these languages do have words for "green" "grass", seen in (11), that vaguely resemble those already

discussed. Note that these do not resemble any of the Proto-Mixe-Zoquean (PMZ) words in the maize complex.

(11) a. Proto-Mixean **šokot* "grass"; Proto-Zoquean **so7k* "grass" (Wichmann 1995)
b. Proto-Mixe-Zoquean **cusu(k)* "green"

The Proto-Zoquean word in (11) is probably the source of Proto-Yucatecan (Mayan) *suHk* "zacate" (Kaufman 2003). A good deal of exchange between Mixe-Zoquean languages and Mayan can be identified; another example with sibilants in the "corn/green/grass" domain is Proto-Zoquean **tzutu* "espiga de maíz", probably from Proto-Mayan **tz'utu* "flor de milpa". In Mayan, the only word of potential interest in the "green" series is Proto-Mayan **xaq* "leaf", resembling PUA**saka* "willow, grass" – and the Mixean and Zoquean words in (11a) are also reminiscent of the PUA word.

To summarize this section, if I were a competent long-ranger, I would be looking for a deep-ancestral Mesoamerican language with a word for something like "grass", "green", "leaf," manifesting an initial sibilant. I would be trying to identify the sources of the many apparent derivational elements that have accrued to this word. I would be wondering if any of these derivations could be shown to be a word for *teosinte*. But I hope that I have shown here that untangling the resemblances that are due to contact, especially contact during the period when the Mesoamerican ethnobotanical vocabularies were restructuring to accomodate domesticated plants, and the resemblances that are descended from this hypothetical deep common ancestor, is a formidable task. The six thousand years or so since the domestication of maize is a very long time. The ten thousand years or so since something like the present-day biogeography of central Mexico began to take shape is even longer. An immense amount of noise must have been introduced into the trace signals left by the founding languages in the region, and the only suggestion in which I have any confidence is the one made in the first few sections of this essay, for the possibility that some UA words for "maize" are borrowed from an early stage of Western Otomanguean, probably about 4500 years ago.

References

Bartholomew, Doris. 1965. *The Reconstruction of OtoPamean*. University of Chicago Ph.D. Dissertation (cited in Rensch 1976).

Bellwood, Peter. 1997. Prehistoric cultural explanations for widespread linguistic families. *Archaeology and Linguistics: Aboriginal Australia in Global Perspective*, Ed. by P. McConvell & N. Evans, 123–134. Melbourne: Oxford University Press.

Benz, Bruce. 2001. Archaeological evidence of teosinte domestication from Guila Naquitz, Oaxaca. *Proceedings of the National Academy of Sciences, U.S.A*. 98: 4. 1324–1326.

Berry, Claudia F. & Michael S. Berry. 1986. Chronological and Conceptual Models of the Southwestern Archaic. *Anthropology of the Desert West: Essays in Honor of Jesse D. Jennings,* Ed. by C.J. Condie & D.D. Fowler, 253–327. Salt Lake City: *University of Utah Anthropological Papers*, no. 110.

Campbell, Lyle 1979. "Middle American Languages." *The Languages of Native America: Historical and comparative assessment*, Ed. by L. Chambell & M. Mithun, 902–1000. Austin, Tex. & London: University of Texas Press.

Fowler, Catherine S. 1983. Lexical clues to Uto-Aztecan prehistory. *International Journal of American Linguistics* 49. 224–257.

Greenberg, Joseph H. 1987. *Language in the Americas*. Stanford: Stanford University Press.

Gudschinsky, Sarah C. 1959. *Proto-Popolocan, A Comparative Study of Popolocan and Mixtecan*. Indiana University Publications in Anthropology and Linguistics, num. 5. Baltimore, Md.: Waverly Press.

Hale, Kenneth L. 1958. Internal diversity in Uto-Aztecan: I. *International Journal of American Linguistics* 29. 101–107.

Hale, Kenneth L. 1962. Jemez and Kiowa correspondences in reference to Kiowa-Tanoan. *International Journal of American Linguistics* 28. 1–5.

Hale, Kenneth L. & David Harris. 1979. Historical Linguistics and Archaelogy. *Handbook of North American Indians, Southwest*, Vol. 9, Ed. by A. Ortiz, 170–177. Washington, DC: Smithsonian Institution.

Hill, Jane H. 2001. Proto-Uto-Aztecan: A community of cultivators in central Mexico? *American Anthropologist* 103: 4. 913–934.

Hill, Jane H. 2002. Toward a linguistic prehistory of the Southwest: "Azteco-Tanoan" and the Arrival of Maize Cultivation. *Journal of Anthropological Research* 58: 4. 457–476.

Hill, Jane H. 2008. "Northern Uto-Aztecan and Kiowa-Tanoan: Evidence for contact between the Proto-languages?" *International Journal of American Linguistics* 74: 155–88.

Hopkins, Nicholas A. 1984. Otomanguean linguistic prehistory. *Essays in Otomanguean Culture History*, Ed. by J. Kathryn Josserand, Marcus Winter & Nicholas A. Hopkins, 25–64. Nashville: Vanderbilt University Publications in Anthropology No. 31.

Huckell, Bruce B. 2005. McEuen Cave and the origin of Southwestern U.S. agriculture. Paper presented at the 70th Annual Meeting of the Society for American Archaeology, Salt Lake City, Utah, March 31, 2005.

Kaufman, Terry. 1990. Early Otomanguean homelands and cultures: Some premature hypotheses. *University of Pittsburgh Working Papers in Linguistics* 1. 91–136.

Kaufman, Terry, with the assistance of John Justeson. 2003. A Preliminary Maya Etymological Dictionary. Electronic document at http://www.famsi.org/reports/01051/pmed.pdf, accessed January 20, 2005.

Longacre, Robert E. 1957. *Proto-Mixtecan*. Indiana University Research Center in Anthropology, Folklore and Linguistics V.

Miller, Wick R. 1983. Uto-Aztecan Languages. *Handbook of North American Indians, Volume 10, Southwest*, Ed. by A. Ortiz, 113–124. Washington, DC: Smithsonian Institution.

Rensch, Calvin R. 1976. *Comparative Otomanguean Phonology*. Language Science Monographs Volume 14. Bloomington: Indiana University.

Suárez, Jorge A. 1980. Review of Comparative Otomanguean Phonology (Rensch). *International Journal of American Linguistics* 46. 49–54.

Taube, Karl. 2000. Lightning celts and corn fetishes: The formative Olmec and the development of maize symbolism in Mesoamerica and the American Southwest. *Olmec Art and Archaeology*,

Ed. by J. Clark & M. Pye, 297–337. Washington DC. & New Haven, Conn.: National Gallery of Art/Yale University Press.

Wichmann, Søren. 1995. *The Relationship among the Mixe-Zoquean Languages of Mexico*. Salt Lake City: University of Utah Press.

Wichman, Søren. 2002. Contextualizing proto-languages, homelands and distant genetic relationship: Some reflections on the Comparative Method from a Mesoamerican Perspective. *Examining the Farming/Language Dispersal Hypothesis,* Ed. by P. Bellwood & C. Renfrew, 321–329. Cambridge: McDonald Institute for Archaeological Research.

Wills, W.H. 1995. Archaic foraging and the beginning of food production in the American Southwest. *Last Hunters, First Farmers,* Ed. by T.D. Price & A.B. Gebrauer, 215–242. Santa Fe. School of American Research.

Whorf, Benjamin L. & George L. Trager. 1937. The Relationship of Uto-Aztecan and Tanoan. *American Anthropologist* 39. 609–624.

Historical interpretations of geographical distributions of Amerind subfamilies

Larry Lepionka
College of Charleston, South Carolina

Former student and perpetual friend

As another Hal is reputed to have said: "We Happy Few . . ."

An outline of historical relationships between Amerind language subfamilies inferred from their geographic distributions, with the intent of providing testable hypotheses, is provided. It is contended that the major subfamilies of Amerind are not all independent derivations from Proto-Amerind. In brief, Equatorial and Andean languages are of independent origin in South America, whereas Tucanoan and Macro-Carib are derived from Equatorial, with the latter giving rise to Ge-Pano. Central Amerind and Chibchan are derivatives of Hokan, with Paezan derived from Chibchan. Hokan is a very early derivative of Penutian, as is Almosan at a later interval. Mosan and Keresiouan languages are derived from Almosan.

1. Introduction

Issue X of *Mother Tongue* contained an article by William Davey, in which the accompanying maps bore the above title. With due respect to his published priority, I do not hesitate to employ the same title here: It has the value of precisely describing the intended content, and a certain ironic value in that it precisely replicates a title I used myself in an unpublished study I wrote several years ago – if both of us have arrived at it independently, it must be good. The present paper is a revised version of that earlier work.

The general concept of this paper is similar to Davey's – that patterns of spatial distribution are in themselves artifacts of history, and thereby can reveal some of that history. Too much can, of course, be made of this. Conclusions drawn with consistency and with respect for Occam's Law may produce entirely reasonable results, but

without further evidence to substantiate them, they remain just that – "reasonable" – but hardly proven.

The proper purpose of such an analysis is not to obtain conclusive results, because from the geographical evidence alone that is impossible. Rather, it is to propose hypotheses that then may be tested directly against the linguistic evidence, or more indirectly, against archaeological and ecological evidence.

The conclusions proposed in this paper differ from those of Dr. Davey, but no space will be expended upon belaboring those differences. What both of us have done is to provide reasonable constructs whose merits do not lie in how they compare with one another, but in how each might be used to generate the aforesaid testable hypotheses. Put another way, the value of this paper, if it has such, is in the questions that it provokes and in the tests that may be made of it.

A brief *apologia*: The only time I have ever been accused of being a linguist was in graduate school – Hal Fleming read my paper on dialect structure in Shona and said "It reads like a linguist wrote it!" Hal is known for his enthusiasms, and I will do my best here to maintain the illusion, but the reader should be forewarned that I am not a linguist.

The data for this paper (i.e., linguistic classifications and geographical distributions) are drawn from Merritt Ruhlen's treatment (1987) of Greenberg's classification of Amerind languages. Said classification is not universally accepted. Nevertheless, it is a holistic one, an essential feature for the kind of geographical analysis herein conducted.

2. Theoretical considerations

Certain assumptions have been made, of which the first is the contrast between concentrated and scattered linguistic distributions. Within a given geographical space occupied by two (or more) language families, that family which has the more continuous distribution is assumed to be of later dispersal than a family that consists of isolated pockets. The fragmentation of the latter is, in fact, a result of the intrusion by the former. It is acknowledged that exceptions may occur, but the argument is generally valid, and has been used extensively in linguistic analyses. A corollary of this is that language diversity is greatest in the region of longest occupation.

This raises the important point that age within language families is relative. If a set of languages do hark back to a single ancestral tongue, then all such derivatives, at a given time horizon, are equally old. Obviously, however, it is possible to say that one language has temporal priority over another in terms of its presence in a given region. More importantly, language groups can be recognized as differing in age if one is derived from another.

Linguists recognize a hierarchy of linguistic relations, extending downwards from the most inclusive family to the dialect level. Under conditions of uniform dispersal one might assume a more or less continuous seriation of differences across time and space, analogous, perhaps, to an ideal Hardy-Weinberg genetic dispersal. But, as in the Hardy-Weinberg analogy, conditions are never ideal. We cannot, as of yet, state as precisely as geneticists do the factors that upset this ideal distribution, but obviously they would include such issues as communication density ("gene flow", to continue the analogy, including notably such factors as relative geographical isolation), founder effect (like unto gene drift in small populations), language borrowing (a different order of "gene flow"), adaptations to radically different natural and social ecologies (perhaps a species of "selection"), and spontaneous sound and grammatical shifts ("mutation"). It is emphasized that this is *analogy*, not *homology*, but the comparisons are still, I think, useful.

Subjected to these influences, some derived languages will retain sufficient similarities to be grouped together, at one level or another, by linguists. But other derivatives of the same proto-language may have undergone sufficient change that linguists will recognize them as separate entities. The more successful of these in terms of population growth and dispersal will become the "nodes" or "proto-languages" from which "new" language families spring. It is in this sense that some languages may be said to be "younger" than others, i.e., derivatives from the common source that have altered sufficiently to be recognized as distinct linguistic taxa, are of later origin as recognized entities.

Greenberg's classification, as expressed in Ruhlen (pp. 291–378) is a classical linguistic hierarchy of language relations. It is my contention that it is highly unlikely that the various major subfamilies of Amerind are all independently derived from the original proto-language. That is to say, some recognized subfamilies have developed out of others.

At face value, such a statement could be taken as a negative critique of the Greenberg classification. It is not so intended. Hierarchies are limited by their structure as to the information they may express, and linguists are limited by the nature of their data. In defining taxa they necessarily must maintain some consistency, or uniformity, in how they go about recognizing those distinct taxa. If language A and language B are very similar to one another, while language C is distinct in many ways from both, the linguist would be hard pressed to classify them as "equally" related. There is some room to maneuver here, some of it arbitrary, in terms of which elements of language a given linguist believes to be most indicative of genetic relationship, and some standard patterns of differentiation that are recognized, e.g., creole languages. Low country Gullah is an English language, despite the fact that 99% of English speakers cannot understand it.

Let us consider another biological analogy: *Afrotheria*. Classical taxonomy, based upon anatomical homologies, could not and did not arrive upon this taxon. DNA

analysis and the subtleties of cladistics has made it possible. But let us give full credit to those classical methods and their practitioners – they did recognize the counter-intuitive linkage of elephants and hyraxes and sirenians, and one has the sense that they were discomfited by the odd status of "disconnected" animals like aardvarks and by such catch-all groupings as *Insectivora*. And now aardvarks and certain "insectivores" have gained their rightful status with their proboscidean and marine cousins, not to mention the little hyraxes.

I know of no linguistic equivalents to DNA, but perhaps there is some possibility of employing the concepts of cladistics within linguistics. I would assert that the classical methods of linguistics (*sensu latus* – I do include Greenberg's among these) have a great deal of sensitivity if used with insight and some degree of flexibility, and that just as classical biological taxonomy accomplished great things, so too can the classical methods of linguistics.

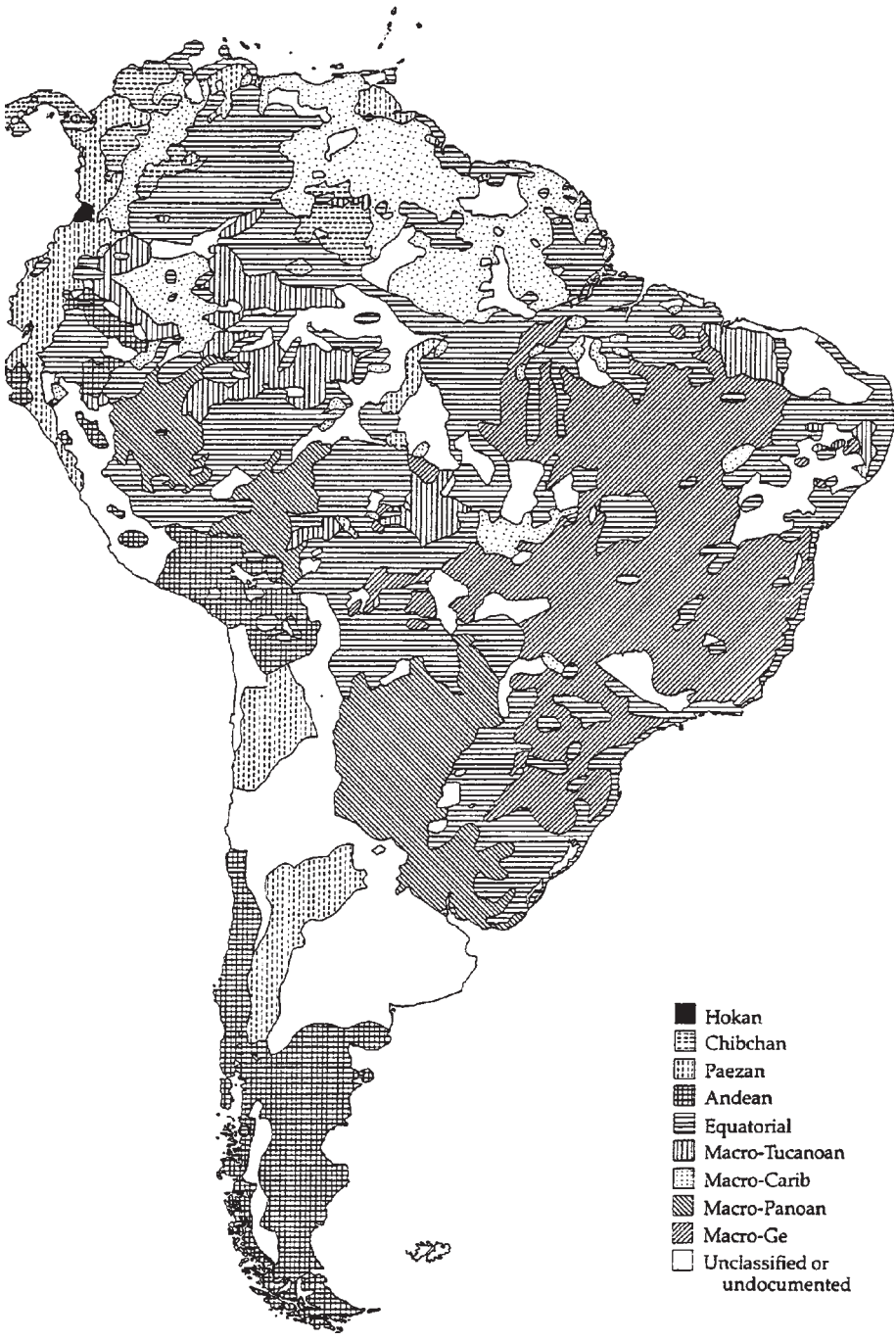
The linguist has to know where to look. As noted above, the speculative relationships that are posited in this paper may or may not be of substance (experts in particular fields may even find some absurd), but their purpose is to direct linguistic (and other investigations) toward particular hypothesized linkages.

3. Analyses of Amerind subfamilies

The units of analysis for this study are either first or second tier groupings as presented in Ruhlen 1987. These are, in order of presentation in this paper:

Equatorial-Tucanoan
Ge-Pano-Carib
Andean
Chibchan-Paezan
Central Amerind
Hokan
Penutian
Almosan-Keresiouan

These classifications are taken at face value. It is assumed that each of these is a linguistic taxon in the classical sense; i.e., each of them originates from a unitary proto-language spoken at some point in time within a specific region by a more or less homogeneous population. Subtaxa will be referenced as is pertinent within the analysis of each group. As a matter of convenience, geographical distributions as depicted in Ruhlen's maps are referred to in the ethnographic present. The maps (South America and North America) are replicated in this paper, as their presence is essential if my


Hokan
Chibchan
Paezan
Andean
Equatorial
Macro-Tucanoan
Macro-Carib
Macro-Panoan
Macro-Ge
Unclassified or undocumented

arguments are to be followed. For reasons of space and in order to allow the geographical arguments to stand on their own, references to other sources of information are kept to a minimum. As one aspect of the geographical analysis, however, certain generic ecological adaptations are postulated.

3.1 Equatorial-Tucanoan

This phylum, ubiquitous east of the Cordillera and north of the southern cone, is posited to be the oldest extant group in South America (see qualification under discussion of Andean). This claim is based upon a distribution pattern consisting of isolated pockets and peripheral locations (especially those on the Atlantic coast and in the Caribbean). It has clearly been truncated by Macro-Carib in the north, by Macro-Ge in the center, and by Macro-Panoan in the south, and on the eastern slopes of the Cordillera by Chibchan-Paezan and Andean from Equador through Bolivia.

If this hypothesis is true, then Equatorial-Tucanoan should have greater internal diversity than other South American subfamilies. Such an analysis will be complicated by the presence of diversity arising from the truncated distribution itself. The linguist must attempt to distinguish between early differentiation and post-truncation changes. Lexical borrowings of known origin, for example, should be excluded in arriving at an estimate of overall variation.

A critical element in early prehistoric settlement was ecological adaptation. The distribution pattern of Equatorial-Tucanoan is riverine and coastal, with both regions frequently characterized by rain forest. There is very limited expansion into the Cordillera, and virtually none on its west slope; nor is there any present evidence for these languages in the Pampas or Patagonia.

In particular, this group is concentrated along the Amazon, its tributaries, and the Orinoco valley. This argues for an early and continuingly successful adaptation to the rain forest/riverine environment, which served as origin point for further expansion to the south and east. Evidence for early complex cultures in the Amazon valley (e.g., Gibbons, *Science* 248: 1488–1490, 22 June 1990) suggests a possible mechanism for the expansion of populations in this region. It is proposed, on the basis of current distributions, that those populations spoke earlier versions of Equatorial-Tucanoan languages. If the Amazon valley is the center of dispersal, then the greatest linguistic diversity should be there, with lesser diversity to be found toward the south and southeast.

This is borne out to some extent by the presence of both Equatorial and Tucanoan languages in the western Amazon valley. Tucanoan is otherwise restricted to two small enclaves in the bulge of Brazil, surrounded by Equatorial languages. Even in its central area of distribution Tucanoan is truncated by Equatorial languages, suggesting that it is the older language variant. However, for reasons cited below, this hypothesis is rejected.

This concentration in the western Amazon valley of both sister taxa raises an interesting question. How is it that the language subfamily herein posited as the oldest

in the continent has its homeland of origin in the center of that continent? How did they get there?

The obvious answer, with respect to the continent as a whole, is via the isthmus of Panama. The alternative (leaving out unsubstantiated Atlantic crossings) is the western maritime route. The attraction of this model to its supporters is that rich coastal ecologies would have sufficed to meet all the needs of migrant populations, and expansion would be limited to the two dimensional world of the coast, with the unknown interior largely ignored. It has three problems: (1) Passage along a thousand miles or so of glaciated coast in southern Alaska, British Columbia, and Washington; (2) potential evidence of sites have been largely flooded out by postglacial sea level rise; and (3) lack of any other physical evidence. If I seem to be belaboring the point, let it be noted that continental migration is a critical assumption later in this paper.

It is assumed here that migration southward was continental, on both sides of the Cordillera (for which, it will be admitted, there is also no evidence). It is not unreasonable to expect that there were hundreds, in time thousands, of prehistoric Balboas who had their moments of "wild surmise" upon a peak in Darien. There is no reason to believe that one coast had priority over another. Once in South America, it is clear that the Cordillera is a massive barrier, separating populations and their migratory routes, but neither the Pacific nor the Caribbean coasts have any logical priority over one another.

We posit that there was migration eastwards, taking advantage of the coastal ecology, which indeed is rich. Equatorial languages are still spoken around Lago de Maracaibo and, importantly, westwards in the vicinity of the mouth of the Orinoco. Adaptations to coastal ecologies are logically transferable to riverine ecologies, and the headwaters of the Orinoco are contiguous with the headwaters of the Rio Negro, the major northern tributary to the Amazon.

This coastal habitation then provides access to two regions – the central Amazon valley and to the Caribbean. The major center of dispersal for Equatorial-Tucanoan languages may have been the central Amazon, but that area itself was derived from the Orinoco valley and the Caribbean coast. This suggests that Macro-Arawakan is the elder and therefore presumably more diversified taxon, from which all else is derived, with Tucanoan a presumed very early western Amazon offshoot. Caribbean Island Arawakan is either an early or late derivative of coastal Arawakan. Time restraints for this dispersal are perhaps best established by archaeology. Geographical logic, however, demands that it occurred prior to expansion of Macro-Carib.

3.2 Ge-Pano-Carib

The distribution of Ge-Pano-Carib is much more compact than that of Equatorial-Tucanoan, strongly suggesting that it is the result of a later series of three expansions. The most compact of these is Macro-Ge, located in the eastern bulge of Brazil, where it encloses several Equatorial pockets or confines them to the Atlantic

shoreline. It stretches south to make virtual contact with Macro-Panoan, though again an Equatorial enclave intervenes.

Macro-Panoan has a trilateral distribution, but each unit is itself compact (southeast Brazil to Paraguay, Bolivia, and the eastern slope of Peru into Brazil). The blank area of the map in northern Argentina was possibly largely occupied by Macro-Panoan speakers, which would have lent more continuity to the distribution of this language group.

Macro-Carib is distributed in Venezuela, the Guianas, and adjacent Brazil, with one outlier on the east slope of the Cordillera in Colombia, a second located towards the Equador-Peru border, and another in central Brazil, adjacent to Macro-Ge. In its primary area of distribution it also encloses several Equatorial languages.

Ge-Pano-Carib has spread at the expense of Equatorial-Tucanoan, as is evident in the numerous relict enclosures and peripheral coastal zones, and is therefore a later dispersal. This expansion is possibly attributable to a superior technological adaptation to the savanna environment, since all three subgroups are concentrated in this ecological zone. (Alternatively, this may have initially been a forced adaptation by peripheral populations excluded from rich riverine resources.)

Macro-Carib is posited to be the earliest of the subgroups, with Ge-Panoan its derivative (and "younger" sister group in the classificatory hierarchy). The differentiation on linguistic grounds is supported by the greater geographic splintering of Macro-Carib, a factor that may result from greater age.

The region of origin of Ge-Pano-Carib then probably lies within the principal Macro-Carib zone, i.e., Venezuela and the Guianas. Major expansion began with southern movement into the Brazilian highlands, where Ge-Panoan differentiated from the ancestral Carib language, with expanding populations to the east (Macro-Ge) and to the south and west (Macro-Panoan).

The origin of the subfamily thus becomes the question of the origin of Macro-Carib. Some offshoot of an Equatorial language would seem to be the most likely candidate. Chibchan-Paezan is not excluded, but such an origin would shift the probable homeland westwards to the eastern slope of the Cordillera in Colombia, hardly a savanna environment.

This analysis is rife with potential hypotheses. Is Ge-Pano-Carib less linguistically diverse than Equatorial-Tucanoan? Is Ge-Panoan less diverse than Macro-Carib? Does Macro-Carib have any special affinities with Equatorial or with Chibchan-Paezan languages? Archaeologically, is the occupation of the savanna and highland ecologies later than riverine habitation?

3.3 Andean

The claim of greatest antiquity for Equatorial-Tucanoan is challenged by the status of Andean languages. The Panama isthmus joins South America on the western side of

the Cordillera but, as noted above, this does not require that western migration has any significant priority of over eastern migration along the Caribbean shore. Nevertheless, the largest area of continuous Andean language distribution is in the southernmost part of the continent, and this in itself suggests a great time depth, as do early dates in Chile (Monte Verde).

The likely route of migration would have been along the western slope of the Cordillera, with a possible early concentration on coastal resources. Much of this route is inhospitable, thereby requiring larger foraging areas and consequently accelerating the rate of migration. This western route is supported by the present geographical distribution which, except in Patagonia, is concentrated on the western slope.

There are two alternative homelands for Proto-Andean, the zone of contiguous occupation in southern Chile and Patagonia, or Peru. Both areas have sizeable blank spots where no data are available. It is suggested, however, that the southern origin is more likely. If so, the languages of this region should have greater diversity than those that extend from Bolivia through Peru into Ecuador.

The geographical justification for the southern homeland is the distribution of Paezan languages, extending southwards from Colombia through Equador and into northern Chile and Argentina. It is suggested that this was once a continuous distribution (aside from the ubiquitous pockets of Equatorial languages in the north).

Peru and its peripheries (hereafter shortened to "Peru") were, from an early time, a region of high culture. The dispersals of Akkadian, Aramaic, and Arabic in the Near East possibly serve as a model for linguistic events in Peru, where the wealth and productivity of a region served as a magnet for outside groups (a similar phenomenon occurred in Mexico), including peoples speaking southern Andean languages.

We propose the following mode for the linguistic occupation of Peru with respect to known language groups: (1) Ancestors of populations that were to develop Proto-Andean in the far south passed through; (2) Equatorial language speakers infringed from the east (at one point even reaching the Pacific coast); (3) Paezan speakers moved southward, as far as northern Chile and Argentina, absorbing or maginalizing relict groups from (1) and (2) above; (4) migrant southern (Andean) and eastern (Macro-Panoan) peoples moved in, the latter on the eastern periphery and the former taking over much of the western slope; and (5) subsequent internal population shufflings brought about through the activities of state systems.

The critical element here is the relationship between Andean and Paezan. One mode of testing their relative positions would be to examine loan words. Word borrowing would of course be an ongoing process, and the languages of dominant state systems would be a major source of such loans. However, if there is any way to distinguish earlier borrowings from later ones, an assessment of such early loan words would be a strong indication as to which subfamily, Andean or Paezan, had priority in the region. Place name analysis is also in order, and in fact should be a potent tool for examination of occupational priority among all language groups.

In earlier discussion it was maintained that Ge-Pano-Carib is derivative from a deviant Equatorial language. No such argument can be made for Andean; it is a co-equal descendant, along with Equatorial-Tucanoan, from some stage of Proto-Amerind differentiation – in biological terms, it is a true sister group.

3.4 Chibchan-Paezan

This language subfamily originated from node formation in Meso-America, with a very early split that yielded Proto-Paezan in South America. That is to say, it is suggested that Chibchan has temporal priority over Paezan, with the latter developing from Chibchan. Thus we have older and younger "sister" languages. Chibchan has three discontinuous areas of distribution: The southwest coast of Mexico (approximately Jalisco through Michoacan), Central America from Honduras and El Salvador through Panama, and a section of the Caribbean coast and interior of Colombia. It is suspected that the Meso-American enclaves were at one time continuous, and have been truncated by subsequent Central Amerind and Penutian penetrations. Chibchan may at one time have extended considerably farther north in Mexico. In contrast, the South American enclaves are the result of later Chibchan migrations, following the establishment of Paezan, with the former truncating the latter along the Caribbean coast.

The homeland of Chibchan is posited to be Mexico, though not necessarily in the same location as its present distribution. The source from which Proto-Chibchan was derived is possibly Hokan. Hokan languages are present along the north Gulf coast, in Baja California, and as small relict territories near the Guerrero-Oaxacan border, northern Honduras, and the Pacific coast of Colombia. Like Chibchan, Hokan has undergone fragmentation through Central Amerind and Penutian penetrations, but Hokan distribution seems to have suffered from the earlier Chibchan expansion as well.

Such a derivation raises the question as to what ecological or social factors might have occurred to change a given language to the degree that it becomes the node or proto-language of a taxon that is sufficiently distinct to be recognized as a separate unit. We have encountered this problem earlier, with the proposal that Ge-Pano-Carib is derivative from an Equatorial language. In the latter case, it was suggested that an adaptation to the savanna environment may have been an instrumental factor. In the case of Hokan and Chibchan, possibly the latter results from cultural adaptations commencing with the earliest food production.

If the hypothesis of derivation from Hokan is correct, linguistic analysis can possibly determine if the two taxa have a "special" relationship, i.e., are closer to one another than they are to other Amerind language groups (but see discussion of Central Amerind below). As for the timing and circumstances, investigation should be made of terms associated directly or indirectly with agricultural economies.

Paezan has a far-flung distribution, ranging from central Chile and Argentina in the south to Florida in the north, with patches along the Caribbean coast, and one enclave deep in central Brazil. The homeland of Proto-Paezan was certainly the Colombian coast, from which they moved both east and south. The Caribbean distribution truncated an earlier Equatorial distribution, and was later itself truncated by the expansion of Macro-Carib.

There is no logical explanation for the central Brazilian manifestation of Paezan, but similar examples of far-flung isolates occur elsewhere in the Americas – southwestern Na-Dene or Penutian Zuni. People did on occasion just up and move long distances. We have the historic example of the North Carolina Tuscarora moving north to rejoin their linguistic brethren.

Primary movement was toward the south, along the Pacific coast. As has been previously argued, it is suspected that at one time this distribution was continuous, extending southwards from its origin in Colombia as far as central Argentina and Chile, with subsequent truncation of this distribution by Andean languages. If this pattern is real, then northern Paezan languages should have a higher degree of diversity than those in the south.

Florida? How did they ever get to Florida? There are two possible routes. The easternmost known extent of Paezan is to the Venezuelan coast adjacent to Trinidad. From here short voyages could follow the Lesser Antilles up the Greater Antilles, and make a transit to Florida from Cuba, a voyage that has frequently been made with all kinds of improbable craft in our own time.

The argument against this is that the Antilles were already occupied by speakers of Equatorial languages, who no doubt did use this route (as did later Caribs). But if Paezan passed intact through these inhabited islands, they have left no evidence. This is a bit uncertain because of the early demise of Caribbean populations, with little note taken of their linguistic habits. Nevertheless, it is unlikely that Paezan passed this way.

A stronger argument consists of the cultural resemblances between the north Colombian coast and the North American Atlantic coast from Florida through South Carolina. Fiber tempered pottery is present on the Colombian coast as early as 3100 B.C., constituting the earliest ceramics in that region. Similarly, fiber tempered wares are the earliest ceramics on the American southeast coast (or, for that matter, anywhere else in the country), extending back as far as 2500 B.C. Site distribution in both regions is coastal and riverine, with extensive exploitation of shellfish. The shell was itself utilized, in the construction of large ring structures. These rings are likewise present along the southeast coast, where they are co-extensive with the distribution of fiber tempered pottery (at least on the coast; so far as I am aware, shell rings are not associated with the interior riverine sites).

No convincing explanation has been offered concerning shell ring function. It is evident, however, that they are purposeful constructions, and presumably are the

result of some ideological aspect of the culture. There is certainly nothing unique about exploitation of coastal resources, and it could be argued that both fiber tempered pottery and shell rings are coincidental and independent developments. This becomes less convincing when we add to the equation the presence of a Colombian Paezan language in Florida (Timucua), and when we consider the congruence of dates between the two regions.

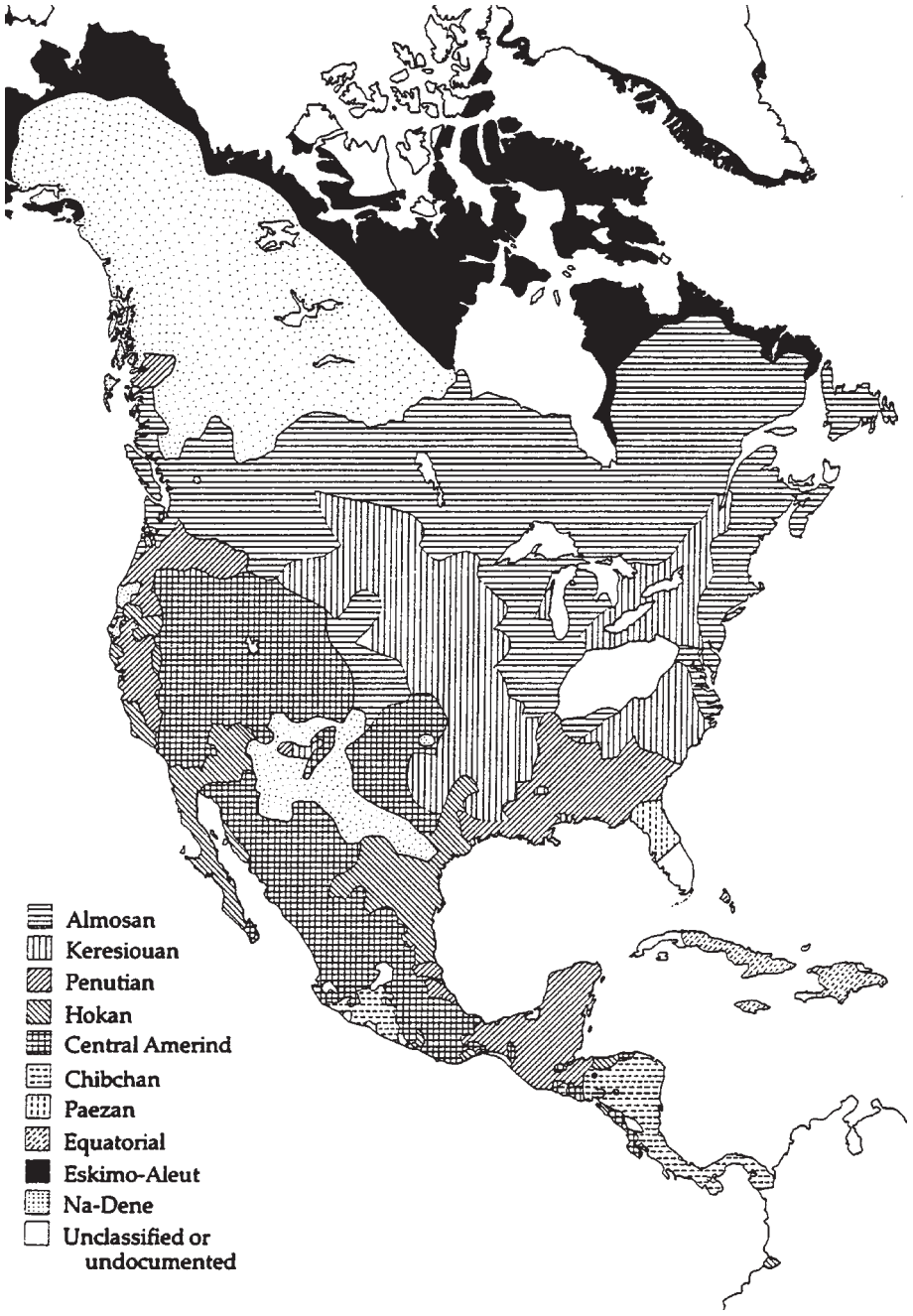
Prevailing ocean currents also fit into the pattern. The flow that becomes the Gulf Stream is westerly along the Colombian coast, and shifts northward as it approaches the Gulf of Mexico, finally turning eastward to swing around Cuba and then move northward along the Florida Atlantic shore.

Voyages may have been, and probably were, accidental. Large sea going vessels were no doubt confined to later periods, but people with a culture of coastal exploitation would have at least had simple canoes. With fishing gear on board, a modicum of mild weather, and perhaps a fortunate rain squall or two for drinking water, it is entirely feasible that such vessels could make the journey.

How could such minimal groups have had any significant influence in an already occupied land? – enough influence to maintain their language over the generations and to spread their cultural accomplishments of shell rings and fiber tempered ceramics? There is no way to arrive at any certainty on such issues, but the historical presence of Timucua in Florida and the relatively sudden appearance of pottery and shell rings farther north, with no other known antecedents, certainly did occur. A probable involved factor was that local populations were small and dispersed, practicing a Late Archaic foraging culture. There would be no significant food production for another two thousand years in this region. Ceramics are more commonly associated with Neolithic and later societies, but once introduced were possibly perceived by aboriginal populations as a valued innovation. Utilization of estuarine resources in the region allowed for and some ways required more permanent habitations that were more amenable to the use of heavy and fragile objects. Typical vessel types are moderately large, open pieces, and may have been useful for short term preservation of shellfish. The diffusion of an ideological concept such as shell rings is not so easily rationalized, but they certainly had the raw material in abundance.

3.5 Central Amerind

Central Amerind is continuously distributed throughout the southwestern United States and into Mexico as far as Tehauntepec, with one major break caused by subsequent migrations of Na-Dene speaking peoples. Proto-Central Amerind developed in the American southwest, possibly as part of cultural adaptations to increasingly arid postglacial conditions, a process greatly accelerated by the initiation of food production. The incursion into central Mexico is analogous to the pull of high cultures


Almosan
Keresiouan
Penutian
Hokan
Central Amerind
Chibchan
Paezan
Equatorial
Eskimo-Aleut
Na-Dene
Unclassified or undocumented

in Peru, with such groups as the Aztec retaining some traditions of their northern origins. Limited penetrations beyond Tehuantepec were late movements. On the whole, they did not prevail against the already resident Penutian populations in Yucatan and Guatemala, whereas they did come to dominate former Chibchan regions, reducing the areal distribution of the latter.

The most interesting relationship, however, is with Hokan. Pockets of Hokan speakers are found in Colombia and Honduras, on three sides of Central Amerind in Mexico, and in California. The geographical evidence strongly suggests that Central Amerind has displaced Hokan throughout much of this region. It also suggests that Proto-Central Amerind is derived from a Hokan language. Evidence of this must be sought linguistically, beginning with the northern expressions of Hokan. As in all such cases, there is every possibility that the particular Hokan languages that are closely related to the one that metamorphized into Proto-Central Amerind are no longer extant.

The northern origin is bolstered by the distribution of the three major subdivisions of Central Amerind. Tanoan is limited to the United States, Uto-Aztecan is present there and in Mexico, while Oto-Manguean is Mexican. The Aztec retained some memory of their northern origin, indicating a relatively recent migration, thereby placing all or most of two of these subdivisions in the northern region.

3.6 Hokan

Hokan originated in North America west of the Cordillera, and at one time was continuously distributed from California to the Gulf of Mexico and throughout Meso-America. Proto-Hokan and its immediate progeny are the closest approach that we have to western Paleo-Indian. The hunters who left behind their exquisitely shaped blades at Clovis and Folsom mumbled to each other in a Hokan dialect as they butchered their prey.

This is, of course, a rather extreme claim, but consider the present distribution: Pockets in north to central and south California; all of Baja California, with an extension into southern Arizona and the coast of Sonora; the largest enclave, on the Texas coast, extending into Mexico as far as the Vera Cruz border, with deep interior penetration along the valley of the Rio Grande and its hinterland; two small pockets on the Guerrero-Oaxaca and Oaxaca-Chiapas borders, on the Pacific coast, with a third in El Salvador; the north coast of Honduras; and, finally, a small enclave on the Colombian Pacific coast. This distribution encloses and contains that of Central Amerind and, for that matter, Chibchan. It has earlier been proposed that these subfamilies are, in fact, derivative from particular unknown Hokan laguages, with Central Amerind having its homeland in the American southwest, while Proto-Chibchan originated in Mexico.

There is no other extant language group in the region that has priority over Hokan. Other explanations are, of course, possible, with special pleading, but it is difficult to

imagine that all those isolated enclaves are products of separate migrations. It is more in accord with Occam's razor to believe that they are the remnants of a former more or less continuous distribution.

So, then, whence Hokan?

As noted previously, we reject the concept of the Pacific coastal route of migration, and work with the assumption that initial migration into North America was via the corridor between the Cordilleran and Laurentide ice sheets. It is acknowledged that physical evidence for this is minimal, but such evidence is nonexistent for other modes of approach. In any event, this is not the place to argue this particular question. For purposes of this paper we have assumed it to be true.

Once past the southern edge of the glacial sheet, one is confronted on the west with high mountain ridges and continuing alpine glaciation; on the east there is open land, rolling plains with occasional rugged uplands and, as one moves south, one river after another flowing east. To the east were herds of mammoth and horse, both in the periglacial tundra and in the steppe lands to the south. I do not believe there is too much question which route the migrants followed.

That is to say, eastern colonization preceded that of the west, with the eastern ridge of the Rocky Mountains constituting the line of demarcation. Early populations likely enough did occupy the periglacial region. It was, after all, the ecology to which they had become inured, and it was, at least seasonally, a rich one. Early and well validated dates in Pennsylvania support this contention.

But as groups expanded, sooner or later they would reach the tree line, and would find in the American southeast another kind of paradise, likewise full of game and plant resources. Cultural and technological innovations would be required to adjust to it, but this could be managed readily enough.

It may be suggested that this is empty rhetoric and scientific hogwash, and not without some justification. We are, however, attempting to establish here a theoretical model of settlement, which the critics are then free to lambaste as they will. And there is some evidence aside from the circumstantial (e.g., the undoubted ecological richness of the southeastern woodlands). There are early sites in the east: Meadowcroft in Pennsylvania (12,000 to 15,000 radiocarbon years), Saltville (13,000 to 14,500 radiocarbon years) and Cactus Hill (15,000 to 17,000 radiocarbon years) in Virginia, and possibly Topper (16,000–20,000 years) in South Carolina (Goodyear 2005).

Distribution patterns also suggest that the Clovis point was a southeastern innovation. Their highest density occurs in the southeast. There is also present the related Suwannee point, more or less a Clovis point without the fluting. It is suggested, albeit on the basis of woefully limited evidence, that the southeast was the heartland for the formation of the continent-wide classical Paleo-Indian period, complete with its Clovis points.

This is not to exclude the possibility that some of the earliest migrants went directly south, or that some, once they were far enough south for it to be feasible – say, New Mexico, crossed the Cordillera westwards (but see discussion of Penutian). Early dates in South America strongly suggest that the early inhabitants of that continent did not linger long in North America. So far as I am aware, however, early sites in the southwest are Paleo-Indian, complete with fluted points. If said points did indeed originate in the southeast, there is an argument that their spread to the southwest was the result of migration out of the region of origin.

If this proposition is correct then its corollary is that western Proto-Amerind (Proto-Hokan) is a derivative of Eastern Proto-Amerind, otherwise known as Proto-Penutian. Hokan and Penutian thereby constitute sister taxa, with the former a bit younger than the latter. Linguists might consider if there are any special features that unite the two, particularly ancient elements, if such can be isolated.

It is suggested that the initial Hokan adaptation was to a somewhat more salubrious southwestern environment characterized by pluvial lakes in the late glacial period. Differentiation of Central Amerind from Hokan was possibly initiated as part of the cultural response to the increasing aridity of the Early Archaic period (e.g., the Cochise Culture). Primitive domesticated maize found in the southwest also suggest that elements of the Central Amerind group may have been further defined by early adaptation of Neolithic borrowings out of Mexico.

3.7 Penutian

Proto-Penutian is the aboriginal language of eastern North America – it was eastern Paleo-Indian. Immediately above it has been referred to as eastern Proto-Amerind, but that terminology becomes a matter of definition. In one sense, Proto-Amerind was that language spoken by the very first immigrants that passed through central Beringia. The first descendants of those immigrants who passed beyond the bounds of the Cordilleran and Laurentide ice sheets possibly spoke a very similar language. The unilinear dimensions of the ice free corridor would have channeled expansion into fresh exploitation territories ever southward, such that the full passage would have been made in a minimal number of generations.

Once south of the glacier, expansion could have been multi-directional, though as noted in previous discussion, probably not toward the west (to be qualified below). As people spread out and encountered differing environmental challenges and underwent linguistic "gene drift" in their small band organizations, Proto-Amerind would have rapidly diversified. Penutian, it is suggested, is derived from the particular variations of Proto-Amerind that entered into the southeastern forest zone.

Penutian distribution includes: The southeastern United States, with continuity from east Texas to southern South Carolina, extending up the Mississippi to Missouri,

excepting most of Florida; the west coast, discontinuously from the Alaskan panhandle to southern California; an isolated pocket in the southwest (Zuni); isolated pockets in southern Vera Cruz; and a continuous distribution from Tehauntepec to Guatemala, including all of the Yucatan Peninsula.

The gulf region (i.e., southeastern United States) is, as noted above, the homeland of Proto-Penutian. Gulf languages are the subtaxon of Penutian presently occupying this region. As such, they should have the greatest degree of diversity within the subfamily.

The west coast expressions of Penutian produce some real oddities. Ruhlen (1987: 235–236) notes some of the California languages are not only Penutian, but are specifically Gulf languages, and that there are cultural attributes that further link these peoples to recent Gulf groups. One quoted source (Melton) suggests a New Orleans (or thereabouts) origin. It almost sounds like a Bush administration effort at hurricane relief. What it means, however, is that these groups are a relatively recent displacement, not part of some ancient expansion out of Proto-Penutian. The southwest isolate, Zuni, is, at least, a linguistic isolate as classified within the larger Penutian subfamily.

Nor is there any really viable explanation for the more distantly related Pacific coast Penutian languages. The Canadian, Washington, Oregon, and Plateau groups are each classified as separate groups, at the same taxonomic level as Gulf or Mexican Penutian, implying there is no particular internal linkage among them – that they are, in fact, all separately derived from Proto-Penutian.

One solution to this is that Proto-Penutian really is Proto-Amerind – i.e., the language of the people who first moved south of the glaciers. Counter to my previous argument, some of those folk did go west, at a very early time. Relief maps suggest a viable route south of the Bighorn Mountains, thence to the Snake River valley and its salmon fishery. Once on the coast, there could be movement south into California and, as the glaciers retreated, movement north, in both cases seeking further salmon fisheries.

If this is the case, these northwestern languages should have as high a degree of diversity as Gulf languages. A study of place names and of borrowings by Na-Dene and Mosan speakers is in order, to see if there is any evidence for priority of the Penutian speakers in the region.

Linkage between Gulf and Mexican Penutian languages is geographically more straightforward. All that separates them is a stretch of Texas and North Mexico coast line that came to be occupied by Hokan speakers, probably rather sparsely. Much of this territory is marginal land, and never would have been densely occupied. Nevertheless, population pressure in the ecologically rich southeast might have made this Texas route a natural avenue of expansion. The unusual phenomenon of Poverty Point offers no direct evidence of southern connections, but certainly proves the existence of far-flung connections at an early date. Possibly this Texan-northeast Mexico territory was shared by Hokan and Penutian speakers in earlier times.

Subsequent events (expansion of Central Amerind) led to the truncation of the Vera Cruz Penutian areas from Mayan Penutian in Yucatan.

3.8 Almosan-Keresiouan

Almosan-Keresiouan stretches across the entire breadth of North America. Taken as a single unit, it is the most contiguously organized of all the Amerind subfamilies. In this, it resembles Na-Dene with its monolithic spread across the northwest (ignoring for the nonce its southwestern offshoot), or the apparent unity of Eskimo-Aleut across the far north. And for much the same reason. All three expansions are post-glacial in date, and all three are relatively late.

Almosan-Keresiouan is, in fact, a response to glacial retreat, an occupation of the now exposed northern lands. As such, it was initially a movement out of the greater southeast (i.e., a rather broader territory than that occupied by Penutian languages in later times). Proto-Almosan-Keresiouan was derived from some Penutian stock that was involved in the earliest phases of this northward movement. More properly, we should refer to this language stock as Proto-Almosan, as it is our contention that Mosan and Keresiouan languages are subsequent developments out of Almosan, constituting "younger" sister branches.

Proto-Almosan migration was initially directed primarily toward the northeast. Moving northward into Canada, they expanded as far as the tree line, in time filled the Atlantic coast, and moved westward through Canada. Occupation of northern Quebec (Ungava) was delayed, as glacial conditions lingered there for a considerable time. Westward migration was hastened by the more limited resources of the northern woodlands, succeeded by the prairies, both of which required larger exploitation territories for group survival.

The first great barrier to western expansion was the western Cordillera, which is still partially glaciated today. Just beyond those ridges, however, lie the headwaters of the salmon streams of the west slope, notably, the Fraser River, a stream whose entire valley is still occupied by Mosan speakers. Once having reached the coast, they presumably truncated what had been a continuous Penutian distribution, leaving only Tsimshian in the north. It is suggested that the crossing of the Cordillera and the resultant isolation from the east is the event that marks the split between Almosan (eastern) and Mosan (western) language groups. This is borne out by the presence of the Blackfoot, certainly a far western tribe, but one on the eastern slope of the Cordillera, whose tongue is the sister language of Algonquin.

If the western migration had continued as a response to glacial retreat, and that retreat was closely followed, it is likely that Mosan speakers arrived on the coast prior to speakers of Na-Dene, unless the latter came by sea. There are no other geographical

indicators of priority. Again, loan word and place name analysis might be useful in resolving this question.

The major conundrum in Almosan-Keresiouan distribution is the splitting of Keresiouan into western and eastern divisions, with an outlying Almosan region west of the primary Keresiouan territory. Our suggestion is that Proto-Keresiouan is a late phenomenon beginning with the introduction of plant domestication and resultant agrarian settlement of the river valleys of the Great Plains. Proto-Keresiouandeveloped out of some Almosan language, probably somewhere along the upperMississippi drainage.

The isolated Almosan speaking region in the general vicinity of Colorado could have been derived from earlier generic Almosan expansion across the Plains, prior to the differentiation of Keresiouan. More likely, however, it arises from elements of Northern Almosan who moved into this region at later times. These western languages (e.g., Cheyenne, Arapaho) are closely related to Northern Cree and to languages of the American midwest.

The eastern contingent of Keresiouan are the Iroquoian languages, stretching from the St. Lawrence valley on the north through New York and, with a minor break, into Virginia, the Carolinas, and immediately adjacent areas. Siouan languages are present in the Carolinas and as small enclaves elsewhere. It is suggested that these were all relatively late and massive movements of whole peoples that succeeded in cutting deeply into both Penutian and eastern Algonquin territories.

Acknowledgments

The author wishes to thank Dennis Adams for editorial commentary and Emmitt F Bufkin for technical assistance in the production of this paper. I also appreciate very much the numerous discussions that I have had with linguists over the years, especially Hal Fleming and the late Sergei Starostin. Merritt Ruhlen assisted me with background information and sources, notably his book, when I was writing the original version of this paper several years ago, for which I am continuingly grateful I, of course, am solely responsible for the content of the paper.

References

Davey, William G. 2005. "South American Migrations." *Mother Tongue* 10. 163–171.

Gibbons, Ann. 1990. "New View of Early Amazonia." *Science* 248. 1488–1490.

Goodyear, Albert. 2005. Page proof of article for inclusion in *Paleoamerican Origins*, Ed. by Bonnichsen, Lepper, Steele, Stanford, Warren, Gruhn, Texas A&M University Press.

Ruhlen, Merritt. 1987. *A Guide To The World's Languages, Vol 1: Classification.* Stanford, Calif.: Stanford University Press.

PART V

Human origins, language origins, and Proto-Sapiens language

Current topics in human evolutionary genetics

Stephen L. Zegura
University of Arizona

Primate genomic data have become essential for the understanding of a number of topics related to our evolutionary history. For instance, they have provided new models for the actual speciation process that led to the divergence of the chimpanzee clade from our own evolutionary branch, a process that may have required millions of years and entailed extensive hybridization. New nuclear genetic data have also raised the possibility that some gene flow actually occurred between our species and other hominin groups as our ancestors colonized Eurasia. Hundreds of regions of our genome show the effects of natural selection over the last 50,000 years as people adapted to new environments during their global trek. Finally, genetic studies of the biological basis of language are accumulating rapidly and hold promise for identifying the ensemble of genetic changes responsible for what many linguists consider to be our chief behavioral apomorphy as a species … spoken language.

1. Introduction

Hal Fleming is an unusually wide-ranging synthetic scholar whose competencies range far beyond his major interests in historical comparative African linguistics and linguistic paleontology. Indeed, his use of an ethnographic approach and his appreciation for and understanding of data from archaeology, paleoanthropology, biological anthropology, and human evolutionary genetics epitomize the vision of *Anthropology* as a cooperative, holistic discipline with subdisciplines interacting synergistically to address issues of common concern. Since I am primarily an anthropological geneticist, my contribution will focus on recent developments from the field of genetics that pertain to the main topical concerns of human biology . . . the explanation of human origins and human variation within the context of modern evolutionary theory. The following paragraph provides an overview of the specific areas that I have chosen for discussion because I think Hal might find them both interesting and provocative.

Genetic data have become essential for the unraveling of the story of human evolution (Jobling et al. 2004). Recent comparisons of human and chimpanzee genome

sequence data have begun to pinpoint the genetic differences that permitted hominins to diverge from their African ape ancestors and have even provided a model for the actual speciation process (Mikkelsen et al. 2005; Patterson et al. 2006). The term "hominin" refers to all taxa on the human lineage post-dating the split from our common ancestor with our sister group, the chimpanzees. Competing models for the origin of *Homo sapiens* are now being tested with autosomal and X-chromosome data in addition to markers from haploid systems (i.e., the maternally inherited mitochondrial DNA and the paternally inherited Y-chromosome) (Templeton 2005; Garrigan et al. 2005). Details of the subsequent global dispersal of our species will soon be forthcoming at a much greater resolution than heretofore achieved thanks to the international Genographic Project (Shreeve 2006a). Genetic signals of relatively recent Darwinian natural selection at the regional and local levels have also emerged form the International Haplotype Map (HapMap) database (Voight et al. 2006; Wang et al. 2006). Other studies based on samples from the Coriell Institute for Medical Research have shown that two human genes that regulate brain size have swept to high frequencies under strong positive selection since our dispersal from Africa (Evans et al. 2005; Mekel-Bobrov et al. 2005). Although the first gene (*Fox P2*) associated with what many scholars believe to be the key human behavioral autopomorphy, spoken language, was discovered in 2001, we still do not know what specific genes this transcription factor actually regulates (Lai et al. 2001). However, it is most probably just a matter of time before the ensemble of genetic changes that facilitated the origin of human language will be identified (Marcus 2006).

2. The chimpanzee connection

The September 1, 2005 issue of *Nature* was devoted to The Chimpanzee Genome. The suite of papers underscored our close genetic kinship with *Pan*. Humans and chimpanzees differ by only 1.23% in terms of nucleotide substitutions (i.e., there are approximately 35×10^6 single-nucleotide changes between the two taxa); however, they also differ in 5×10^6 insertion/deletion events comprising approximately 3% of both genomes as well as in various chromosomal rearrangements (Li & Saunders 2005; Mikkelsen et al. 2005). Thus, Britten's (2002) prediction that 95% would be a good overall estimate for the base pairs shared between human and chimpanzee DNA demonstrated great foresight. In terms of individual chromosomes, the highest divergence was found for the Y-chromosome while the lowest divergence was found for the X-chromosome, thereby underscoring the higher germline mutation rate in males in both taxa (Li & Saunders 2005).

The crucial problem of homing in on the developmental genetic changes, both structural and regulatory in nature, that have made us human was succinctly summarized

by Carroll (2003: 249) when he asked: "How can we identify the smoking guns of human genetic evolution from neutral ticks of the molecular evolutionary clock?" In other words, what genetic and phenotypic systems demonstrate different expression regimes and/or dissimilar signals of natural selection in humans and chimpanzees? Examples that have been found so far include genes associated with the immune response against various pathogens, cell cycle progression and apoptosis, embryonic development, early brain and heart development, the peripheral nervous system, olfaction, taste, hearing, collagen formation, amino acid metabolism, reproduction (pregnancy, fertilization, and spermatogenesis), endorphin synthesis, craniofacial musculature, and speech (Varki 2001; Carroll 2003; Stedman et al. 2004; Balter 2005a; Mikkelsen et al. 2005; Zegura 2005). Additional human/chimpanzee genomic differences that may impact gene expression include DNA methylation patterns and recombination hot spots (Enard et al. 2004; Pennisi 2004a; Jorde 2005; Winckler et al. 2005). Khaitovich et al. (2004) presented data on regional patterns of gene expression in human and chimpanzee brains utilizing probes to about 10,000 human genes. Approximately 10% of the genes differed in expression in at least one region of the brain and more genes were up-regulated in humans than in chimpanzees. One of their most surprising results was that, *contra* expectation, no major change in gene expression was detected in Broca's Area. In a similar vein, Uddin et al. (2004) examined gene expression profiles in the anterior cingulate cortex which has been functionally associated with human grammar and vocal calls in non-human primates and found that the chimpanzee profile was more similar to human profiles than to the gorilla profile (Zegura 2005). Khaitovich et al. (2005) also compared patterns of gene transcription and expression for the human and chimpanzee brain, heart, liver, kidney, and testis. The brain showed the most stringent selective constraints and, therefore, the fewest differences while the liver showed the most. Moreover, genes active in the brain accumulated more changes on the human lineage in accord with previous work. The authors conclude with a cautionary note by emphasizing that regulatory changes may not be more important than structural protein changes during evolution, *contra* the landmark King and Wilson (1975) regulatory gene hypothesis. Rather, they stress that both types of changes have likely acted in concert (Khaitovich et al. 2005: 1853).

Another interesting question involves the relationship of genetic change to the actual speciation process that produced the hominin lineage. Patterson et al. (2006) recently presented a complex model for this divergence, perhaps the most famous speciation event of all (Rieseberg & Livingstone 2003). Based on about 20×10^6 base pairs of aligned sequence data from humans, chimpanzees, gorillas, orangutans, and two monkey species, the Patterson et al. (2006) model hypothesized an initial divergence date of about 10 million years ago between the proto-hominin and proto-chimpanzee lineages followed by an extensive period of hybridization that would have formed a third (hybrid) population. Subsequently, the hybrid population became extinct and eventually the human and chimpanzee lineages completed their divergence no earlier

than 6.3 million years ago and more likely in the neighborhood of 5.4 million years ago. The wide range of genetic divergence dates for different parts of the genome (i.e., a range of more than 4 million years) led to a speculative demographic scenario which would accommodate the fact that humans and chimpanzees are most closely related throughout their X-chromosomes. Thus, hybrid males may have been infertile, while a viable hybrid population could only have arisen if fertile hybrid females backcrossed to one of the ancestral populations (e.g., a population on the chimpanzee lineage), thereby producing fertile male hybrids when they transmitted X-chromosomes derived primarily from that ancestral population (Patterson et al. 2006). The postulated existence of this hybrid population has important implications for understanding the early hominin fossil record.

Indeed, one of the hallmarks of all pre-australopithecine hominin taxa is the presence of a combination of primitive apelike characteristics and more derived traits reminiscent of later hominins (Jurmain et al. 2005). Could one (or more) of these taxa actually have been a hybrid population? This would certainly explain the extensive mosaic evolution encountered in the fossil record. Especially pertinent to this discussion are the dating estimates for both the early African fossil hominins and for the human-chimpanzee splitting times based on genetic data. *Sahelanthropus tchadensis* lived sometime between 6.5 and 7.4 million years ago, a bit earlier than the majority of genetics-based divergence times for the proposed ancestral human-chimpanzee speciation. *Orrorin tugenensis* and *Ardipithecus kadabba* are both thought to have lived between 5.2 and 6.2 million years ago, even closer to the majority of genetic divergence dates. Although all three of these taxa are generally considered to be hominins and on "our" side of the human-chimpanzee bifurcation, things may be much more complicated than previously envisioned by many paleontologists (Patterson et al. 2006). Incidentally, the Patterson et al. (2006) paper exemplifies a recent trend in publishing in high-profile journals such as *Science, Nature*, and the *Proceedings of the National Academy of Sciences*, *etc*.. Although the actual paper is only 6 pages in length, the accompanying Supplementary Information runs to a full 40 pages!

The possibility of a relatively long-term hybrid population creating a rather "messy" evolutionary history for our lineage has been articulated in the past. In 2003 two papers presented a model of chromosomal speciation wherein the human-chimpanzee split represents parapatric speciation with gene flow (Navarro & Barton 2003; Rieseberg & Livingstone 2003). In fact, Rieseberg and Livingstone (2003) hypothesized that hybridization may have occurred for up to half the time of divergence between the two lineages (i.e., for up to 3 million years)!

These hybridization models based on genetic data clearly resonate with Jolly's (2001) insightful extrapolation from present-day hybrid zones in baboons to possible analogous situations in the hominin fossil record. Not only did he speculate that basal hominin populations could have produced viable and fertile hybrids among

themselves, he also entertained the possibility that basal hominins could have interbred with species of the proto-chimpanzee lineage. He also emphasized that in contrast to most standard models of speciation and hominin origins, "the fuzzy zone in which evolution is somewhat reticulate as well as divergent can be prolonged well beyond the point of adaptive, and paleontologically documented, morphological divergence" (Jolly 2001: 192).

3. The origin of *Homo sapiens*

In 1996 when Hammer and Zegura reviewed the five leading models for the origin of *Homo sapiens*, the preponderance of genetic data then available pointed to the African Replacement Model as the most likely explanatory framework for our species' inception (Hammer & Zegura 1996). According to this model, which from a genetics perspective was based primarily on mitochondrial DNA evidence, humans began in Africa sometime between 140,000 and 200,000 years ago as a result of a cladogenetic (true speciation) event from an ancestral hominin species whose identity has been hotly debated in the literature of paleoanthropology for many decades (Cann et al. 1987; Stringer & Andrews 1988; Aiello 1993). One of the major tenets of this model is that after the speciation event in Africa, *Homo sapiens* began dispersing to the rest of the world without any interbreeding with other hominin taxa that were encountered as Asia, Australia, and Europe were colonized. This absence of inter-taxa gene flow leads to a strong prediction: if the African Replacement Model is correct, there will be no ancient (i.e., pre-*Homo sapiens*) hominin genes from Europe, Australia, or Asia in today's human gene pool. The discovery of such genes would falsify the African Replacement Model.

Three of the five models in the Hammer and Zegura (1996) review permit various levels of hybridization as older hominin groups were encountered by our species and a steadily growing body of genetic evidence supports the idea that this gene flow did actually occur. Relethford (2001) concluded that the best description of our species' evolutionary history is that our ancestors and their genes came "Mostly Out of Africa." This model bears a strong resemblance to Bräuer's (1989) fossil-based African Hybridization and Replacement Model that postulates limited hybridization between migrating early humans and indigenous premodern populations and, thus, permits the existence of some non-African ancient genes in today's primarily African-derived human gene pool. The other two models that incorporate hybridization (i.e., Wolpoff et al.'s 1984) Multiregional Evolution Model and Smith et al.'s (1989) Assimilation Model) actually call for more extensive gene flow and a much higher percentage of genes that trace back to earlier Eurasian hominids in today's gene pool (Hammer & Zegura 1996). Templeton (2002, 2005) has also questioned the accuracy of models

that forbid gene flow as our species expanded its geographic range. Using nested clade phylogeographic analysis of genome-wide data, he concluded that the out-of-Africa expansion event was not a replacement event, but rather, was characterized by interbreeding (Templeton 2002). A more recent statistical analysis with an expanded genetic data base has overwhelmingly confirmed this rejection of the African Replacement Model (Templeton 2005).

Ancient DNA recovered from Neandertals has provided no evidence that Neandertal mitochondrial DNA ever entered the human gene pool (Jobling et al. 2004). In 2006 the first nuclear sequences from a Neandertal were reported and the principal finding so far is that the Neandertal Y-chromosome is substantially more different from the human and chimpanzee Y-chromosomes than are the other chromosomes, suggesting that little (if any) interbreeding occurred with other hominins (Dalton 2006). However, the hunt for Neandertal genetic material in modern human populations has really just begun. For example, it has been hypothesized that Neandertals may have contributed a gene tied to several neurodegenerative diseases to people of European ancestry (Dalton 2006). It is also germane to point out that two of the three subspecies comparisons for *Pan troglodytes* based on mitochondrial DNA data from *Hypervariable Region I* sequences showed larger observed nucleotide differences than those seen between Neandertals and humans (Krings et al. 1999). Thus, for this particular part of the genome we are genetically closer to Neandertals than two of the three common chimpanzee subspecies are to each other. This makes the possibility of gene flow from Neandertals to humans a reasonable conjecture that will be tested repeatedly as more ancient DNA is recovered. Indeed, the 24,000 year old Lagar Velho child from Portugal may well have been the result of admixture between a terminal Neandertal population and early modern humans as suggested by the authors of the original report (Duarte et al. 1999).

Templeton (2005) presents estimated coalescent times for 25 human DNA regions and all of the X-linked and autosomal markers have TMRCA (Time to the Most Recent Common Ancestor) dates older than the two haploid systems (i.e., mitochondrial DNA and the non-recombining portion of the Y-chromosome). This is to be expected because the effective population size for autosomal systems is four times that of the haploid systems while the X-chromosome's effective population size is three times that of the haploid systems. The relevant population genetics theory postulates that the systems with larger population sizes can look farther into the past and will likely have older TMRCA's (Jobling et al. 2004). Intriguingly, some of these older genes trace back to places other than Africa (Templeton 2005)! The best candidate for the "smoking gun" that may effectively demolish the African Replacement Model is the *ribonucleotide reductase M2 subunit pseudogene 4* (*RRM2P4*) system published by Garrigan et al. (2005). Their analysis concluded that this pseudogene clearly had roots in East Asia and had a most recent common ancestor approximately 2 million years ago,

close to the time *Homo erectus* was expanding into Eurasia. A global survey of modern humans for this locus led to the hypothesis that this ancient lineage was a remnant of later introgressive hybridization as modern humans emerging from Africa met archaic hominin populations in Eurasia who possessed the Asian basal lineage of the *RRM2P4* system (Garrigan et al. 2005). The work of Bräuer (1989); Relethford (2001); Templeton (2002, 2005), and Garrigan et al. (2005) all points to the same conclusion: humans today possess a vast majority of their genes from African sources; however, a few Eurasian genes have made the journey into our genome via rare gene flow events in our past.

4. The geographic expansion of *Homo sapiens*

Literally hundreds (if not thousands) of genetics-based papers have appeared in the past few decades documenting the spread of our species throughout the globe. Some concentrate on local population movements while others are regional, *intra*-continental, *inter*-continental, or global in scope. Numerous books have attempted global syntheses, often using different parts of the genome as their databases. For instance, Cavalli-Sforza et al.'s (1994) magisterial compendium, *The History and Geography of Human Genes*, was based primarily on traditional serogenetic and enzyme systems from the nuclear (autosomal) compartment of the genome. Sykes' (2001) *The Seven Daughters of Eve* used a mitochondrial DNA perspective for his maternally-based migratory scenario written for a popular audience. Wells' (2002) *The Journey of Man* gave a paternal perspective to the human odyssey by concentrating on Y-chromosome data. Oppenheimer's (2003) *The Real Eve* then combined both haploid systems for his scenario of how and when people first left Africa and their subsequent global trekking. Two major points of dispute involve the initial route taken for the first successful (i.e., lasting) expansion and the timing of this migration (Forster & Matsumura 2005). Early models favored a northward route through the Levant; however, more recent data favor a more southerly route from the Horn of Africa across or around the Red Sea to Southern Asia (Cavalli-Sforza et al. 1994; Oppenheimer 2003; Forster & Matsumura 2005). Mitochondrial DNA date this expansion at between 55,000 and 85,000 years ago (Forster & Matsumura 2005). Although Oppenheimer (2003) favors the earlier date which would predate the huge Toba volcanic eruption on Sumatra by about 10,000 years, recent Y-chromosome evidence from the Genographic Project points to a Post-Toba date of around 60,000 years ago (S. Wells, personal communication 2006).

The aforementioned five-year Genographic Project jointly funded by the National Geographic Society, IBM, and the Watt Family Foundation began in 2005 (Pennisi 2005). Under the direction of Spencer Wells, the project will involve 10 research groups from across the world. In addition to collecting DNA samples from

about 1,000 indigenous populations and ancient DNA from around the world, public participation has been encouraged. As a result, the initial goal of obtaining 100,000 samples to elucidate the migratory history of our species was surpassed in less than a year! The project has already amassed over 130,000 samples with no end in sight, particularly because of the enthusiastic response by the general public (M. Kaplan, personal communication 2006). The project was projected to cost about $40 million dollars and has already far outstripped the earlier Human Genome Diversity Project in both scope and sample collection (Pennisi 2005). Shreeve (2006a, 2006b) presents an overview of the project and lists a series of unanswered questions that the project is currently addressing. It promises to provide an heretofore unequalled baseline for anthropological synthesis that will eventually involve archaeological, skeletal, and linguistic data.

5. Recent natural selection signals in the human genome

Voigt et al. (2006) display the results of a genome-wide search for signals of recent positive selection in our species, identifying over 700 regions reshaped by natural selection within the last 5,000 to 15,000 years. These still incomplete selective sweeps provide information about the adaptation of modern humans to local, regional, and (to a lesser extent) global environmental conditions. The three HapMap population categories utilized represent people of African, East Asian, and Western European ancestry. Among those genes with the strongest signals of positive selection are reshaped genes involving taste, smell, and digestion among East Asians. The timing of these genetic changes coincides with the adoption of rice farming and may mark genetic adaptation to the Neolithic revolution as East Asians switched form wild to domesticated foods, much as a change in the lactase gene was a northern European genetic response to the adoption of dairy farming (Bersaglieri et al. 2004; Wade 2006). Other noteworthy examples of strong selection found by the Voight et al. (2006) team involve reproduction, metabolism of numerous carbohydrates, lipids, and phosphates, vitamin transport, skin pigmentation, bone morphogenesis, hair formation and patterning, brain function, and parts of the electron transport system associated with the breakdown of many pharmaceutical agents and with salt-sensitive hypertension.

Other recent genetic surveys of the human genome that have detected signals of positive, background (negative), and balancing selection include Clark et al. (2003), Vallender and Lahn (2004), Balter (2005b), Kamal et al. (2006), and Wang et al. (2006). The inescapable conclusion from these surveys is that *Homo sapiens* is still evolving biologically, especially in the realm of immune defense against novel pathogens. Wang et al. (2006) detected approximately 1800 human genes that exhibited the genetic architecture of selection in their study of 1.6×10^6 single-nucleotide polymorphisms.

They contend that most of these selective events occurred in the last 10,000–40,000 years and they speculate that gene-culture interactions directly or indirectly shaped our genomic architecture as humans spread throughout the globe and began regional shifts from hunter-gatherer to agrarian societies.

More specific system-associated examples of natural selection that have been recently documented focus on the loss of olfactory receptor genes, the evasion of malarial parasites, and tooth development (Gilad et al. 2003; Wang et al. 2003; Pereira et al. 2006). The Wang et al. (2003) study is especially important because the glycophorin loci (*GPA* and *GPB*) that produce receptors for the binding ligand(s) for *Plasmodium falciparum* also are responsible for the antigens underlying the *MN* and *Ss* blood group systems. In fact more than 40 human blood types are caused by glycophorin variation. Wang et al. (2003) propose that the rapidly evolving glycophorin system may contribute to the evasion of the malarial parasite via positive selection.

Lamason et al.'s (2005) landmark study identified a zebrafish pigment locus and its human homologue which actually accounts for between 25% and 38% of the skin color difference between Europeans and Africans (Balter 2005c). The gene is thought to be involved in cation exchange, probably across the melanosome membrane. Melanosomes are the melanin-containing organelles in melanocytes responsible for body coloration in birds and mammals. The distribution of the two alleles in humans is striking in that the evolutionarily conserved ancestral allele predominates in African and East Asian populations, while the newly derived allele associated with light pigmentation is present in 98% of the Europeans tested, thereby suggesting that there has been a recent selective sweep among the ancestors of modern Europeans (Balter 2005c; Lamason et al. 2005). These results are consistent with the hypothesis that positive selection for light skin occurs among Europeans because light pigmentation facilitates the absorption of sunshine so adequate vitamin D can be synthesized at the less sunny European latitudes (Balter 2005c).

Finally, two gene loci that regulate brain size underwent strong positive selection within the last 40,000 years (Evans et al. 2005; Mekel-Bobrov et al. 2005). Both genetic systems can lead to primary microcephaly due to recessive mutations. The *microcephalin* (*MCPH1*) locus most likely controls the proliferation and/or differentiation of neuroblasts during neurogenesis, while the *abnormal spindle-like microcephaly associated (ASPM)* locus may regulate neural stem proliferation and/or differentiation during brain development by mediating spindle assembly during cell division (Zegura 2006). The *MCPH1* locus exhibits a high frequency haplotype (i.e., haploid genotype) with a derived allele that originated *ca*. 37,000 years ago, a date that coincides with the introduction of modern humans into Europe, as well as with the increased presence of art and symbolism characteristic of the Upper Paleolithic (Balter 2005d; Evans et al. 2005). Frequencies of the selectively favored haplotype were highest in Eurasia and the Americas, but were lowest in Africa leading to speculation

that this haplotype may have originated in Eurasia due to local adaptation (Evans et al. 2005). The selectively favored derived haplotype at the *ASPM* locus originated *ca*. 5,800 years ago and has noticeably higher frequencies in Europeans and Middle Easterners (Mekel-Bobrov et al. 2005). Provocative biocultural and geographical correlates that are temporally congruent include the spread of domestication from the Middle East, the development of cities and early state formation, the origin of written language, and dramatic population size increase (Balter 2005d; Mekel-Bobrov et al. 2005; Yoffee 2005). Obviously these speculative connections are intriguing; however, it should be emphasized that their significance has neither been studied in detail nor in any way established (Zegura 2006).

6. Genes and the origin of language

The genetic secrets that led to the development of human language are still locked up in our brains. As Hauser et al. (2002) and Marcus (2006) so aptly stress, the abstract computational capacity of language may consist of a novel evolutionary reconfiguration of many ancestral components. The genetic changes responsible for the synergistic emergence of the behavioral evolutionary hallmark of our species, spoken language, were probably both regulatory and structural in nature. A suite of reshaped genetic systems is more likely than the major gene hypothesis articulated by Klein whereby a fortuitous mutation some 40,000–50,000 years ago permitted the capacity for human speech (Holden 1998; Klein 1999; Fleming 2005; Marcus 2006). Evolutionary "tinkering" and exaptation are generally thought to be more widespread in nature than are macro-mutational shifts and language may turn out to reflect this ongoing dynamic Darwinian pattern of genetic descent with modification (Marcus 2006).

Some genetic changes which may be associated with the eventual development of language have already been identified. For instance, Chenn and Walsh (2002) demonstrated that a mutation in the *β-catenin* gene has the ability to enlarge the brain of transgenic mice. The most striking changes involved the greatly increased cerebral cortical surface area and the formation of complex folds resembling the sulci and gyri of primate brains on the heretofore smooth surface of the mouse brain (Vogel 2002). Human brain asymmetry can be traced to the differential expression of 27 genes in the left versus right cerebral hemispheres (Sun et al. 2005). One of these genes, *Lim Domain Only 4* (*LMO4*), turned out to be a transcription factor that is differentially expressed in the left and right perisylvian cortex (Sun et al. 2005). This is especially germane since for most people language function is predominantly localized to a distributed network in the left perisylvian cortex (Sun et al. 2005). The most famous language-associated gene is the *FOX P2* (forkhead box P2) transcription factor implicated in both speech articulation and grammatical competence (Lai et al. 2001;

Balter 2002; Enard et al. 2002). Enard et al. (2002) made a convincing case that the gene has been a target for natural selection with the fixation of the beneficial human variant occurring sometime within the last 200,000 years. Individuals with disruption of *FOX P2* suffer from an impairment of selection and sequencing of fine motor movements. Since speech can be understood as an intricate motor activity, the ability to improve the control of orofacial movements (thereby permitting more proficiency in spoken language) may have been the phenotype under selection associated with the beneficial human *FOX P2* variant (Enard et al. 2002; Holden 2004).

The questions of how and why language evolved continue to inspire heated debate (Holden 2004). Can the origin of language be traced to gestures, vocal imitation, "motherese," seduction, gossiping, grooming, an increased ability to form and manipulate abstract symbols, the capacity for syntactic recursion, anatomical changes in the vocal tract, specific changes in the brain, and/or genetic changes and in what combination (Deacon 1997; Hauser et al. 2002; Holden 2004; Miller 2006)? The presence of mirror neurons that facilitate imitation in both Broca's Area (the ventral pre-motor cortex *in sensu stricto*) and Wernicke's Area suggests that these neurons may provide a crucial link between movement and speech control (Deacon 1997; Holden 2004; Miller 2006). The supplementary motor cortex adjacent to the primary motor cortex controls the physical constraints on vocal expression and is probably involved in the production of the click sounds characteristic of the Khoisan language family (Holden 2004). Many linguists believe that the first human language was a Khoisan click-containing language spoken somewhere in East or South Africa (Pennisi 2004b). Since Ethiopia is the location of the two oldest anatomically modern human fossil sites (i.e., Omo at 190,000–200,000 years ago and Herto at 154,000–160,000 years ago), it is an attractive candidate (Zegura 2006). Alternatively, the oldest human mitochondrial DNA (with a TMRCA date of *ca*. 170,000 years ago) comes from three present-day click-speaking groups: the Hadzabe and Sandawe from Tanzania and the San who presently reside in southern Africa but whose ancestors may have lived farther to the northeast (Pennisi 2004b). Interestingly, clicks are made without involvement of the larynx, while the rest of the sounds in the phonological repertoire of the Khoisan languages reflect the participation of the larynx as is true for all other human languages (Holden 2004). In a cultural evolutionary sense, non-click phonemes out-competed click sounds by the time people began their global diaspora.

7. A concluding addendum

In terms of *The Great Archaeological Debate on Human Origins* recently discussed in *Mother Tongue*, it seems that a substantial amount of evidence has accumulated in favor of the position taken by Brooks (McBrearty & Brooks 2000; Fleming 2005).

Humans were anatomically and culturally modern long before they reached Europe (Wong 2005). Afro-Pacific languages diverged from a Khoisan stock and were the first languages spoken outside of Africa. Indeed, people colonized Australia and many parts of Asia thousands of years before they arrived in Europe (Oppenheimer 2003). The hypothetical major genetic changes invoked by Klein (1999) to explain the supposed sudden appearance of the European Upper Paleolithic as well as human language remain hypothetical and are weakly supported at best, notwithstanding Balter (2005d) and Evans et al.'s (2005) generalizations about the *MCPH1* locus. In a similar vein, publication of Diamond's (1997) Great Leap Forward Model predated much of the African Middle Stone Age evidence that directly contradicts the position that so-called modern behavior essentially began about 40,000 years ago in Europe (Wong 2005). Of the 14 material culture components of modern behavior discussed in McBrearty and Brooks (2000), 13 have now been shown to antedate the 40,000 year threshold with only cave art images appearing after that date (Wong 2005). The shattering of the 14-element list is reminiscent of another list: the 13 design features of language proposed by Hockett (1960). If one adds the property of recursion as Hauser et al. (2002) rather emphatically suggest, we have another 14-element construction, one which has had a similar fate. Hockett (1960) asserted that 9 (or 10 in another part of the paper) of these design features were shared with other communication systems. Since then, all of the supposedly unique features of human language have also been encountered in non-human communication contexts. Gentner et al.'s (2006) recent claim that European starlings exhibit recursive syntactic song pattern learning completes the "demolishment" of the 14-member linguistic list, although I must confess that when I first read the paper I used the simpler finite-state grammar approximation rather than the center-embedding context-free grammar strategy to classify the patterns of rattles and warbles. Hal might be amused (or bemused) to discover that this author seems to be less linguistically sophisticated than some male songbirds! Hal's noteworthy synthetic faculties are also sure to be engaged as the forthcoming results of the Genographic Project provide comparative data for a long-awaited forging of a global consilience of human prehistory (Wilson 1998). I look forward to your continued musings, Hal … and hearty congratulations for the past 80 intellectually stimulating years!

References

Aiello, L.C. 1993. "The fossil evidence for modern human origins in Africa: A revised view." *Am. Anthropol.* 95. 73–96.

Balter, M. 2002. "'Speech gene' tied to modern humans." *Science* 297. 1105.

Balter, M. 2005a. "Expression of endorphin gene favored in human evolution." *Science* 310. 1257.

Balter, M. 2005b. "Are humans still evolving?" *Science* 309. 234–237.
Balter, M. 2005c. "Zebrafish researchers hook gene for human skin color." *Science* 310. 1754–1755.
Balter, M. 2005d. "Are human brains still evolving?" *Science* 309. 1662–1663.
Bersaglieri, T., Sabeti, P.C., Patterson, N. & 6 co-authors. 2004. "Genetic signatures of strong recent positive selection at the lactase gene." *Am. J. Hum. Genet*. 74. 1111–1120.
Bräuer, G. 1989. "The evolution of modern humans: A comparison of the African and non-African evidence." *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans*, Ed. by P. Mellars & C. Stringer, 123–154. Edinburgh: Edinburgh University Press.
Britten, R.J. 2002. "Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels." *Proc. Natl. Acad. Sci. USA* 99. 13633–13635.
Cann, R.L., Stoneking, M. & Wilson, A.C. 1987. "Mitochondrial DNA and human evolution." *Nature* 325. 31–36.
Carroll, S.B. 2003. "Genetics and the making of *Homo sapiens*." *Nature* 422. 849–857.
Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. 1994. *The History and Geography of Human Genes*. Princeton: Princeton University Press.
Chenn, A. & Walsh, C.A. 2002. "Regulation of cerebral cortical size by control of cell cycle exit in neural precursors." *Science* 297. 365–369.
Clark, A.G., Glanowski, S., Nielsen, R. & 14 co-authors. 2003. "Inferring nonneutral evolution from human-chimp-mouse orthologous trios." *Science* 302. 1960–1963.
Dalton, R. 2006. "Neanderthal DNA yields to genome foray." *Nature* 441. 260–261.
Deacon, T.W. 1997. *The Symbolic Species*. New York: W.W. Norton and Company.
Diamond, J. 1997. *Guns, Germs, and Steel*. New York: W.W. Norton and Company.
Duarte, C., Maurício, J., Pettitt, P.B. & 4 co-authors. 1999. "The early Upper Paleolithic human skeleton from the Abrigo do Lagar Velho (Portugal) and modern human emergence in Iberia." *Proc. Natl. Acad. Sci. USA* 96. 7604–7609.
Enard, W., Przeworski, M., Fisher, S.E. & 5 co-authors. 2002. "Molecular evolution of *FOXP2*, a gene involved in speech and language." *Nature* 418. 869–872.
Enard, W., Fassbender, A., Model, F. & 3 co-authors. 2004. "Differences in DNA methylation patterns between humans and chimpanzees." *Current Biol*. 14. R148–R149.
Evans, P.D., Gilbert, S.L., Mekel-Bobrov, N. & 6 co-authors. 2005. "*Microcephalin*, a gene regulating brain size, continues to evolve adaptively in humans." *Science* 309. 1717–1720.
Fleming, H.C. 2005. "The great archeological debate on human origins." *Mother Tongue* 10. 14–39.
Forster, P. & Matsumura, S. 2005. "Did early humans go north or south?" *Science* 308. 965–966.
Garrigan, D., Mobasher, Z., Severson, T. & 2 co-authors. 2005. "Evidence for archaic Asian ancestry on the human X chromosome." *Mol. Biol. Evol*. 22. 189–192.
Gentner, T.Q., Fenn, K.M., Margoliash, D. & 1 co-author. 2006. "Recursive syntactic pattern learning by songbirds." *Nature* 440. 1204–1207.
Gilad, Y., Bustamante, C.D., Lancet, D. & 1 co-author. 2003. "Natural selection on the olfactory receptor gene family in humans and chimpanzees." *Am. J. Hum. Genet*. 73. 489–501.
Hammer, M.F. & Zegura, S.L. 1996. "The role of the Y chromosome in human evolutionary studies." *Evolutionary Anthropol*. 5. 116–134.
Hauser, M.D., Chomsky, N. & Fitch, W.T. 2002. "The faculty of language: What is it, who has it, and how did it evolve?" *Science* 298. 1569–1579.
Hockett, C.F. 1960. "The origin of speech." *Sci. Am*. 203. 89–96.

Holden, C. 1998. "No last word on language origins." *Science* 282. 1455–1458.
Holden, C. 2004. The origin of speech. *Science* 303. 1316–1319.
Jobling, M.A., Hurles, M.E. & Tyler-Smith, C. 2004. *Human Evolutionary Genetics*. New York: Garland Science.
Jolly, C.J. 2001. "A proper study for mankind: Analogies from the Papionin monkeys and their implications for human evolution." *Yrbk. Phys. Anthropol*. 44. 177–204.
Jorde, L.B. 2005. "When we're hot, they're not." *Science* 308. 60–62.
Jurmain, R., Kilgore, L. & Trevathan, W. 2005. *Introduction to Physical Anthropology*. Belmont: Thomson Wadsworth.
Kamal, M., Xie, X. & Lander, E.S. 2006. "A large family of ancient repeat elements in the human genome is under strong selection." *Proc. Natl. Acad. Sci*. USA 103. 2740–2745.
Kaplan, M. 2006. Personal communication.
Khaitovitch, P., Muetzel, B., She, X. & 12 co-authors. 2004. "Regional patterns of gene expression in human and chimpanzee brains." *Genome Research* 14. 1462–1473.
Khaitovich, P., Hellmann, I., Enard, W. & 6 co-authors. 2005. "Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees." *Science* 309. 1850–1854.
King, M.C. & Wilson, A.C. 1975. "Evolution at two levels in humans and chimpanzees." *Science* 188. 107–116.
Klein, R. 1999. *The Human Career*. Chicago: University of Chicago Press.
Lai, C.S.L., Fisher, S.E., Hurst, J.A. & 2 co-authors. 2001. "A forkhead-domain gene is mutated in a severe speech and language disorder." *Nature* 413. 519–523.
Lamason, R.L., Mohideen, M-A.P.K., Mest, J.R. & 23 co-authors. 2005. "*SLC24A5*, a putative cation exchanger, affects pigmentation in zebrafish and humans." *Science* 310. 1782–1786.
Li, W-H. & Saunders, M.A. 2005. "The chimpanzees and us." *Nature* 437. 50–51.
Marcus, G.F. 2006. "Startling starlings." *Nature* 440. 1117–1118.
McBrearty, S. & Brooks, A.S. 2000. "The revolution that wasn't: A new interpretation of the origin of modern human behavior." *J. Hum. Evol.* 39. 453–563.
Mekel-Bobrov, N., Gilbert, S.L., Evans, P.D. & 5 co-authors. 2005. "Ongoing adaptive evolution of *ASPM*, a brain size determinant in *Homo sapiens*." *Science* 309. 1720–1722.
Mikkelsen, T.S., Hiller, L.W., Eichler, E.E. & 64 co-authors. 2005. "Initial sequence of the chimpanzee genome and comparison with the human genome." *Nature* 437. 69–87.
Miller, G. 2006. "Probing the social brain." *Science* 312. 838–839.
Navarro, A. & Barton, N.H. 2003. "Chromosomal speciation and molecular divergence: Accelerated evolution in rearranged chromosomes." *Science* 300. 321–324.
Oppenheimer, S. 2003. *The Real Eve*. New York: Carroll & Graf.
Patterson, N., Richter, D.J., Gnerre, S. & 2 co-authors. 2006. "Genetic evidence for complex speciation of humans and chimpanzees." *Nature* advance online publication, 17 May 2006 (doi:10.1038/nature04789).
Pennisi, E. 2004a. "The case of the disappearing DNA hotspots." *Science* 304. 1590.
Pennisi, E. 2004b. "The first language?" *Science* 303. 1319–1320.
Pennisi, E. 2005. "Private partnership to trace human history." *Science* 308. 340.
Pereira, T.V., Salzano, F.M., Mostowska, A. & 10 co-authors. 2006. "Natural selection and molecular evolution in primate *PAX9* gene, a major determinant of tooth development." *Proc. Natl. Acad. Sci. USA* 103. 5676–5681.
Relethford, J.H. 2001. *Genetics and the Search for Modern Human Origins*. New York: John Wiley & Sons.

Rieseberg, L.H. & Livingstone, K. 2003. "Chromosomal speciation in primates." *Science* 300. 267–268.
Shreeve, J. 2006a. "The greatest journey." *National Geographic* 209. 60–69.
Shreeve, J. 2006b. "Reading the secrets of the blood." *National Geographic* 209. 70–73.
Smith, F.H., Falsetti, A.B. & Donnelly, S.M. 1989. "Modern human origins." *Yrbk. Phys. Anthropol.* 32. 35–68.
Stedman, H.H., Kozyak, B.W., Nelson, A. & 7 co-authors. 2004. "Myosin gene mutation correlates with anatomical changes in the human lineage." *Nature* 428. 415–418.
Stringer, C.B. & Andrews, P. 1988. "Genetic and fossil evidence for the origin of modern humans." *Science* 158. 1263–1268.
Sun, T., Patoine, C., Abu-Khalil, A. & 6 co-authors. 2005. "Early asymmetry of gene transcription in embryonic human left and right cerebral cortex." *Science* 308. 1794–1798.
Sykes, B. 2001. *The Seven Daughters of Eve*. New York: W.W. Norton & Company.
Templeton, A.R. 2002. "Out of Africa again and again." *Nature* 416. 45–51.
Templeton, A.R. 2005. "Haplotype trees and modern human origins." *Yrbk. Phys. Anthropol.* 48. 33–59.
Uddin, M., Wildman, D.E. Liu, G. & 6 co-authors. 2004. "Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain expression profiles." *Proc. Natl. Acad. Sci. USA* 101. 2957–2962.
Vallender, E.J. & Lahn, B.T. 2004. "Positive selection on the human genome." *Hum. Mol. Genet.* 13. R245–R254.
Varki, A. 2001. "Loss of N-glycolylneuraminic acid in humans: Mechanisms, consequences, and implications for hominid evolution." *Yrbk. Phys. Anthropol.* 44. 54–69.
Vogel, G. 2002. "Missized mutant helps identify organ tailors." *Science* 297. 328.
Voight, B.F., Kudaravalli, S., Wen, X. & 1 co-author. 2006. "A map of recent positive selection in the human genome. *PLoS Biol.* 4". 446–458/e72.
Wade, N. 2006. "The twists and turns of history, and of DNA." *New York Times* March 12, 2006 p. 14.
Wang, E.T., Kodama, G., Baldi, P. & 1 co-author. 2006. "Global landscape of recent inferred Darwinian selection for *Homo sapiens*." *Proc. Natl. Acad. Sci. USA* 103. 135–140.
Wang, H-Y., Tang, H., Shen, C-K.J. & 1 co-author. 2003. "Rapidly evolving genes in human. 1: The glycophorins and their possible role in evading malarial parasites." *Mol. Biol. Evol.* 20. 1795–1804.
Wells, S. 2002. *The Journey of Man*. Princeton: Princeton University Press.
Wells, S. 2006. Personal communication.
Wilson, E.O. 1998. *Consilience*. New York: Alfred A. Knopf.
Winckler, W., Myers, S.R., Richter, D.J. & 8 co-authors. 2005. "Comparison of fine-scale recombination rates in humans and chimpanzees." *Science* 308. 107–111.
Wolpoff, M.H. 1999. *Paleoanthropology*. Boston: McGraw-Hill.
Wong, K. 2005. "The morning of the modern mind." *Sci. Am.* 292. 86–95.
Yoffee, N. 2005. *Myths of the archaic state*. Cambridge, UK: Cambridge University Press.
Zegura, S.L. 2005. "Anthropology." *2005 Britannica Book of the Year*, Ed. by KJ Sparks, 147–148. Chicago: Encyclopaedia Britannica, Inc.
Zegura, S.L. 2006. "Anthropology." *2006 Britannica Book of the Year*, Ed. by KJ Sparks, 147–148. Chicago: Encyclopaedia Britannica, Inc.

A wild 50,000-year ride

Philip Lieberman
Brown University

The study of languages in prehistory critically involves the ever changing sound patterns of human languages, in short speech. Speech is the default medium for language, and it most likely has a long evolutionary history. However, the fully human speech anatomy which allows us to produce the most common, "universal" vowels [i], [u] and [a], first appears in the fossil record in the Upper Paleolithic (about 50,000 years ago) and was absent in both Neanderthals and earlier humans. Recent genetic evidence also suggests the appearance of the neural substrate that is necessary to regulate human speech within the past 100,000 years. Thus, we have had a wild ride – creating thousands of languages since the Upper Paleolithic start-point for fully human linguistic capacity.

Most theoretical linguists, preoccupied with the study of syntax and phonology treated as though it were a branch of syntax, ignore speech. However, speech is the only aspect of language, which even closely related living species lack. Speech, moreover, has properties that allow us to transmit information at rates that make it possible to transmit and comprehend long, complex utterances form part of the neural substrate involved in syntax and other aspects of cognition.

If we are to understand why fully human phonetic ability appears in the fossil record at the comparatively recent date of 50,000 years in the Upper Paleolithic, some aspects of the anatomy and physiology involved must be reviewed.

1. Primitive and derived features of human speech

It is obvious that animals cannot talk. However, the vocal signals of all terrestrial mammals, including humans, are generated by means of similar processes. The lungs provide a current of air that for the most part is inaudible having audio frequency components below the limits of human perception. Human speech and most mammalian vocal signals are produced on the expiratory, outflow of air from the lungs. As humans talk, they use muscle groups that inflate the lungs during quiet inspiration to present a more or less level air pressure to the vocal cords of the larynx over the course of a long utterance. The air pressure would otherwise start at a high level and gradually fall, causing the vocal cords to produce phonation starting at a high fundamental frequency,

then falling throughout the course of a sentence. Comparative evidence is presently absent that would establish whether living non-human species perform these maneuvers, which involve "knowing" at some internalized mental level, the length of a sentence that one intends to utter. The larynges of all animals, from frog-like ones onwards, act as "transducers," transposing the energy of the airflow from the lungs to higher, audible frequencies through the process of phonation. The air pressure against the vocal cords (which are often termed "folds" – they are complex structures, neither folds nor cords)and tensions of laryngeal muscles determines the rate at which they rapidly open and close, releasing "puffs" of air. The rate at which these puffs of air occur is the fundamental phonation (F0). The perceived pitch of a person's voice is closely related to the average F0. These and other aspects of speech physiology and the evolution of speech are discussed in detail in Lieberman (2006).

If the activity of the larynx and lungs were all that was involved in speech production all animals would possess the physiologic bases for human speech. Variations in F0 over the course of a syllable differentiate words in tone languages and are used to convey referential information by non-human primates (e.g., Cheney & Sayfarth 1990) However, the key element to fully human speech involves the "supralaryngeal vocal tract" (SVT) – the airway above the larynx. The acoustic signal generated by the larynx during phonation is filtered by the SVT. The positions assumed by the tongue, lips, soft palate (which controls the opening between the oral and nasal passageways), and larynx determine the frequencies at which maximum energy can pass through the SVT. These frequencies, which come as "sets" owing to the principles of physical acoustics, are termed "formants." They are local potential energy maxima that continually change as the tongue, lips and larynx move Vowel and consonantal quality is largely conveyed by formant frequency patterns, enhanced by distinctions in duration (Fant 1960). As we talk, the airway above the larynx, the supralaryngeal vocal tract (SVT), continually changes its shape as we talk, producing a time-varying formant frequency pattern. Aperiodic noise generated at a constriction along the SVT as in the fricative [s] can also serve as a source of acoustic energy that is filtered by the portion of SVT above it.

2. Formant frequencies

To summarize, the larynx provides the source of acoustic energy for vowels and other phonated speech sounds; the SVT act as an acoustic filter that determines the phonetic quality of the sounds. A given SVT shape will let more acoustic energy through at a set of particular formant frequencies, local energy maxima, occurring in inharmonic combinations. The lowest formant frequency is identified by the notation F1, the next highest as F2, the third as F3. For example, the vowels [i] and [u] of the words "see" and "sue", can be produced with identical F0's, – different formant frequencies specify these vowels.

The relationship between formants, the laryngeal source and speech signals is in some ways similar to that between the source of light, the lenses of a pair of sunglasses and the tint of the world seen through the sunglasses. The lenses achieve the perceived tint by "attenuating," i.e., reducing the amount of light energy throughout a range of frequencies. Dies do not yield a single spectral peak, they produce local energy maxima at a set of electromagnetic frequencies that our visual system perceives as a color. The combination of frequencies that are least attenuated determines the tint that the sunglasses impart. The same "source" of light, sunlight will provide a blue or pink world when viewed using different sunglasses. The formant frequencies of the SVT are the acoustic frequencies that pass through it with minimum attenuation.

3. Human and non-human supralaryngeal vocal tracts

The filtering action of the SVT is determined by its shape and length in much the same manner as the tube of a woodwind instrument. It's shape and total length, the variation in cross-sectional area along its length range determine the formant frequencies that it can generate. In the 18th and 19th centuries tubes that were formed into shapes that were thought to represent the shape of the SVT were used to model the SVT. The tubes acted as acoustic filters; reeds were the source acoustic energy. Over the past half-century, computer-implemented models have been developed that can be used to systematically determine the formant frequencies that a particular SVT can produce (e.g., Stevens 1972; Baer et al. 1991). The models allow the investigator to determine the formant frequencies that a particular SVT shape would produce; the range of shapes is limited by a particular SVT's anatomy.

The human SVT after age eight years has a tongue having an almost circular sagital (midline) contour forming two segments, a "horizontal" section delimited by the portion of the tongue in the oral cavity (SVTh), and a vertical section delimited by the section of the tongue extending down the neck into the pharynx (SVTv). The human tongue when seen in profile has a round posterior contour and forma a right-angle bend at its midpoint – the junction of the oral cavity and pharynx. Movements of the undistorted tongue in the space defined by the oral cavity and pharynx can produce the abrupt midpoint 10:1 area function discontinuities necessary to produce the formant frequency patterns of the universal "point" or "quantal" vowels [i], [u] and [a] whose properties will be discussed below. In contrast, computer modeling shows that the SVTs of living primates, whose tongue are almost entirely within their mouths, inherently cannot produce quantal vowels because they cannot produce the necessary abrupt, 10:1 midpoint area function discontinuities(Lieberman, Klatt & Wilson 1969; Lieberman, Crelin & Klatt 1972).

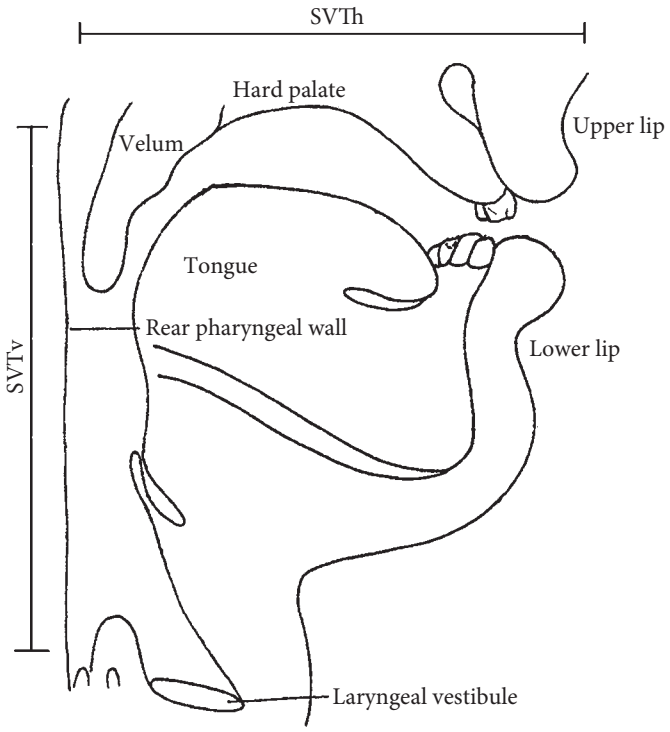


Figure 1. Adult human supralaryngeal vocal tract (SVT) in a cross-sectional view. The human tongue has an almost circular posterior (rear) contour. The "horizontal,"(SVTh)oral portion, and "vertical," (SVTv) pharyngeal portion, bounded by the tongue, have almost equal lengths. There is a natural discontinuity formed at the intersection of SVTh and SVTv, allowing speakers to form abrupt changes in the cross sectional area of the human SVT at its midpoint without material changes in the shape of the tongue.

Acoustic analyses of the vocalizations of non-human primates (e.g., Lieberman 1968; Rendall et al., in press) consistently show that they produce schwa-like vowels (the vowel of the word "bub") because their tongue are positioned almost entirely in the mouth. The soft tissue of their tongues which have a thin cross-section cannot be abruptly deformed to produce 10:1 area function discontinuities. Alternate mechanisms can be used to produce signals that to an unaided ear "sound" like quantal vowels.

4. How do [i], [u], and [a] bear on the evolution of human speech? quantal vowels

Speech communication would be possible without the point, quantal vowels that for decades have been recognized as being among the few attested universals of human

language (Greenberg 1963). Moreover, since as we shall see, the anatomy that makes it possible to produce these vowels also increases the risk of death by choking, there would have been no selective advantage for retaining the mutations involved in the evolution of the human SVT, unless some form of speech were already being used by our hominid ancestors. Stevens (1972) coined the term quantal to characterize sounds that have two useful properties. Quantal vowels and consonants (which also occur) have acoustic properties that are perceptually salient and can be produced with a degree of articulatory sloppiness They also facilitate the complex process by which we "decode" speech signals.

The task of speech production would be simpler if it were possible to produce a stable acoustic signal without having to control tongue placement precisely. The task of speech perception also would be less subject to errors if the resulting acoustic signals are maximally distinct. Quantal speech sounds meet these criteria. Quantalness can be illustrated by means of the following analogy. Suppose that a restauranteur wishes to have waiters who will signal the diners orders by means of acoustic signals. Shall he employ waiters who attempt to bow notes on violins or wave handbells? Although violins might be more elegant, handbells which each produce a distinct acoustic signal without having to use precise manual gestures are the obvious choice if the goal were minimizing errors.

The perceptual salience of the quantal vowels [i], [u], and [a] results from the convergence of two formant frequencies, yielding spectral peaks (Fant 1960). For [i] the second and third formants, F2 and F3, converge at a high frequency; for [a] F2 and F1 converging at the midpoint of the frequency spectrum; for [u] F1 and F2 converge at a low frequency. Using "quantal" vowels when talking is similar to signaling messages using flags having saturated colors. The formants of other vowels do not converge and communicating with them is analogous to signaling with flags differentiated by pastel colors. An abrupt area function discontinuity occurs at the midpoint of the SVT, the tongue can move as much as 1 cm back or forth without changing the formant frequencies appreciably (Stevens 1972). The exact position of the speaker's tongue with respect to the midpoint constriction for [i] does not have to be precise. Cineradiographic studies of tongue movements during speech have confirmed Steven's theory (Beckman et al. 1995).

Radiographic and MRI studies show that the tongue body has a circular midsagital posterior contour and is almost undeformed when vowels are produced. An [i] involves moving the tongue upwards and forward, [a] involves moving the tongue back and down. Studies dating back to Russell (1928) consistently show that this is the case (c.f., Lieberman 2006) Moreover, the human tongue and those of most mammals are "hydrostats" (Stone & Lundberg 1996); it can not be squeezed into a smaller volume. The intrinsic muscles of the tongue are sometimes bunched up when speakers produce an [u]or an [i], but the tongue's shape is usually a segment of a circular arc when vowels are produced.

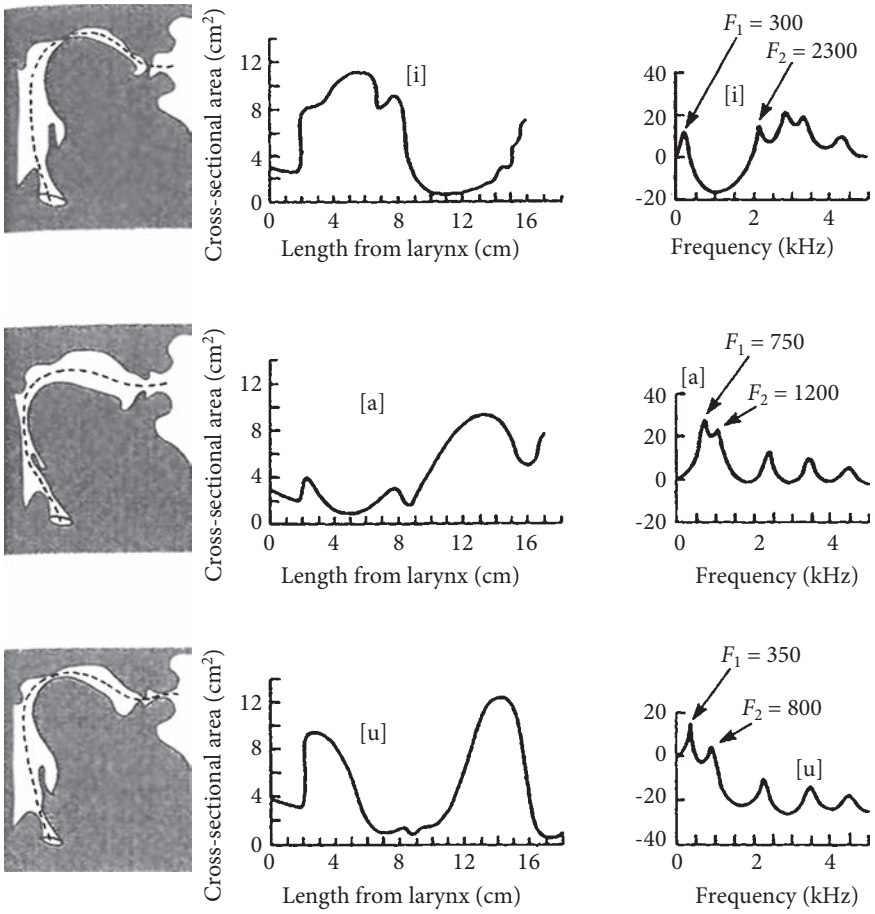


Figure 2. Midsagital views of an adult human SVT for the quantal vowels [i], [a] and [u] and the resulting formant frequency patterns. Note the peaks in the frequency spectrum that follow from the convergence of two formant frequencies. The 10 to one discontinuity at the midpoint of the SVT allows speakers to be both imprecise and still generate vowels that have spectral peaks.[i]

5. Decoding speech – estimating the length of a speaker's SVT

The vowel [i] also makes estimating the length of a speaker's SVT more certain. A long SVT produces lower formant frequencies than a short SVT for the same vowel or consonant. It became evident from the start of work on speech recognition by machine first started that the absolute values of the formant frequencies of the same sound produced by different persons vary (Peterson & Barney 1952). A "normalizing" process that takes account of SVT length is a critical step in speech perception.

Peterson and Barney (1952) first recorded seventy-six adult male, adult female and adolescent male and female speakers carefully reading aloud English words having the form [hVd], such as "heed," "hid," "had" and so on. The sound [h] in these words is an noise-excited, aspirated prelude of the vowel "V" and has the same formant frequencies – this made measurements of formant frequencies possible for voices having high F0s given the technical limitations of the sound spectrograph apparatus that was then the state-of-the-art.

The spoken words were identified by panels of listeners who had to identify each token without previously listening to a long stretch of speech produced by each particular speaker. This was achieved by having the listeners identify a set of all of the words in random order produced by ten different speakers in random order. A listener would not know whose voice or what word was coming next. There were only two errors in 10,000 trials in identifying [i]s.; [u]'s were misidentified six times, other vowels hundreds of times

The measured formant frequencies of the vowels explained some of the perceptual errors. For example, many speakers [e] vowels had the same formant frequencies of other speakers' [I]s The Peterson and Barney study has been replicated, using computer-implemented formant analysis by Hillenbrand et al. (1995).

How then can human listeners unscramble the overlapping formant frequencies? We can use different means to estimate a SVT's length. Ladefoged and Broadbent (1957, for example, showed that the same tape-recorded word was perceived as "bit," "bat," or "but" depending on the average formant frequencies of a preceding phrase. But we generally do not need to hear a person talking before we identify a word; there are immediate normalization cues in the speech signal. Nearey reasoned that the formant structure of [i] might make it possible for listeners to use it as an anchor point for vocal tract normalization.

Nearey predicted that a token of a formant frequency pattern in the [i] range will always be heard as an [i] produced by a SVT that had a particular length. If this were so, a listener would immediately "know" the length of a speaker's SVT and would correctly associate formant frequency patterns with the vowel that the speaker intended to convey. In a controlled experiment, listeners first heard a "calibrating" [i] followed by a synthesized formant frequency pattern that could correspond to any vowel produced by either a short or a long SVT, followed by the same calibrating [i](Nearey 1979, pp. 98–149). Nearey used two different calibrating [i]s, one that could be produced by an adult male's long SVT, one an [i] produced by an adolescent child's shorter SVT. Juxtaposed with the calibrating [i]'s were vowels having formant patterns that ranged over almost the total possible range of vowels for adult speakers and adolescent children. Listeners heard isolated sequences that had the for [i]-V- [i], where the [i]s were either long or short SVT [i]s and V the test stimuli. The listeners were told to identify each intermediate vowel "V." The listeners were also asked to rate the naturalness

of the vowel V of each [i-V-i] sequence that they heard. There were four categories of "naturalness judgment," from "OK" to "very bad."

The listeners' responses showed that they were "normalizing" SVTs using the single token of an [i], changing their identification of the identical formant frequency pattern when they heard it between long or short SVT [i]s. The listeners' "naturalness" responses demonstrated that they interpreted these synthesized speech stimuli using a mental procedure that "knew" the range of formant frequencies that can be produced by the calibrating [i]'s SVT length. For example, formant frequency patterns that could be produced by a short SVT were judged to be "natural" when they were embedded with an [i] produced by a short SVT but were judged to be non-speech stimuli when they were embedded with an [i] from a long SVt that inherently could not produce such high formant frequencies. The V vowels clearly were perceived in a "speech-mode," using neural processing that took account of the speech producing capabilities of the human SVT.

Other speech sounds can be used for vocal tract normalization (c.f., Lieberman (2006) for relevant studies), but the vowel [i] is an optimal calibrating sound. It's usefulness for SVT length estimation follows from its unique formant pattern (high frequency converging F2 and F3) and constraints on the vocal tract maneuvers that can be used to produce an [i]. Whereas alternate gestures can be used to generate the formant frequency patterns of virtually all other vowels, the tongue position and lip openings that generate an [i] are constrained (Stevens & House 1955; Nearey 1978). Speakers can protrude and constrict their lips to create the effect of having a longer SVT for most other vowels. Different tongue positions can be used for these vowels; speaker FSC in Nearey's (1978) study, for example, kept his tongue in almost the same position for almost all of his high F2 vowels, except for [i]. Alternate lip and larynx gestures generated his vowels' formant patterns. Fewer possibilities can generate the formant frequency patterns for an [i]. The tongue must be placed forwards and upwards to the point where turbulent noise is sometimes generated in the constricted oral passage necessary to produce an [i](Fant 1960). The vowel [i] is an "honest" signal that specifies the speaker's actual SVT length. And it is one of the speech sounds that a non-human SVT cannot produce.

The neural mechanisms for perceiving formant frequencies and deriving SVT length appear to have a long evolutionary history Other species appear to use formant frequencies to estimate the size of a conspecific. Fitch (1997) used a simple metric obtained by the subtracting the frequency of F1 from F3 to estimate a monkey's SVT length, which is highly correlated with its body weight and length. This metric works for other species (Fitch 2000) However, Fitch's metric works only because these animal vocalizations are similar to the neutral "schwa" vowel of English in which F3 is approximately equal to 5(F1)(see the compilation for non-

human primate species in Riede et al (2005)). If the same metric were applied to human speech it would yield different estimated SVT lengths for the same speaker, depending on the vowel analyzed since the formant patterns produced by humans diverge from the schwa vowel.

6. Why is speech the default mode for human language?

Why do we talk? Why don't we use manual gestures? The answer rests in some obvious factors, and one that became apparent through research conducted in the 1960s. Vocal communication frees a speaker's hands, can occur in darkness, and doesn't require looking at the individuals who are signaling. A less obvious reason is the speed at which information can be transmitted by speech. Speech allows humans to transmit phonetic distinctions at rates of up to 20 to 30 "segments" per second. Other auditory signals merge into a continuous buzz at rates exceeding 15 items per second. Speech achieves this rapid transmission rate because it is an "encoded" signal in which information is transmitted at the slower syllable rate then "decoded" into phonetic segments (Liberman et al. 1967).

For example, the formant frequency patterns that convey the "phonemes" of the word "cat" (approximated by the letters of the alphabet) are melded together into one syllable. As the tongue moves from the syllable-initial consonant, a formant frequency pattern is produced that transitions into that of the vowel, and then to the final consonant. Human speakers plan ahead. As you begin to say the word "too," your lips "round" (protrude and narrow) anticipating the rounded [u] vowel. Your lips are not rounded at the start of the word "tea", because the following vowel is not rounded. The encoding differs somewhat from language to language (Lubker & Gay 1982), and is acquired without conscious effort by children.

7. Choking and the antiquity of speech

Speech must have been present in hominid species who lacked SVTs capable of producing quantal vowels because the human SVT increases the risk of choking to death on food lodged in the larynx. Palmer and his colleagues, reviewing studies of swallowing note that in contrast to nonhuman mammals:

> "normal humans are at risk for inadvertently inhaling food particles both before and after swallowing. Indeed, obstruction of the airway by inhaled food is a significant cause of morbidity and mortality in otherwise healthy individuals." (Palmer et al. 1992).

About 500,000 Americans suffer from swallowing disorders (dysphagia), and deaths from choking are the forth largest cause of accidental deaths in the United States (http; //www.nsc.org/library/report_injury_usa.htm). There would have been no reason for retaining the mutations that resulted in the human SVT, unless speech already was in place in hominids ancestral to humans before the evolution of the human SVT.

8. Tracing the evolution of the human SVT

This brings us to the vexatious problem of reconstructing the soft tissue of the SVT of a fossil when all that remains are bones.

Much attention has been given to the position of the larynx, which as we will see, can rule out hypothetical SVTs. However, studies of the ontogenetic development of the human SVT, discussed below, reveal other factors.

1. The skeletal structure that supports the roof of the mouth rotates towards the back of the skull, effectively shortening the mouth and the "horizontal component of the SVT, SVTh during the first two years of life; the human face is "flat" compared to prognathious present day apes and early hominids such as the Australopithecines (D. Lieberman, Ross & Ravosa 2000).
2. The human tongue gradually descends down into the pharynx, changing its shape from a relatively long flat shape positioned almost entirely in the mouth to a massive form having a posterior rounded shape. This yields the 1: 1 SVTh/SVTv proportions seen in Figure 1. This unique human developmental process is not complete until age 6–8 years (D. Lieberman & McCarthy 1999 Neck length is critical since a larynx positioned below the neck at the level of the sternum (collarbone) would make it impossible to swallow (Palmer et al. 2000; D. Lieberman et al. 2001).

9. Neanderthal speech

As is the case in non-human primates throughout life, the tongue is positioned almost entirely in the mouth in human neonates. In the course of human ontogenetic development, the tongue moves down into the pharynx, carrying the larynx down with it. This process was first described by Victor Negus (1949) who thought that it reflected the:

> … recession of the jaws; there is no prognathous snout …The [human] tongue however retains the size it had in Apes and more primitive types of Man, and in

> consequence it is curved, occupying a position partly in the mouth and partly in the pharynx. As the larynx is closely approximated to its hinder end, there is of necessity descent in the neck; briefly stated the tongue has pushed the larynx to a low position, opposite the fourth, fifth and sixth cervical vertebrae. (Negus 1949, pp. 25–26)

Negus's inferences were correct insofar as extensive facial retraction occurs only in humans. As noted above, it has become clear that the process entails more than the recession of the jaws which occurs in the first two years of life.

The low position of the human larynx is a reflex of the human tongue reshaping and moving down into the pharynx. The position of the human larynx is closely coupled to tongue As the tongue descends down into the pharynx, it carries the larynx down with it. The descent of the tongue into the pharynx, its posterior circular shape and the right angle bend at its midpoint enables the human SVT to produce the major midpoint area function discontinuities necessary for quantal vowels. Thus, the descent and change in the tongue's shape is the key factors in both the development and evolution of the human SVT (Lieberman 1984, pp. 276–280).

10. When did our ancestors begin to say *oy veh*?

A SVT that can produce the full range of human speech must have 1:1 SVTh to SVTv proportions. If SVTh is long, as is the case for Neanderthals, than SVTv must also be long. But the anatomy involved in speech (tongue, hyoid bone, and larynx) must also allow us to eat. The hyoid which supports the larynx, moves upwards and forwards opening the espohagous and placing the larynx into a larynx in the neck can execute these maneuvers. However, a larynx placed in the chest, would be blocked by the sternum bone precluding eating – end of story.

We can determine whether Neanderthals and other fossil hominids could have had 1:1 SVTh to SVTv proportions by examining their basicrania, which provides a measure of SVTh and their cervical vertebrae, which provides a measure of the length of their necks. McCarthy et al. (forthcoming)studied a large a sample of rhesus monkeys, the WT 15000 *Homo ergaster* fossil, three Neanderthal fossils, 82 human specimens, including the Middle Paleolithic Skhul V fossil, eight Upper Paleolithic fossils, and 73 contemporary humans from seven different parts of the world. Neanderthal necks were too short to have fully human SVTs. (A similar conclusion was reached in Lieberman (1984, pp. 290–296). That's also the case for the middle Pleistocene fossil Skhul V (McCowan & Keith 1939), which has often been thought to be fully modern. Skhul V's SVTh is relatively long; its short neck precludes its having a fully human vocal tract. Fully modern speech anatomy is not evident in the fossil record until the Upper Paleolithic, about 50,000 years ago,

coincident with the cultural "revolution" apparent in the archaeological record (Klein 1999). So the wild ride of language change that Hal has been studying began almost yesterday – 50,000 years ago.

Happy Birthday Hal.

References

Baer, T., Gore, J.C., Gracco, L.C. & Nye, P.W. 1991. "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels." *Journal of the Acoustical Society of America* 90. 799–828.

Beckman, M.E., Jung, T-P., Lee, S-H., De Jong, K., Krishnamurthy, A.K., Ahalt, S.C., Cohen, K.B. & Collins, M.J. 1995. "Variability in the production of quantal vowels revisited." *Journal of the Acoustical Society of America*. 97. 471–489.

Cheney, D.L. & Sayfarth, R.M. 1990. *How monkeys see the world; Inside the mind of another species*. Chicago:University of Chicago Press.

Chiba, T. & Kajiyama, J. 1941. *The vowel: Its nature and structure*. Tokyo: Tokyo-Kaisekan Publishing Co.

Fant, G. 1960. *Acoustic theory of speech production*. The Hague: Mouton.

Fitch, W.T. 1997. "Vocal tract length and formant frequency dispersion correlate with body size in macaque monkeys. *Journal of the Acoustical Society of America.* 102. 1213–1222.

Fitch, W.T. 2000. "Skull dimensions in relation to body size in nonhuman mammals: The causal bases for acoustic allometry." *Zoology* 103. 40–58.

Greenberg, J.H. 1963. *Universals of language* Cambridge, Mass.: MIT Press.

Hiiemae, K.M., Palmer, J.B., Medicis, S.W., Hegener, J., Jackson, B.S. & Lieberman, D.E. 2002. "Hyoid and tongue movements in speaking and eating." *Archives of Oral Biology* 47. 11–27.

Hillenbrand, J.L., Getty, A., Clark, M.J. & Wheeler, K. 1995. "Acoustic characteristics of American English vowels." *Journal of the Acoustical Society of America* 97. 3099–3111.

Ishida, R., Palmer, J.B. & Hiiemae, K.M. 2002. "Hyoid motion during swallowing; factors affecting forward and upward displacement." *Dysphagia* 17: 262–272.

Klein, R.G. 1999. *The human career*, 2nd edition. Chicago: University of Chicago Press.

Ladefoged, P.J. & Broadbent, D.E. 1957. "Information conveyed by vowels." *Journal of the Acosutical society of America* 29. 98–104.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. 1967. "Perception of the speech code." *Psychological Review* 74. 431–461

Lieberman, D.E. & McCarthy, R.C. 1999. "The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions." *Journal of Human Evolution* 36. 487–517.

Lieberman, D.E., Ross, C.F. & Ravosa, M.J. 2000. "The primate cranial base: Ontogeny, Function and Integration." *Yearbook of Physical Anthropology* 43. 117–169.

Lieberman, D.E., McCarthy, R.C., Hiiemae, K.M. & Palmer, J.B. 2001. "Ontogeny of postnatal hyoid and laryngeal descent: implications for deglutition and vocalization." *Archives of Oral Biology* 46. 117–128.

Lieberman, P. 1968. "Primate vocalizations and human linguistic ability." *Journal of the Acoustical Society of America* 44. 1157–1164.

Lieberman, P. 1984. *The Biology and Evolution of Language*. Cambridge, Mass: Harvard University Press

Lieberman, P. 2000. *Human language and our reptilian brain: The subcortical bases of speech, syntax, and thought.* Cambridge Mass: Harvard University Press.

Lieberman, P. 2006. *Toward an evolutionary biology of language*. Cambridge Mass: Harvard University Press.

Lieberman, P., Klatt, D.H. & Wilson, W.H. 1969. "Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates." *Science* 164.185–1187.

Lieberman, P. & Crelin, E.S. 1971. "On the speech of Neanderthal man." *Linguistic Inquiry* 2. 203–222.

Lieberman, P., Crelin, E.S. & Klatt, D.H. 1972. "Phonetic ability and related anatomy of the newborn, adult human, Neanderthal man, and the chimpanzee." *American Anthropologist* 74. 287–307.

Lubker, J. & Gay, T. 1982. "Anticipatory labial coarticulation: Experimental, biological, and linguistic variables." *Journal of the Acoustical Society of America* 71. 437–438.

McCarthy, R.C., Strait, D.S., Yates, F. & Lieberman, P. forthcoming. *The Recent Origin of Human Speech.*

McCowan, T.D. & Keith, A. 1939. "The stone age of Mount Carmel." Volume 2 of *The fossil remains from the Levalloiso-Mousterian.* New York: Clarendon Press

Nearey, T. 1978. *Phonetic features for vowels*. Bloomington, Ind.: Indiana University Linguistics Club.

Negus, V.E. 1949.*The comparative anatomy and physiology of the larynx*. New York: Hafner.

Nishimura, T., Mikami, A., Suzuki, J. & Matsuzawa, T. 2003. "Descent of the larynx in chimpanzee infants." *Proceedings of the National Academy of Sciences* 100. 6930–6933.

Palmer, J.B., Rudin, N.J., Lara, G. & Crompton, A.W. 1992. "Coordination of mastication and swallowing." *Dysphagia* 7. 187–200.

Peterson, G.E. & H.L. Barney. 1952. "Control methods used in a study of the vowels." *Journal of the Acoustical Society of America* 24. 175–184.

Rendall, D., Kollias, S., Ney, C. & Loyd, P. in press. "Pitch (Fo) and formant profiles of human and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry." *Journal of the Acoustical Society of America.*

Riede, T., Bronson, E., Hatzikirou, H. & Zuberbuhler, K. 2005. "Vocal production in a non-human primate; morphological data and a model." *Journal of Human Evolution* 48. 85–96.

Russell, G.O. 1928. *The Vowel*. Columbus, Oh.: Ohio State University Press.

Stevens, K.N. 1972. "Quantal nature of speech." *Human communication: a unified view,* ed. E.E. David Jr. & P. B. Denes, 51–66. New York: McGraw Hill.

Stevens, K.N. & House, A.S. 1955. "Development of a quantitative description of vowel articulation." *Journal of the Acoustical Society of America* 27. 484–493.

Stone, M. & Lundberg, A. 1996. "Three dimensional tongue surface shapes of English consonants and vowels." *Journal of the Acoustical Society of America* 99. 3728–3736.

Can Paleolithic stone artifacts serve as evidence for prehistoric language?

Ofer Bar-Yosef
Department of Anthropology,
Peabody Museum, Harvard University

In this paper a hypothesis is offered concerning the geographic distribution of a Paleolithic language. Two models are commonly used for explaining invention and spread of new techno-typologies: emergence in a core area (invention), and appearance simultaneously and independently in various places (diffusion). The spread of new techniques, tool types, and ideas assumes the existence of verbal communication. The current evidence indicating that a Mousterian industry with Levalloisian techniques was the "mother-culture" from which the Initial Upper Paleolithic emerged, may hint that an unknown but hypothetical language was the source of the Upper Paleolithic languages. This proposal is based on the common knowledge among today's flintknappers that the learned skill of making stone tools is acquired not only through watching and mimicking the artisan's gestures, but also through oral explanations that accompany the teaching process.

It is common knowledge that the oldest written documents and recent linguistic studies serve as the basis for reconstructing prehistoric languages. At best most current hypothetical considerations reach back to the Terminal Pleistocene. The idea that languages spread with people and especially with the agricultural package is hotly debated (Renfrew 1987, 2002; Bellwood 2005). Genetic evidence collected among recent populations reflects past dispersals but also indicates a certain degree of admixture between endemic populations and recent newcomers. While the more recent dispersals benefit from our knowledge of historical events, it is somewhat difficult to date the older dispersals, in spite of calculations of mutation rates.

Explanations concerning the "out of Africa" exodus of Modern humans (e.g., Underhill et al. 2001; Underhill 2002; Forster 2004) try to sort out the chronological steps and geographic routes of humans into Eurasia. Indeed, many archaeological problems concerning this period, known as the "Middle to Upper Paleolithic transition", remain unsolved. Though the records for western Eurasia are better known than the archaeological sequences and the chronology of human cultures in the vast

regions of Central and Eastern Asia, our patchy knowledge commonly leaves room for speculations and major disagreements. Still, one can safely state that the complex pattern of the Old World colonization by Modern humans left its imprint on the history of prehistoric languages.

Difficulties in interpretations arise when comparisons between genetic data, human fossils, and archaeological remains are derived from the remote prehistoric periods such as the Middle to Upper Paleolithic transition. Disclosing clear patterns of movements of foragers is hampered by lack of detailed chronologies, poorly studied regions, and certain ambiguities concerning the published archaeological information. Moreover, it seems that there is a certain reluctance to use the studies of stone artifacts for advancing working hypotheses about patterns of dispersal, and/or as potential sources in reconstructing past languages.

In this short paper I will try to do just this. I will offer a hypothesis concerning the geographic distribution of a Paleolithic language. This proposal is based on the common knowledge among today's flintknappers that the learned skill of making stone tools is acquired not only through watching and mimicking the artisan's gestures, but also through oral explanations that accompany the teaching process.

The history of making stone tools can be divided into several phases. Clearly the earliest assemblages from *ca*. 2.6–1.7 Ma represent mostly simple rock breaking, in order to obtain sharp edges on detached pieces or the entire nodule (or pebble). This approach to making stone tools, sometimes called the "core and flake industry," continues in several geographic regions up to *ca.* 300,000 years ago and even later in eastern and south-eastern Asia. It has been suggested that we compare its earlier phase, i.e., the Oldowan, to stone manipulation by apes (e.g., Wynn & McGrew 1989) although study of experimental stone-tool making by the chimp Kanzi does not support this proposal (Toth et al. 1993). Hence, the making of "core and flake industries" is not explained by comparison to apes as genetically programmed, and the presence of such simple stone tools (during the Upper Pleistocene in East Asia, for example) has no bearing on physical or cognitive capacities of humans.

The new tool types that emerged around 1.7 MA are the bifaces (or hand axes) that characterize the Acheulian industries. Here again, the mode of production based on shaping the two faces of the piece (whether a nodule or thick flake) continues for quite some time. It is present all through the Middle Paleolithic, re-appearing during the Upper Paleolithic in certain areas, and is well known in Neolithic assemblages. If we confine ourselves to the production of Acheulian bifaces that were often thicker and larger than the foliates or bifacial points of the Middle Paleolithic, the sequence is interrupted by the introduction of the Levallois technique some 300/250,000 (the beginning of the Middle Paleolithic) or at an earlier time around 400,000 years ago.

The Levallois technique is seen as a suite of production methods based on the artisan's intention to obtain a particular type of blank (detached piece) from a nodule.

Often the nodules or pebbles searched for in this technique were relatively flat because the volumetric shape of the raw material may determine the type of detached product. However, observations by many lithic analysts indicate that the options for obtaining the desired detached pieces vary. Once the raw material was selected, the knapper would start the process of detachment by direct percussion. Given the various methods he or she may obtain different types of products (for detailed explanations see e.g., Boëda 1995; Van Peer 1992).

The most basic production methods of the Levallois include the radial or centripetal detachment, the convergent *recurrent*, the convergent for obtaining triangular points, and the uni- or bi-directional. The first method produces large oval or rectangular flakes. The second method removes a series of flakes and some blades while the core is being constantly reshaped; the third method results in triangular flakes often called Levallois points, and the fourth method produces blades (pieces whose length is larger than twice their width). There is some grading between the "convergent *recurrent*" and the "uni-bi-directional" but the "radial" and "convergent for points" result in producing easily recognizable distinct forms.

Contrary to the common belief in the "simplicity" of stone tool making, the production of Levallois artifacts requires more expertise than that needed for the production of simple blades that characterized the later Upper Paleolithic industries. Those who believe that Neanderthals were not capable of having complex stone techniques, as compared with Modern humans, are proven wrong when faced with typical Levallois products recovered from Middle Paleolithic (often referred to as the Mousterian cultures) contexts in Western Europe or the Levant.

Literature on lithic techniques suggests that the Levallois technique evolved from the bifacial shaping of hand axes (bifaces). If this were the case, then the constant and frequent making of bifaces for almost 1.7 million years should have culminated in the appearance of the Levallois technique many hundreds of thousands years ago. However, this did not happen, and systematic production of Levallois blanks began only some 400–300,000 years ago.

The difficulties in dating sites of Middle Pleistocene age (780–130,000 years ago) led archaeologists to suggest that the Levallois technique, which is a set of several methods of obtaining flakes, blades, and points, was invented independently in various regions in Africa and western Eurasia. However, this technique is not known from eastern or southeastern Asia where the simple "core and flake" technique of obtaining flakes was employed during the Middle and Upper Paleolithic periods. In spite of the difficulty in dating the early appearances of the Levallois technique in the archaeological record we need to consider whether an invention of a well-designed knapping technique requires an answer to a frequent question: "do new inventions emerge in a particular geographic area, or is it possible that such a knapping technique emerged contemporaneously in various regions and is an example of cultural convergence?"

We often use two models for explaining invention and spread of new techno-typologies in our world. We see the new technique, tool type, idea, or the like, both as emerging in a core area, and appearing simultaneously and independently in various places. The same explanations were advanced for the Industrial Revolution. Once ideas of applied scientific research and the making of machines emerged in England it took just a few decades for these new technologies to spread and be adopted in France, Holland and other countries (e.g., Mokyr 2002). It should be noted that the transfer of the technical achievements of this period to North America was aided by the use of the common English language.

Going back in time, I would like to mention two current archaeobotanical proposals that view the emergence of cultivation as taking place in different locales across the Levant (Willcox 2005; Weiss et al. 2006). While the general evidence seems to support their contention, a detailed chronology for testing the presence of cultivation in the various locales is missing. This means that in a relatively small region on a continental scale (ca. 1000 km long and ca. 100–250 km wide), with documented obsidian trade or exchange from Anatolia down to Jericho, it is hard to accept that there was no communication between the different villages on route. Hence, it is not surprising that cultivation as a subsistence strategy, in addition to gathering in the wild, began at the same general time (within one or two centuries around 11,500 cal B.P.) across the entire region. This assumes either the existence of verbal communication, whether through one common language with various dialects, or a shared knowledge of various languages among the people of the region.

Earlier on in time we can note that the Magdalenian culture (ca. 19–14,000 cal B.P.) spread across western and central Europe. Its most amazing features are the use of the same symbols expressed in art objects and utilitarian stone objects. Whether incised or painted, animals such as the ibex, horses, female figures, or abstract symbols are always depicted in the same style (e.g., Djindjian et al. 1999). The best interpretation is that the creators of all these artistic representations shared the same views of the world (cosmology) and probably spoke the same language (G. Bosinski, personal communication). Like every specific population it seems that the Magdalenian emerged in a particular region and being a successful culture of foragers, expanded its territory. The early phase of the Magdalenian is considered a local phenomenon of the Franco-Cantabrian zone, but the later Magdalenian phases spread over larger regions of Western Europe including Iberia and parts of Central Europe.

One of the most discussed human migrations is the "out of Africa" dispersal of Modern humans mentioned above. It is documented through genetics, and supported by the archaeological evidence from western Eurasia depicting a chrono-geographic trajectory of the colonization from east to west.

If we consider a similar attitude towards the Levallois methods, these being the common stone making techniques among the same people, one possible explanation

of such a scenario is that the particular knapping designs were invented within a specific region of the Old World and only later spread all over to be shared by many other groups, enriched in due course by a series of additional technical improvements.

In this scenario we can assume that constant gene flow facilitated the transmission of the knowledge of the different Levallois methods. This could have happened within the realm of a single expanding population, with the expectation of exchange and mobility between groups and migrations across populations. In the latter case we should be able to notice exchange of technical knowledge and/or processes of acculturation. The transmission of the Levallois methods could have been the most fundamental result of inter-group contacts all across vast areas in the Old World.

This particular explanation for the emergence and spread of the Levallois methods does not appear in most of the published reports from the different African and Eurasian regions. Archaeologists often stress certain attributes observed in core-reduction strategies easily explained through equifinality or the limitations imposed by raw material. Although the "devil is in the details," perhaps this is what prevents us from seeing the broader picture that may have implications for reconstructing the prehistory of certain Paleolithic languages. This assumption is based on my contention that we definitely can identify different "paleo-cultures" across the Middle and Upper Paleolithic world. Efforts in this direction have already been published, though obviously they are mostly related to the rich records of the Upper Paleolithic period (Bocquet-Appel & Demars 2000a, 2000b). However, we also have to take into account that the disappearance of several "paleo-cultures" may correspond to lineage extinctions, and that the abandonment of particular areas was a common phenomenon during the long Middle Paleolithic period (ca. 300/250,000–35,000 B.P.).

When we consider the Levallois *suite* of knapping methods I would like to suggest a hypothesis that the significance of the knowledge of technical methods for making stone tools resided in the social level of human groups forming a biologically viable population, speaking the same language. The survival of those who were heavily dependant upon the Levallois set of techniques led to a series of expanding populations. If dates for an earlier appearance of even a single Levallois method were to converge around 300–250,000 years ago, it is not impossible that the bearers were those particular early Modern human groups which moved from Africa. The human fossils of Omo Kibish and Herto-Bouri in Ethiopia, or the Skhul/Qafzeh skeletal remains in the Levant, could represent a population of Levallois users that kept moving in different directions, and by spreading across several regions of the Old World they also spread their common language. The current evidence indicating that a Mousterian industry with Levalloisian techniques was the "mother-culture" from which the Initial Upper Paleolithic emerged, may hint that this unknown language was the source of the Upper Paleolithic languages.

Acknowledgments

This paper is dedicated to Hal Fleming who endlessly kept asking me about prehistoric cultures and what can be learned from them. I thank him for forcing me to think about the archaeology of language. Thanks to A. Belfer-Cohen (Hebrew University, Jerusalem) for her editorial help. I am solely responsible for this far-fetched hypothesis.

References

Bellwood, P. 2005. *First Farmers: The Origins of Agricultural Societies*. Oxford: Blackwell.

Bocquet-Appel, J.-P. & Demars, P.Y. 2000. "Neanderthal contraction and modern human colonization of Europe." *Antiquity* 74: 285. 544–552.

Bocquet-Appel, J.-P. & Demars, P.-Y. et al. 2005. "Estimates of Upper Paleolithic meta-population size in Europe from archaeological data." *Journal of Archaeological Science* 322: 11. 1656–1668.

Boëda, E. 1995. "Levallois: A volumetric construction, methods, a technique." *The Definition and Interpretation of Levallois Technology,* ed. by H.L. Dibble & O. Bar-Yosef, 41–68. Madison, Wis.: Prehistory Press.

Forster, P. 2004. "Ice ages and the mitochondrial DNA chronology of human dispersals: a review." *Philosophical Transactions of the Royal Society London* B 359. 255–264.

Mokyr, J. 2002. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton and Oxford: Princeton University Press.

Renfrew, C. 1987. *Archaeology and Language: The Puzzle of Indo-European Origins.* Cambridge, UK: Cambridge University Press.

Renfrew, C. 2002. "'The Emerging Synthesis': the Archaeogenetics of Farming/Language Dispersals and other Spread Zones." *Examining the farming/language dispersal hypothesis,* ed. by P. Bellwood & C. Renfrew, 3–16. Cambridge, UK: McDonald Institute for Archaeological Research.

Rolland, N. 1995. "Levallois technique emergence: Single or Multiple: A review of the Euro-African record." *The Definition and Interpretation of Levallois Technology,* ed. by H. Dibble & O. Bar-Yosef, 333–360. Madison, Wis.: Prehistory Press.

Toth, N., Schick, K.P. et al. 1993. "Pan the Tool-Maker: Investigations into the stone tool-making and tool-using capabilities of a Bonobo (*Pan paniscus*)." *Journal of Archaeological Science* 20: 1. 81–92.

Underhill, P.A. 2002. "Inference of Neolithic Population Histories using Y-chromosome Haplotypes." *Examining the farming/language dispersal hypothesis,* ed. by P. Bellwood & C. Renfrew, 65–78. Cambridge, UK: McDonald Institute for Archaeological Research.

Underhill, P.A., Passarino, G. et al. 2001. "The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations." *Annual of Human Genetics* 65: 43–62.

Van Peer, P. 1992. *The Levallois Reduction Strategy*. Madison, Wis.: Prehistory Press.

Willcox, G. 2005. "The distribution, natural habitats and availability of wild cereals in relation to their domestication in the Near East: multiple events, multiple centers." *Vegetation History and Archaeology* DOI 10 1007//s00334-005-0075-x: 17 pages.

Weiss, E., Kislev, M.E. et al. 2006. "Autonomous cultivation before domestication." *Science* 312: 1608–1610.

Wynn, T. & McGrew, W.C. 1989. "An ape's view of the Oldowan." *Man (N.S.)* 24: 383–398.

The origin of language

Symbiosism and symbiomism

George van Driem
Leiden University

Symbiosism is a Darwinian model of language and its emergence. Symbiotic Theory operates on the Leiden definition of memes as isofunctional neuroanatomical entities corresponding to linguistic signs in the Saussurean sense, not on the Oxonian conception of memes as units of imitation. Symbiosism treats linguistic forms as vehicles for the reproduction of meaningful elements in the hominid brain and so transcends the obsolete discord between the functionalist or European structuralist conception of language, whereby linguistic forms are seen as instruments used to convey meaningful elements, and the formalist or generative approach, whereby linguistic forms are treated as abstract structures which can be filled with meaningful elements. Symbiomism is the philosophy of life which grew out of Symbiosism and which understands our individual and collective human identity as symbiomes of a biological host and a semiotic symbiont.

Alfredo Trombetti's 'la nostra dottrina monogenistica' (1925: 151) was, as I once previously hastened to point out, precisely that, a doctrine, because demonstrating that all human languages stem from a single original mother tongue lies beyond the bounds of what is accessible to empirical testing, at least by conventional comparative linguistics. Allan Bomhard, Michael Fortescue, Joseph Greenberg, Laurent Sagart and the late and universally beloved Sergei Starostin have gone beyond where many traditional comparative linguists dare to tread and so ventured through what I have characterised epistemologically as the 'monogeneticist time warp' (van Driem 2001: 145). Yet the work of these intrepid scholars – but perhaps first and foremost amongst them, Harold Fleming – has shown that it may be possible to breach this time warp in an empirically defensible way and so responsibly to indulge in 'gazing beyond the event horizon' of conventional comparative linguistics.

This essay too goes well beyond this horizon and treats of the origin of language itself. The empirical basis of the Leiden conception of language is language's own lingering legacy in the shape of the very neuroanatomical and semiotic workings of meaning. Yet the aim here is not to repeat what has been explained in the previous work of exponents of the Leiden school of language evolution, but to point out an essential difference between Kortlandt's and Wiedenhof's views of language and my

own, viz. Kortlandt (1985, 1998, 2003), Wiedenhof (1996) vs. van Driem (2001a, 2001b, 2003, 2004, 2005). Some critics had already noticed that the Leiden school is not a single coherent view of language evolution, but three coherent views of language evolution, albeit largely congruent ones.

Symbiosism is a Darwinian model of language and its emergence. Symbiosism is a variety of Symbiotic Theory, which treats linguistic forms as vehicles for the reproduction of meaningful elements in the hominid brain. The symbiotic view transcends the obsolete discord between the functionalist or European structuralist conception of language, whereby linguistic forms are seen as instruments used to convey meaningful elements, and the formalist or generative approach, whereby linguistic forms are treated as abstract structures which can be filled with meaningful elements. Symbiotic Theory shows naming and syntax to be two faces of the same phenomenon.

Syntax arose from meaning. The first primaeval holistic utterances with a meaning in the linguistic sense inherently constituted a projection of reality with a temporal dimension. First-order predication arose when such a holistic utterance was split. This point of view was argued by Pierre de Maupertuis (1756, III: 444) and Hugo Schuchardt (1919a, 1919b) and contrasts with the naïve view that syntax arose from the concatenation of labels or names. The splitting of a signal for 'The baby has fallen out of the tree' could have yielded meanings such as 'That which has fallen out of the tree is our baby' and 'What the baby has done is to fall out of the tree'. Mária Ujhelyi (1998) has considered long-call structures in apes in this regard.

Moreover, Symbiotic Theory operates not on the Oxonian conception of a meme as a 'unit of imitation', but on the Leiden definition of memes as isofunctional neuroanatomical units corresponding to linguistic signs in the Saussurean sense, corresponding to single morphemes or monomorphemic words. The neuronal correlate of a meaning along with the neuronal representation of its associated phonological form or grammatical manifestation is a *meme*, whereas in Leiden school terminology a unit of imitation is a *mime*. The symbiotic model of the human mind is based on an understanding of language as a semiotic organism which has arisen and evolved in the hominid brain. The empirical basis of the Leiden conception of language is language's own lingering and tangible evolutionary legacy in the shape of the neuroanatomical and semiotic workings of meaning.

The Leiden school is not a single view of language evolution, but three largely congruent views of language evolution. Kortlandt (1985, 1998, 2003) and Wiedenhof (1996) conceive language to be a parasite, based on the correct insight that natural meanings have the properties of non-constructible sets in the mathematical sense. Symbiosism, however, distinguishes the mutualist nature of language as such from the workings of individual meanings (van Driem 2001a, 2001b, 2003, 2004, 2005). Language has greatly augmented our reproductive fitness to the detriment of countless other macroscopic species. Language, therefore, is a mutualist symbiont.

Our language-driven pre-eminence has made us the blight of the biosphere in the Holocene period. On the other hand, language-borne ideas can be beneficial or deleterious to the human host.

1. Language as organism is no mere metaphor

The idea that language is a life form in its own right was popular amongst Indo-European linguists in Germany in the early 19th century. Friedrich von Schlegel described language as 'ein lebendiges Gewebe' (1808: 64), and Wilhelm von Humboldt spoke of the 'Organismus der Sprache' (1812: 8). Later, inspired by Ernst Haeckel's popularisation in Germany of Darwin's 1859 book *On the Origin of Species*, August Schleicher formulated a lucid statement on the organismal nature of language. Some have misinterpreted this conception of language as an organism as no more than a metaphor. Yet Schleicher's statement about language as a life form was unequivocally literal: 'Die Sprachen sind Naturorganismen, die, ohne vom Willen des Menschen bestimmbar zu sein, entstunden, nach bestimmten Gesetzen wuchsen und sich entwickelten und wiederum altern und absterben; auch ihnen ist jene Reihe von Erscheinungen eigen, die man unter dem Namen »Leben« zu verstehen pflegt' (1863: 6–7).

The conceptualisation of language as an organism remained popular, but the notion was reinterpreted by historical and comparative linguists in ways that differed from the intimations of more semiotically inclined thinkers. In retrospect, the latter category of thinkers blazed the trail for the Leiden school of language evolution. The indologist Max Müller and later the mathematician Bertus Brouwer had profound and often disturbing insights into the nature of linguistic meaning and the effects of language. Müller and Brouwer can be identified as Frederik Kortlandt's intellectual precursors. In the early 1980s, Kortlandt's tutorials led to the growth in Leiden of a new school of thought on language evolution, with Jeroen Wiedenhof and myself at the time as his principal disciples. His now famous article on the language parasite, which appeared in 1985, explained that the nature of the organism stemmed from the fact that meanings, the replicating units of language, were non-constructible sets in the constructivist mathematical sense.

Kortlandt's view was a radical departure from earlier views on language as an organism, for his semiotic approach cut to the chase in identifying meaning and its behaviour as the crux of linguistics and language evolution. In Kortlandt's Leiden school, the inherently dynamic character of meanings is seen as a direct function of their neuroanatomy, as modelled by Hebb (1949) and later elaborated by Changeux (1983) and Edelman (1987). Brouwerian semantics dovetails with neuroanatomical reality and the observable behaviour of categories of meaning as units in the Darwinian process of neuronal group selection.

The Leiden model heralds a revolution in the way we think about ourselves. So, one might predict a natural disinclination to welcome Symbiotic Theory based on what we know about human nature or, alternatively, based on the historiographical model of paradigm shift in scientific revolutions developed by Thomas Kuhn (1962, 1977). Kuhn underscored the social dimensions and psychological imperatives of change, though it would be imbalanced to overlook the role of sheer serendipity in science and the individual character of insight in scientific thought. Some have taken cognizance of, but chosen to ignore Symbiotic Theory, whilst Salverda (2003) has drawn attention to several attempts to surreptitiously scoop the Leiden school by secondarily promulgating its insights. Another, intrinsically far more intriguing cause for a reluctance to accept the symbiotic view, I believe, may lie in a natural resistance built into our minds against recognising the linguistic symbiont for what it is. Language may not want to be found out. Our mind, caught in the web of language, is neither inclined nor even well-equipped to discern its own linguistic soul.

2. Pessimistic vs. optimistic linguistics?

The perceived difference between Kortlandt's view of language and my own symbiosist view has often been phrased, even by Kortlandt himself, along the lines of the master viewing language as a parasite, whereas his pupil sees language as a symbiont. Part of the confusion is terminological in nature, for technically a parasite too is a symbiont. Symbiosis is when two phylogenetically distinct organisms live together in some sort of intimate relationship. Symbiotic relationships abound in nature and take on many forms. The most far-reaching form of symbiosis is a relationship in which both organisms cannot live without the other and effectively become as one life form. Most life forms on the planet today originated as symbiotic relationships. An early understanding of the role of symbiosis in evolution dates back to the same period in the history of biology that evolution by natural selection first came to be understood by Charles Darwin and Alfred Russel Wallace.

Pierre Joseph van Beneden, professor at the Catholic University at Leuven, adopted the term *mutuellisme,* brandished by the French social reformer Pierre-Joseph Proudhon for his ostensibly benign variety of communism, to apply to mutually beneficial relationships between species. The Belgian marine biologist later popularised the idea in his 1876 book *Les commensaux et les parasites*, which also appeared in German and English translations that same year. He distinguished various types of symbiotic relationship, i.e., parasite, free-living commensal, resident or obligate commensal and mutualist. Van Beneden stressed that beneficial reciprocity was as prevalent as commensalism. He described in detail how commensalism and mutualism contrasted

strongly with the deleterious effects of parasitism and likewise carefully distinguished between various forms of commensalism and the intimate and reciprocally beneficial interdependency which characterised mutualism.

Van Beneden's work inspired the German botanist Heinrich Anton de Bary, who in 1879 popularised the word *Symbiose* 'symbiosis', an already extant term of Greek origin, in a public address to German biologists and physicians at Cassel as a cover term to designate all forms of 'Zusammenleben ungleichnamiger Organismen', i.e., the living together of organisms with different names, viz. belonging to differently named taxa. Symbiosis included 'der *vollständige Parasitismus*' (viz. full-fledged parasitism, which de Bary held to be the 'most exquisite' form of symbiosis), various types of commensals, and 'van Beneden's *Mutualisten*', which were neither parasitic nor commensal. De Bary's most fascinating examples were lichens. All lichens are symbiomes of fungi known as ascomycetes with either algae or cyanobacteria. His description of these fascinating symbiomes made lichens the emblematic classroom example of symbiosis.

Friedrich Schmitz, professor of botany in Bonn, observed that the chloroplasts of eukaryotic algae, along with their associated starch-accumulating structures called pyrenoids, were not fabricated anew in the cytoplasm, but reproduced independently by division within individual cells (1882). Schmitz first made this observation in 1880 'für eine Anzahl von Algen … während eines Aufenthaltes an der Zoologischen Station zu Neapel', but within two years he had established that the independent reproduction of *Chromatophoren* or chloroplasts was a feature of all eukaryotic algae.

This observation regarding the autonomous nature of chloroplasts in eukaryotic algae inspired botanist Andreas Schimper, who in 1883 showed that *Chlorophyllkörner* or chloroplasts in green plants too 'nicht durch Neubildung aus dem Zellplasma, sondern durch Theilung aus einander entstehen' (1883: 106). This discovery led Schimper to venture that all green plants had originated through an original symbiotic association of two unlike organisms: 'Möglicherweise verdanken die grünen Pflanzen wirklich einer Vereinigung eines farblosen Organismus mit einem mit Chlorophyll gleichmäßig tingierten ihren Ursprung'.[1] In a similar vein, the botanist Albert Bernard Frank (1885) soon afterwards recognised mycorrhiza too to be a symbiotic relationship between terrestrial plants and subterranean fungi which subsist on their roots and provide these plants with essential nitrogen and minerals.

---

1. Recently, gene sequencing has provided the first genomic evidence that all plastids form a monophyletic group and that a single endosymbiotic event gave rise to a unified but highly diverse phylum comprising all primary photosynthetic eukaryotes, viz. green plants, red algae and glaucophytes (Rodríguez-Ezpeleta *et al.* 2005).

It was in Russia that the term *symbiosis* began to acquire a new anodyne flavour. Andrej Sergeevič Famintsyn studied the ontogeny of chloroplasts in green plants (1889, 1893, 1907). His studies inspired Constantin Mereschkowksy to make the same observation in 1905 that Schmitz had made in 1880 and Schimper in 1883, namely that chloroplasts are not assembled from scratch in the cytoplasm, but are cytoplasmically inherited and replicate themselves autonomously within the host cell. Mereschkowksy went a step further than Schmitz and Schimper, however, in claiming that chloroplasts remained genetically independent of the nucleus. Mereschkowksy also argued that *'Cyanophyceae'* or cyanobacteria, which until relatively recently used to be called blue-green algae, were basically free living chloroplasts that had not entered into the cytoplasma of a host cell, where they had taken up a reduced symbiotic existence and rendered the host cell autotrophic. For the genesis of a new life form through symbiosis, Mereschkowsky coined the term *symbiogenesis* in 1909.

Famintsyn felt that the term *simbioz"* should be reserved for relationships that were mutually beneficial, i.e., that *simbioz"* be used in the sense of van Beneden's mutualism rather than in the sense of de Bary's symbiosis. Famintsyn's symbiosis therefore excluded parasitism, which de Bary had considered to be the 'most exquisite' form of symbiosis. Since then, numerous types of symbiosis have been identified and analysed, and an elaborate terminology has evolved to designate different types of symbiotic relationship, e.g., parasymbiosis, social parasymbiosis, phoresy, inquilinism, symbiotrophism (Henry 1966; Margulis & Schwartz 1988). This rich arsenal of precise terminology contrasts with the feel-good 'New Age' flavour which the term *symbiosis* has acquired today in popular lay usage. Not all symbiotic relationships are mutually beneficial, but in lay parlance and even sometimes in biological discourse *symbiosis* is used to refer to mutually beneficial relationships. This connotation can be traced back to Famintsyn, who gave the term a favourable twist and thus set in motion a shift in meaning away from de Bary's original usage of *Symbiose* as a cover term for all forms of intimate inter-species relationship.

Recapitulating, the label 'Symbiotic Theory', which I introduced for the Leiden model of language evolution, can be applied to both Kortlandt's view of language and my own. Besides using the term *symbiosis* strictly in its original Flemish and German sense, i.e., free of value judgment, I have discussed the language organism whilst knowingly suggesting the originally Russian, now popular pleasant connotation of *symbiosis* as mutualism. Kortlandt, however, has insisted that language is a parasite. A terminologically more precise rephrasing of the difference between Kortlandt's view of language and mine, therefore, would be whether language is a parasite, and thus an organism deleterious to its hominid host, or a mutualist, and so a partner in a mutually beneficial symbiotic relationship. Kortlandt has called his version of Symbiotic Theory 'the parasitologist's view'. I call my more optimistic, mutualist view Symbiosism. Yet we must be careful not to over-simplify either view.

3. The Leiden definition of the meme and its precursors

An idea often takes shape in more than just one human brain. Sometimes the same idea occurs independently to the minds of different individuals at very different times or even recurrently to various people throughout history. Alternatively, the cultural environment may be ripe for an idea which occurs independently to the minds of different individuals at roughly the same time in history. Yet scholars seldom recount the course of events in precisely that way, and the history of ideas is usually told as a tale that does not reflect this more complex reality. The view of culture as a dynamic evolving process in which words and ideas act as the transmitted units of evolution is in fact a rather obvious way of looking at human culture, and so this conception of culture has occurred to many people. Victor Hugo wrote that 'le mot, qu'on le sache, est un Être vivant' (1856, I: ℓ. 675).

The linguist Friedrich Max Müller was a great proponent of evolution by natural selection and applied the theory to language, religion and cultural evolution. Müller wrote: 'A struggle for life is constantly going on amongst the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand, and they owe their success to their inherent virtue' (1870: 257). Darwin himself adopted Müller's view of linguistic evolution and echoed Müller's insights: 'The survival or preservation of certain favoured words in the struggle for existence is natural selection' (1871, I: 60–61). Darwin added 'novelty' to Müller's repertoire of traits that might enhance the appeal and thus survival potential of a word.

Writing in the context of the phylogeny of Niger-Congo languages, Gottlob Adolf Krause claimed: 'Für mich ist jedes Wort ein sprechendes Lebewesen, das seine Geschichte erzählt, sobald ich es kennengelernt habe. Ich sehe die Zeit kommen wo man von einer etymologischen Biologie sprechen wird' (1885: 257). So, already in the nineteenth century, *words* were conceived as the living units of cultural evolution by Hugo and Krause, and Müller and Darwin explicitly saw words as units of evolution subject to natural selection.

In a related but different vein, a zoologist in Germany began to contemplate the notion of transmissible neural entities. Richard Wolfgang Semon coined the term *Mneme*. Semon was born on the 22nd of August 1859 in Berlin. He became Ernst Haeckel's favourite student at Jena, conducted zoological expeditions to Africa and Australia, produced a number of zoological studies, converted from Judaism to Protestantism in 1885, and later became a Monist, all before he developed his mneme theory. Semon published the book *Die Mneme als erhaltendes Prinzip im Wechsel des organischen Geschehens* in Leipzig in 1904, two revised editions of which appeared in 1908 and in 1911. A first sequel to *Die Mneme* appeared in 1909 entitled *Die mnemischen Empfindungen*. Yet Semon never completed the second sequel about 'die Pathologie der Mneme'. Unable to reconcile himself with the defeat of Germany at the end of the

First World War, he shot himself through the head on the 27th of December 1918. His lifeless body was found the following day sprawled out on the old black, white and red German tricolour.[2]

Deeply imbued with the work of Darwin and Haeckel, Semon's conception of the mneme was fundamentally an idea which modern biological theoreticians today would call Lamarckian.[3] Semon developed an epigenetic theory of memory based on the notion of the *Engramm*, a modification in the neural tissues corresponding to a memory triggered by a *Reiz* 'stimulus'. Semon conceived of the mneme as the collective set of *Engrammata* or neural memory traces, whether conscious or subconscious, that he believed were inherited genetically. Semon describes the mneme as 'das für die organische Entwicklung unumgänglich notwendige *erhaltende* Prinzip, das die Umbildungen bewahrt, welche die Außenwelt fort und fort schafft' (1911: 407). Largely forgotten today, Semon's ideas were quite influential in the first half of the twentieth century, and some of his other coinages such as *Engramm, Engraphie* and *Ekphorie* have taken up lives of their own, in German as well as in other languages.

The term 'mneme' was adopted by the Belgian entomologist, poet and playwright Maurice Maeterlinck, whose work was preoccupied with symbolism and who won the Nobel prize for literature in 1911. His entomological works *La vie des abeilles,* first published in 1901, and *La vie des termites*, first published in 1926, were translated as *The Life of the Bee* and *The Soul of the White Ant* respectively. Both books went into numerous printings in English in the first half of the twentieth century. Maeterlinck attempted to explain the workings of memory in termites and ants in terms of engrams, i.e., neural memory traces, which were added 'upon the individual mneme' (1927: 198).

After the discovery of the double helical structure of deoxyribose nucleic acid (DNA) and the chemical identity of genes in the Cavendish lab in Cambridge by Francis Crick, James Watson and Rosalind Franklin in 1953, Müller's view of the natural selection of 'words and grammatical forms' and Darwin's view of 'the survival or preservation of certain favoured words' was rapidly and widely succeeded by a more general public awareness that there must be units of cultural replication analogous to the gene.

The cultural evolutionist Leslie White came up with the term *symbolate* for 'something that results from the action or process of symboling', coarsely conceived as encompassing all 'phenomena dependent upon symboling' (1959: 231, 246). Perhaps unbeknownst to White, the term *symbolate* had already been used by Lady Victoria Welby in

---

2. A more detailed account of his colourful and turbulent life in the historical and scientific context of the day is provided by Jürg Schatzmann (1968).

3. The neo-Darwinians have generally been wont to obscure or underplay the fact that Darwin himself was a Lamarckian.

the sense of 'thing symbolised' (1896: 196). For White, however, symbolates were observable not only as acts and external events, but, in keeping with the inclusive definition of culture, symbolates also include 'concepts, beliefs, emotions, attitudes' within the human brain as well as acts and events mediated by 'symboling' and all external objects and events which are cultural artefacts or in some way the result of human intervention (1959: 235).

A widespread awareness that cultural evolution too must be a Darwinian process prompted Hudson Hoagland in 1962 to suggest what must have been obvious to many people. Hoagland proposed that *ideas* are the units of selection and that 'ideas may be considered to social evolution what genes are to biological evolution' (Huxley 1962: 203). Hoagland saw competing ideas as units of what he called 'psychosocial selection' in cultural evolution. In 1964, Henry A. Murray coined the term *idene* as an analogue in social evolution to the gene in biological evolution (Hoagland 1964: 111).

In 1963, manifestly inspired by the works of Semon and Maeterlinck, Harold Blum coined the term *mnemotype* for a unit of 'information determining the cultural pattern of a society' residing 'in the brains of its members where it is stored as personal sets of memory images'. Blum envisaged the cultural evolution of a society in terms of 'changes in the collective mnemotype', and that these innovations were precipitated by 'changes in the individual mnemotypes which compose it' (1963: 39). Others likewise devised neologisms for the widely assumed unit of cultural selection. Ralph Burhoe coined the term *culturetype* for assemblages of cultural and linguistic information. He saw this as a new type of information in evolution representing a relatively stable, transmissible 'heritage' largely independent of the genotype, but just as subject to natural selection (1967: 83).

In a panoramic treatment of man's natural history, Carl Swanson (1973: 313) proposed the term *socio-genes* for the ideas or cultural molecules as units of selection in a process of cultural evolution governed by the principles of Darwin and Mendel. Swanson also addresses the subjective and illusory nature of 'progress' in biological and cultural evolution. In a similar vein, Cloak wrote of the 'natural selection of cultural things', such as behavioural instructions which he termed *tuitions* and defined as 'the programming of an instruction upon one's hearing a linguistic analogue of that instruction uttered by a conspecific', a process which 'is almost surely unique to humans' (1975: 167). Cloak described *tuitions* as 'corpuscules of culture' residing in the central nervous system.

Just as Maeterlink's *mneme* had been the inspiration of Blum's coinage *mnemotype*, Laurent (1999) identified Maeterlink's *mneme* as the source of Richard Dawkins' *meme* in 1976, although the Oxford zoologist may no longer have been consciously aware of the engram at the time. A meme as defined by Dawkins was not just something essentially different from Semon's *Mneme*. Dawkins' meme also differed from all the cultural replicators for which various labels had already been proposed, viz. *words, symbolates, ideas, mnemotypes, idenes, culturetypes, socio-genes* and *tuitions*.

Whereas all these putative units of cultural selection were either explicitly or implicitly conceived as linguistic or language-mediated entities, Dawkins characterised a meme as 'a unit of cultural transmission, or a unit of *imitation*', with italics supplied by Dawkins to emphasise that a meme was a unit of essentially imitative behaviour (1976: 206). As opposed to earlier views of the unit of cultural selection, Dawkins' *meme* was inspired as much by the mindless mimicry observed in butterflies and by learnt behaviours such as some bird song as it was by human culture. With its single-minded focus on imitation, a deafening silence reigned about the crucial role of language. In comparison to earlier conceptions of the units of selection in cultural evolution, the meme was originally therefore actually a step backward.

In a later edition, Dawkins brought his definition of the meme more into line with earlier conceptions of a unit of cultural selection by adding that a meme was 'a unit of information residing in the brain' (1982). Yet fundamentally Dawkins' meme remained a 'unit of imitation', and therefore something neither specifically human nor necessarily linguistic. This definition found its way into the Oxford English Dictionary as 'an element of a culture that may be considered to be passed on by non-genetic means, esp. imitation'. The Oxonian meme is not essentially a semiotic construct. Blackmore, an orthodox proponent of Dawkins' view of the meme, envisages 'spoken grammatical language' as resulting from 'the success of copyable sounds' and explicitly denies the relevance of the meanings borne by language (1999). The inadequacy of the Oxonian *meme* underlay Kortlandt's choice not to use the term in the early 1980s in his treatment of the replicating units of language, viz. meanings with the propensity of non-constructible sets in the intuitionist mathematical sense.

Pursuant to the discovery of the double helical structure of DNA in 1953, the coinage *gene* – by truncation from *genetic* – aided and abetted the popularisation of the 1976 coinage *meme* so that *meme* soon outcompeted all other coinages. The deficiency of the Oxonian conception of the unit of cultural evolution, however, necessitated either the redefinition or replacement of the term *meme*. In my view, the term's popularity and its interesting lineage made it more expedient to redefine the term rather than to coin yet another neologism. The Leiden definition brought the term back into line with the conception of earlier thinkers by redefining a meme as a neuroanatomical unit corresponding to a sign in the Saussurean sense, i.e., the neuronal correlate of a meaning along with the neuronal representations of its associated phonological form and grammatical manifestation.

In Leiden, a unit of imitation was termed a *mime*. In contrast to a meme, a mime does not as ably meet the criteria of fecundity, high-fidelity replication and longevity required to qualify as a successful life-sustaining replicator. With memes the competition between observable populations of patterns is more fierce than in the case of mimes. Meaning and language account for the difference between the behaviour of pre-linguistic mimes, e.g., the rice washing of Japanese macaques or the elaborate songs

of whales, and the comportment of post-linguistic mimes, e.g., music, clothing fashions, dancing styles. In ethology, the term *culture* has come to be applied to complex learnt behaviours transmitted between conspecifics in numerous species other than our own. This usage is apt, but an essential difference remains between the semiotically enriched culture of our species and the mimetic culture of other species, which are not inhabited by a language organism. Other species lack memes in the Leiden sense of a Saussurean sign.

Mimes behave differently once they are awash in a sea of linguistic meanings with their multitudinous neuronal associations and interconnections. Our patterns of imitation as humans are more elaborate because our mimetic culture has been semiotically enriched and enmeshed with our inordinately more complex language-mediated or memetic culture. Yet the theme of Beethoven's 9th symphony none the less remains a mime, and is not a meme. Music is a paralinguistic phenomenon that is causally intimately connected with the evolutionary emergence of language, but music is not language.

Many people today use the word *meme* as just a trendy word for idea, and it has long been appreciated that ideas spread, and that some ideas spread more successfully than others. Perhaps most people can live quite happily without the *meme* concept. My use of the term *meme* in the Leiden sense, reiterated here, dates from 1983. My teacher Frederik Kortlandt, however, used to discourage my use of the term *meme* because of the inadequacy of the Oxford definition. He disputed the utility of using the term in my redefined sense on account of the availability of other terms, e.g., Saussurean sign, and he warned about the confusion that might arise from using the term in two competing definitions, i.e., *sensu Lugdunensi* vs. *sensu Oxoniensi.* Yet as I have shown above, the Leiden definition is more in keeping with the history of thought about cultural evolution. Moreover, the Leiden redefinition of the meme can heal the fuzziness which debilitates some of the burgeoning discourse on memetics.

Characteristically, meanings travel in packs within which a hierarchical structure obtains. As I have pointed out before, the idea that, for example, 'America is one nation, under God, with liberty and justice for all' is not a meme. This sentence is a syntactically articulate idea composed of a number of constituent lexical and grammatical memes, and this idea and its constituent parts are subject to Darwinian natural selection. The decomposability of units of function, such as words, phrases, sentences and narrative, is a central feature of linguistic phenomena and underscores the need for analysis to be conducted at the different levels of granularity traditionally distinguished in linguistics. The smallest structural units in language, viz. phonemes, are smaller than – though also sometimes equal in size to – the smallest functional units, viz. single morphemes and monomorphemic words.

In a strand of DNA, sets of individual base pairs together form the three-nucleotide sequences known as codons, which each code for a specific amino acid.

Even phonologically streamlined languages like Rotokas, spoken on Bougainville island, and Pirahã, spoken along the Maici river in Amazonia, have phoneme inventories greater than the four-letter alphabet of nucleotides which make up the genetic code, even though both languages have fewer phonemes than what Alfredo Trombetti reckoned to be 'il minimo di suoni che si possono con ogni probabilità attribuire alle più antiche fasi del linguaggio umano' (1905: 209).

Whilst the four nucleotides could perhaps be seen as the analogues of phonological features, the repertoire of 64 possible codons is of more or less the same size as an average phoneme inventory. So, are memes really the precise analogues of genes? Linguistic signs have the nature of non-constructible sets, and the reservoir of linguistic signs is potentially infinite in size. What precisely are genes? Are genes too non-constructible sets? Analogies can lead to muddled thinking if a conceit is overextended. We should not lose sight of the fact that semiotic entities are essentially different in nature and inherently different in their dynamics from macromolecules in numerous ways.

Another useful way of thinking about the language-borne units of cultural evolution was proposed by Kortlandt in 2003. The units of meaning are neuronal configurations which behave like a group of ants in an anthill or like the termites of a termite colony. Ant foraging is perhaps a more apt model for the exploratory behaviour of linguistic meanings in the human brain. Anthills are characterised by intricate patterns of exploratory behaviour which give the appearance of being the outcome of a careful overall pathfinding strategy, but which in fact result from numerous relatively simple responses by individual ants to the availability of potential food supplies. Ants leaving the nest secrete trails of pheremones which they follow back to the nest. Ants who have found food secrete more volatile pheremones. Trails that are not reinforced often enough by pheremones tend to evaporate after a while. As a consequence, some ants regularly stray away from weak trails and wander off in a random fashion.

The connectivity of semiotic neuronal groups is likely to operate on the same exploratory principle as ant foraging. Such an exploratory mechanism may be an optimal way of finding targets in a complex environment. Neuronal groups are continually subject to modifications brought about by changes in the environment. At the same time, neuronal exploration establishes myriads of new linkages between neuronal groups. The opportunities for generating novel contingencies are thus constantly multiplied. The exploratory mechanism fulfils a physiological function that remains vital in a forever changing environment. Rather than meme, the term *deme* has been suggested by Kortlandt to denote a functionally coherent configuration of neuronal groups which constitute the neuroanatomical instantiation of a meaning.

Meanings that have colonised a human brain seek to reproduce through meaningful contacts. Just as an ant or termite that has strayed far afield may discover a new source of food, so too meanings find new conceptual havens from which to proliferate.

The dynamics of this process yields vast repertoires of linguistics meanings. During reproduction in the process of transfer from one host to another, a meaning is reduced and, as it were, stripped of its connotations and associations, which must be constructed anew in the brain of the new host. Just as a human is reduced to a haploid sperm cell in the process of reproduction, the isofunctional set of neuronal configurations constructed in the brain of a new host is unique and microanatomically specific to that individual. Just as the needs and prerogatives of an ant colony supersede those of the individual ant, language and linguistically mediated thought shape human societies and supersede the interests of the individual.

4. The nature of the beast

The beast in the brain is a complex organism in its own right and has a high degree of autonomy. We cannot change the grammatical structure of language or fundamentally change its lexicon by an act of will, even though we might be able to coin a new word or aid and abet the popularity of a turn of phrase. Language changes, but seldom because we want it to. We are inoculated with our native language in our infancy. Like any other life form, language consists of a self-replicating core. The units of this self-replicating core are the isofunctional neuronal correlates of signs in the sense of Ferdinand de Saussure, i.e., of meanings and of their associated phonological forms. So, is language a parasite or a mutualist? The architecture of language and the intricate dynamics of the relationship between the biological host and its semiotic symbiont make the answer a complex one.

Our species has overrun the planet. A conventional measure of success for a species is reproductive fitness, and ours has manifestly been enhanced by language, whilst at the same time language thrives through us. By this criterion, therefore, language is a mutualist symbiont. If language were to be a parasite, then why has it not led to the extinction or at least attrition of our species? As Kortlandt has darkly hinted in this context, time will tell. Moreover, he stresses that language is our own undoing even now, for throughout history and in each of our daily lives our most vexing problems derive from language.

Language remains largely impervious to the well-being of man, and it colours and even stunts the perceptive faculties of its hominid host. Certainly, from the perspective of language, human brains are tools for the reproduction of language. Our grey matter has been recruited for the propagation of linguistic signs through the relentless proliferation from host to host of isofunctional neural constructs. The idea that language exerts an unfavourable effect on perception itself and blinds us to reality is an old idea already espoused by Bertus Brouwer and Frederik van Eeden. Language shapes our conceptual reality, yet there is a complex relationship between language as such and

language-borne ideas. Whether or not the capricious nature of non-constructible sets portends our doom as a species, two other issues are relevant to an understanding of how the relationship between language and man straddles the distinction between mutualism vs. parasitism.

One issue is whether or not language debilitates its hominid host. We humans are inoculated with language at birth. Language infests our brain and stays with us until we are entirely brain-dead. Our brains teem with linguistic signs, and each time a linguistic form with its associated meaning is activated in our brain, a Darwinian generation time has elapsed in terms of the neuronal group selection which characterises the rapid life cycle of linguistic signs. By analogy with biological models, it has been my contention that language itself is a mutualist, whereas not all meanings borne by language are mutualists.

As in any symbiotic relationship, models predict that categories of meaning which are vertically transmitted from the parent host to his or her infant offspring are more likely to be mutualistic in nature. Such are the grammatical categories of a language and much of the core vocabulary which is structural to a given language. These constellations of meaning construct our reality and shape our perceptions in pervasive and insidious ways. Yet by and large the grammatical and lexical core of the language acquired in infancy collectively enhances the reproductive fitness of the hominid host.

In contrast with language as such, categories of meaning borne by language that we acquire subsequently and that are readily transmitted horizontally from host to host within a single host generation are less likely to be beneficial. Whereas some linguistic signs may be highly salubrious, others may be lethal to the host and devastating to the host community at large. Jihad, racial purity, proletariat, religious tenets and various other brands of political correctness are obvious examples of pathological ideas, but in fact all horizontally transmitted thoughts are potentially dangerous and parasitic. We live the myths and ideas that impinge upon us and that wash across our societies. Religion is a disease of language. What else does Symbiotic Theory enable us to predict? Symbiosism predicts God, hypocrisy, suicide, ideologies, rites and rituals, sports, the supernatural, theatre, crusades and jihads and numerous other cultural and psychological phenomena, both delightful and baneful, that result from language and make us uniquely human, marking our species as an anomaly in the biological world.

The distinction between the grammatical and lexical core of a language with which an infant human is inoculated and all the language-borne notions that the person acquires later in life is no sharp dichotomy but a fuzzy gradient. Applied to the language organism, the point of the distinction between the vertical and horizontal propagation of linguistic signs from host to host is merely that language-borne notions of reality such as *infidel*, *Ahnenpaß* or *kosher* will have a greater likelihood of being malevolent than meanings such as the present perfect tense, the zero morpheme for singular number in nouns, or lexical items such as *mother*, *hungry* and *water.* Meanings and

syntactically articulated constellations of meanings may be wholesome, indifferent to the well-being of the host, or debilitating. This can only be judged by the effects of linguistic signs, not by their appeal, which is no more than an index of their contagiousness and no indication of their truthfulness.

5. From symbiosism to symbiomism

> 'When *I* use a word,' Humpty Dumpty said, in rather a scornful tone, 'it means just what I choose it to mean – neither more nor less.' 'The question is,' said Alice, 'whether you *can* make words mean so many different things.' 'The question is,' said Humpty Dumpty, 'which is to be master – that's all.'

A second issue is whether we are at the mercy of language. In his seminal article on the language parasite, Kortlandt stated: 'The view of language as a tool of the human species is less well-founded than its converse. The question is, in Humpty Dumpty's words, which is to be master' (1985: 478). I agree that we are at the mercy of language, but just who are we? Certainly, we are not just the hominid host, as the sad example of feral children teaches us (Ball 1880; Burnett 1784; de la Condamine 1755; Dresserus 1577; Itard 1801, 1894; Mason 1942; Singh & Zingg 1942; Sleeman 1858; Squires 1927; Rauber 1885). These soulless children are not fully human, though they are no doubt entirely hominid.

We are not just flesh and blood, we are what we believe. We are symbiomes of body and soul. Our species constitutes a unique type of symbiome in the natural world because of the singular and still quite primitive nature of the semiotic symbiont, language. The dual biological and semiotic mechanics of the symbiome are the key to understanding human mental health. Symbiomism is the school of philosophy which understands our human identity as symbiomes of a biological and a semiotic symbiont. Man is both the hominid host and the language that dwells in his brain and that mediates much of his thinking. Good health is the state in which both constituent symbionts are healthy and abide in some sort of happy equilibrium.

Our body is that of a particular variety of great ape with all its social primate propensities, equipped with a brain which has grown bloated in a long process of coevolution with language. Our soul is the language organism which resides within our skull along with everything inside our brain that is mediated by language. The moment on the 5th of April 1887, when suddenly and heart-rendingly 'the mystery of language was revealed to' her, Helen Keller would subsequently describe as her 'soul's sudden awakening' (1905: 23). We are incomplete without language. The colonisation of an australopithecine brain by language was the symbiogenesis that yielded the first human beings.

The controversy about parasitism vs. mutualism boils down to the question of what makes us human. On the matter of our identity as a species, Wilhelm von Humboldt

observed: 'Der Mensch ist nur Mensch durch Sprache' (1822: 244). The issue of whether the language organism or its hominid host has the upper hand begs the question of our very identity. When Humpty Dumpty asks who is to be master, how much does it really matter? If it feels good to live in a linguistically constructed reality, can this opium really be so bad for us? Of course, whenever we are driven to immolate ourselves for some abstract ideal, or to kill ourselves and murder others for the sake of some belief system, then this question becomes more pressing.

We are as much our essentially linguistic soul as we are its corporeal hominid host. Being healthy involves keeping both components of a symbiome happy. Our brain houses a consciousness which sustains the illusion of a thinking self with a free will. In reality, our feelings, thoughts, yearnings and behaviour are the outcome of the jostle and interplay of the biological propensities and lust for creature comforts of the human host in symbiotic association with a capricious linguistic symbiont which serves as the vehicle for the ideas waging war within us. So when we speak, who is doing the talking?

References

Ball, Valentine. 1880. *Jungle Life in India, or the Journeys and Journals of an Indian Geologist.* London: T. de la Rue.

de Bary, Heinrich Anton. 1879. *Die Erscheinung der Symbiose: Vortrag gehalten auf der Versammlung Deutscher Naturforscher und Aerzte zu Cassel.* Straßburg: Verlag von Karl J. Trübner.

van Beneden, Pierre Joseph. 1876. *Die Schmarotzer des Thierreichs*. Leipzig: F.A. Brockhaus.

Blackmore, Susan. 1999. *The Meme Machine*. Oxford: Oxford University Press.

Blum, Harold F. 1963. 'On the origin and evolution of human culture', *American Scientist,* 51 (1): 32–47.

Burhoe, Ralph W. 1967. 'Five steps in the evolution of man's knowledge of good and evil', *Zygon,* 2: 77–96.

Burnett, James [Lord Monboddo]. 1784. *Antient Metaphysics, Containing the History and Philosophy of Meng, Vol. III*. Edinburgh: J. Balfour.

Changeux, Jean-Pierre. 1983. *L'homme neuronal.* Paris: Librairie Arthème Fayard.

Cloak, F.T. 1975. 'Is cultural ethology possible', *Human Ecology,* 3 (3): 161–182.

de la Condamine, Charles-Marie. 1755. *Histoire d'une jeune fille sauvage, trouvée dans les bois à l'âge de dix ans.* Paris: Madame Hecquet.

Darwin, Charles Robert. 1871. *The Descent of Man and Selection in Relation to Sex* (2 vols.). London: John Murray.

Dawkins, Richard. 1976. *The Selfish Gene.* Oxford: Oxford University Press.

Dawkins, Richard. 1989. *The Selfish Gene* (new edition). Oxford: Oxford University Press.

Dresserus, Matthaeus. 1577. *De Disciplina Nova et Veteri: Tam Domestica, quàm scholastica; Ad consolandum erudiendumq[ue] parentes, praeceptores, ac liberos.* Basel: Ex Officina Oporiniana.

van Driem, George. 2001a. *Taal en Taalwetenschap.* Leiden: Research School of Asian, African and Amerindian Studies CNWS.

van Driem, George. 2001b. *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region with an Introduction to the Symbiotic Theory of Language* (2 vols.). Leiden: Brill.

van Driem, George. 2003. 'The language organism: The Leiden theory of language evolution', in Jiří Mírovský, Anna Kotěšovcová and Eva Hajičová, Eds, *Proceedings of the XVIIth International Congress of Linguists, Prague, July 24–29, 2003.* Prague: Matfyzpress vydavatelství Matematicko-fyzikální fakulty Univerzity Karlovy.

van Driem, George. 2004. 'Language as organism: A brief introduction to the Leiden theory of language evolution', pp. 1–9 in Ying-chin Lin, Fang-min Hsu, Chun-chih Lee, Jackson T.-S. Sun, Hsiu-fang Yang & Dah-ah Ho, Eds, *Studies on Sino-Tibetan Languages: Papers in Honor of Professor Hwang-cherng Gong on his Seventieth Birthday* (Language and Linguistics Monograph Series W-4). Taipei: Institute of Linguistics, Academia Sinica.

van Driem, George. 2005. 'The language organism: The Leiden theory of language evolution', pp. 331–340 in James W. Minett & William S-Y. Wang, Eds, *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics.* Hong Kong: City University of Hong Kong Press.

van Driem, George. forthcoming. *The Language Organism*.

Edelman, Gerald M. 1987. *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books.

Famintsyn, Andrej Sergeevič. 1889. Beitrag zur Symbiose von Algen und Thieren, *Mémoires de l'Académie Impériale des Sciences de St. Pétersbourg,* XXXVI (16): 1–36.

Famintsyn, Andrej Sergeevič. 1893. 'O sud'bě zeren" xlorofilla v" semenax" i prorostkax" ', *Trudy Botaničeskoj Laboratorii Imperatorskoj Akademii Nauk',* 5: 1–16.

Famintsyn, Andrej Sergeevič. 1907. 'O roli simbioza v" èvoljucii organizmov" ', *Trudy Botaničeskoj Laboratorii Imperatorskoj Akademii Nauk",* 9: 1–14.

Frank, Albert Bernard. 1885. 'Ueber die auf Wurzelsymbiose beruhende Ernährung gewisser Bäume durch unterirdische Pilze', *Berichte der Deutschen Botanischen Gesellschaft*, III: 128–145.

Hebb, Donald Olding. 1949. *The Organization of Behavior: A Neurophysiological Theory.* New York: John Wiley & Sons.

Henry, Sidney Mark. 1966, 1967. *Symbiosis* (two vols.). New York and London: Academic Press.

Hoagland, Hudson. 1964. 'Science and the new humanism: Science and purpose are related to man's unique ability as an ethical animal to control his evolution', *Science,* 143 (3602): 111–114.

Hugo, Victor. 1856. *Les contemplations* (2 vols.). Bruxelles: Lebegue.

von Humboldt, Wilhelm. 1812. 'Ankündigung einer Schrift über die vaskische Sprache und Nation, nebst Angabe des Gesichtpunctes und Inhalt derselben', pp. 1–8 [i.e., eight numbered pages following p. 472] in F. Delbrück, C.G.A. Erfurdt, J.F. Herbart, K.D. Hülmann, J.F. Krause & J.S. Vater, Eds, *Königsberger Archiv für Philosophie, Theologie, Sprachkunde unde Geschichte.* Königsberg: Friedrich Nicolovius.

von Humboldt, Wilhelm. 1822. 'Ueber das vergleichende Sprachstudium in Beziehung auf die verschiedenen Epochen der Sprachentwicklung', *Abhandlungen der Königlichen Preussischen Akademie der Wissenschaften,* historisch-philosophische Klasse, aus den Jahren 1820 und 1821: 239–260.

Huxley, Julian S. 1962. 'Evolution: biological and human', *Nature,* 196 (4851, 20 October 1961): 203–204.

Itard, Jean-Marc Gaspard. 1801 [i.e., an X]. *De l'éducation d'un homme sauvage, ou Des premiers développements physiques et moraux du jeune sauvage de l'Aveyron.* Paris: Goujon fils.

Itard, Jean-Marc-Gaspard. 1894 [posthumous]. *Rapports et mémoires sur le sauvage de l'Aveyron, l'idiotie et la surdi-mutité.* Paris: Félix Alcan.

Keller, Helen Adams. 1905. *The Story of My Life*. New York: Doubleday, Page & Company.

Kortlandt, Frederik Herman Henri. 1985. 'A parasitological view of non-constructible sets', pp. 477–483 in Ursula Pieper and Gerhard Stickel, Eds, *Studia linguistica diachronica et synchronica: Werner Winter sexagenario anno* MCMLXXXIII *gratis animis ab eius collegis, amicis discipulisque oblata.* Berlin: Mouton de Gruyter.

Kortlandt, Frederik Herman Henri. 1998. 'Syntax and semantics in the history of Chinese', *Journal of Intercultural Studies*, 5: 167–176.

Kortlandt, Frederik Herman Henri. 2003. 'The origin and nature of the linguistic parasite', pp. 241–244 in Brigitte Bauer & Georges-Jean Pinault, Eds, *Language in Time and Space: A Festschrift for Werner Winter on the Occasion of his 80th Birthday*. Berlin: Mouton de Gruyter.

Krause, Gottlob Adolf. 1885. 'Die Stellung des Temne innerhalb der Bantu-Sprachen', *Zeitschrift für afrikanische und oceanische Sprachen, mit besonderer Berücksichtigung der Deutschen Kolonien,* I: 250–267.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions* (International Encyclopedia of Unified Science, Foundations of the Unity of Science, Volume II, Number 2). Chicago: University of Chicago Press.

Kuhn, Thomas S. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change.* Chicago: University of Chicago Press.

Laurent, John. 1999. 'A note on the origin of "memes"/"mnemes"', *Journal of Memetics,* 3 (1): 20–21.

Maeterlinck, Maurice Polydore Marie Bernard. 1927. *The Life of the White Ant.* London: George Allen and Unwin.

Margulis, Lynn, and Karlene V. Schultz. 1998. *Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth* (third edition). New York: W.H. Freeman & Company.

Mason, Marie K. 1942. 'Learning to speak after six and one-half years of silence', *Journal of Speech Disorders*, 7: 295–304.

de Maupertuis, Pierre Louis Moreau. 1756. 'Dissertation sur les différents moyens dont les hommes se sont servis pour exprimer leurs idées', pp. 435–468, vol. III in *Œuvres* (nouvelle édition corrigée et augmentée, 4 vols.). Lyon: Jean-Marie Bruyset.

Mereschkowsky, Constantin [i.e., Konstantin Sergeevič Mereškovskij]. 1905. 'Über Natur und Ursprung der Chromatophoren im Pflanzenreiche', *Biologisches Centralblatt*, XXV (18): 593–604.

Mereškovskij, Konstantin Sergeevič. 1909a. *Teorija dvux" plazm" kak" osnova simbiogenezisa, novogo učenija o proisxoždenii organizmov"*. Kazan': Tipografija Imperatorskago Kazan'skago Universiteta [published the same year in instalments in German translation as (1909b)].

Mereschkowsky, Constantin [i.e., Konstantin Sergeevič Mereškovskij]. 1909b. 'Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen', *Biologisches Centralblatt,* XXX (8): 278–288, XXX (9): 289–303, XXX (10): 321–347, and XXX (11): 353–367.

Rauber, August. 1885. *Homo sapiens ferus, oder die Zustände der Verwilderten und ihre Bedeutung für Wissenschaft, Politik und Schule. Biologische Untersuchung von Dr. A. Rauber.* Leipzig: Denicke.

Rodríguez-Ezpeleta, Naiara, Henner Brinkmann, Suzanne C. Burey, Béatrice Roure, Gertraud Burger, Wolfgang Löffelhardt, Hans J. Bohnert, Hervé Philippe & B. Franz Lang. 2005. 'Monophyly of primary photosynthetic eukaryotes: Green plants, red algae and glaucophytes', *Current Biology*, 15: 1325–1330.

Salverda, Reinier. 1998. 'Is language a virus? Reflections on the use of biological metaphors in the study of language', pp. 191–209 in Mark Janse and An Verlinden, Eds, *Productivity and Creativity. Studies in General and Descriptive Linguistics in Honor of E.M. Uhlenbeck*. Berlin: Mouton de Gruyter.

Salverda, Reinier. 2003. Letter to the Editor, *New Scientist* (1 February 2003), 2380: 25.

Schatzmann, Jürg. 1968. *Richard Semon (1859–1918) und seine Mnemetheorie.* Zürich: Juris Druck Verlag.

Schimper, Andreas Franz Wilhelm. 1883. 'Ueber die Entwickelung der Chlorophyllkörner und Farbkörper', *Flora oder Allgemeine Botanische Zeitung,* 41 (7): 105–114, 121–131, 137–146, 153–162.

von Schlegel, Friedrich. 1808. *Ueber die Sprache und Weisheit der Indier: Ein Beitrag zur Begruendung der Alterthumskunde, nebst metrischen Uebersetzungen indischer Gedichte.* Heidelberg: Mohr und Zimmer.

Schleicher, August. 1863. *Die Darwinsche Theorie und die Sprachwissenschaft: Offenes Sendschreiben an Herrn Dr. Ernst Häckel, a.o. Professor der Zoologie und Direktor des zoologischen Museums an der Universität Jena.* Weimar: Böhlau.

Schmitz, Friedrich Karl Johann. 1882. *Die Chromatophoren der Algen. Vergleichende Untersuchungen über Bau und Entwicklung der Chlorophyllkörper und der analogen Farbstoffkörper der Algen*. Bonn: Verlag von Max Cohen & Sohn (Fr. Cohen)

Schuchardt, Hugo. 1919a. 'Sprachursprung. I' (vorgelegt am 17. Juli 1919), *Sitzungsberichte der Preussischen Akademie der Wissenschaften,* XLII: 716–720.

Schuchardt, Hugo. 1919b. 'Sprachursprung. I' (vorgelegt am 30. Oktober 1919), *Sitzungsberichte der Preussischen Akademie der Wissenschaften,* XLII: 863–869.

Semon, Richard Wolfgang. 1911 [1904] *Die Mneme als erhaltendes Prinzip im Wechsel des organischen Lebens* (dritte Auflage). Leipzig: Wilhelm Engelmann.

Semon, Richard Wolfgang. 1922 [1909]. *Die mnemischen Empfindungen in ihren Beziehungen zu den Originalempfindungen* (zweite Auflage) Leipzig: Wilhelm Engelmann.

Singh, Joseph Amrito Lal & Robert M. Zingg. 1942. *Wolf Children and Feral Man*. New York: Harper.

Sleeman, William Henry. 1858. *A Journey through the kingdom of Oude in 1849–1850 by Direction of the Right Honourable the Earl of Dalhousie, Governor-General, with Private Correspondence Relative to the Annexation of Oude to British India*. London, R. Bentley.

Squires, Paul. 1927. '"Wolf children" of India', *American Journal of Psychology*, 38: 313–315.

Swanson, Carl P. 1973. *The Natural History of Man*. Englewood Cliffs, New Jersey: Prentice-Hall.

Trombetti, Alfredo. 1905. *L'unità d'origine del linguaggio.* Bologna: Libreria Treves di Luigi Beltrami.

Trombetti, Alfredo. 1925. *Le origini della lingua basca* (Memoria presentata alla Reale Accademia delle Scienze dell'Istituto di Bologna nella Sessione del 24 Novembre 1923). Bologna: Cooperativa Tipografica Azzoguidi.

Ujhelyi, Mária. 1998. 'Long-call structure in apes as a possible precursor for language', pp. 177–189 in James R. Hurford, Michael Studdert-Kennedy & Christ Knight, Eds, *Evolution of Language: Social and Cognitive Bases.* Cambridge: Cambridge University Press.

Welby [i.e., Welby-Gregory], Victoria Alexandrina Maria Louisa Stuart-Wortley. 1896. 'Sense, meaning and interpretation', *Mind*, v: 24–37, 186–202.

White, Leslie A. 1959. 'The concept of culture', *American Anthropologist*, 61: 227–251.

Wiedenhof, Jeroen Maarten. 1996. 'Nexus and the birth of syntax', *Acta Linguistica Hafniensia*, 28: 139–150.

Some speculations on the evolution of language, and on the language of evolution

Paul Whitehouse
London, England

The need for languages to be complete systems suggests that language evolution will be episodic rather than gradual, with periods of stability punctuated by periods of innovation. The example of nicaraguan sign language suggests that such developmental phases can be very rapid. Rapid evolution can create genetic bottlenecks. The differentiation between Homo sapiens and H. erectus was probably the result of a genetic bottleneck arising from the invention of language by part of the H. erectus population. This occurred in Africa, probably between 200 and 100 kyrs BP. The possession of fully complex language by every living human community suggests that it had already been developed by our H. Sapiens ancestors before they left Africa. The possession of complex language has allowed humans to participate in their own evolution in ways impossible for other animals.

1. Stability versus innovation

Nineteenth century linguists were content to classify languages as more or less primitive according to their typology, but nowadays – not least because of the racist assumptions that underpinned such classifications – linguists treat all languages as being at the same stage of development, and avoid the suggestion that some extant language might be in any way "primitive". Nevertheless, if human language did not come into existence fully formed, it must have undergone some sort of development process, passing through stages to which the term "primitive" *would* apply. If so, what might such "primitive" languages have looked like?

The obvious question is: primitive compared to what? What characterises the stage we consider all modern languages to have reached? One thing they all have in common is that they are all complete systems. They are able to express every relationship of person, time and place that might reasonably arise, and have some means, however clumsy, of resolving those ambiguities that the language may contain. This accounts for a second feature they share in common, which is that they are all stable. The universality

of language change shows that it is a state of dynamic equilibrium, but it is equilibrium nonetheless. One way or another, they all "work".

The reason why there are no languages that do *not* work seems fairly obvious: because the use of a language that does not work would be such a severe handicap. However, if the reason languages work is because they are complete systems, but the transition from one complete system to another must involve a period when the system is incomplete, this suggests that such transitions must have been rapid. It also suggests that there can have been no "working" language available for the speakers to switch to while their own was in transition, other than the earlier "complete" version of their own.

What precisely do I mean by a complete system? One such example would be the use of what might be called "absolute person". By this I mean a system in which everyone and everything is referred to by name, such that Johnny says to Joe "Johnny gave the apple to Joe" rather than "I gave it to you." Such a system is complete in that it is always clear who or what is being referred to, and that makes it stable. The system generates no immediate problems requiring it to be changed. It could even be extended by the limited use of a single all-purpose pronoun, but any further elaboration would need to embrace "relative person", a system in which the pronouns denote different entities depending upon who is speaking. Such a system is unworkable unless it covers every permutation, and what this suggests is that an incomplete pronominal system is inherently unstable and would not be allowed to persist. It would either be expanded to include all permutations or abandoned. I would further argue that the decision would have to be made very quickly.

The same must also have been true of systems expressing "relative time" and "relative place", in which the focal point of what is being said is not necessarily here or now, or other innovations such as the use of hypotheticals. No incomplete system would have been permitted to continue without the gaps being quickly filled in and stability restored.

This has implications for the chronology of language evolution. Stability demands no urgency, and a simple system with no internal contradictions could continue indefinitely, whereas incomplete systems must quickly be made complete or as quickly abandoned. One would therefore expect that the development of language involved periods of rapid change separated by long periods of stasis. The question of how long these various phases lasted is impossible to answer, except to say that the periods of stasis could as easily have lasted tens of years as tens of thousands of years. Periods of change, on the other hand, may have been very short indeed. It is not inconceivable that fundamental shifts between levels of complexity could have been achieved in a single generation.

This is, of course, a description of successful transitions only. There may also have been attempts at linguistic development that failed, either because speakers reverted to the earlier version of the language, or because they persisted in using a transitional

version of the language that didn't work and were penalised for it. Such failures would have left no trace of themselves, but the successful transitions are not documented either. Every language we know of was already a complete complex system at the time of its first documentation, and I know of no language in which a fundamental shift in *functional* complexity has been documented. So, do these arguments have any evidential basis, and, if not, why should they be accepted? One reason may be found in the example of the Nicaraguan sign languages. Although these are signed rather than spoken, there seems no reason to think them irrelevant to how spoken language may have developed; the language processing mechanisms are the same.

In 1980 a large group of deaf children (200 or so) were brought together to be taught speaking (in Spanish) and finger-reading. These children had not previously had any contact with other deaf people and knew only the homesign systems they had learnt at home. Each of these homesign systems was different from every other and all were very crude. The formal tuition was not successful at all, but away from the classroom the children themselves developed a single sign language "richer in content and structure" than any one of the homesign systems the individual children had contributed to the mix (Morford & Kegl 2000, p. 367). The new sign language is known as *Lenguaje de Signos Nicaragüense*, or LSN. As new cohorts of children arrived and were exposed to LSN from a much younger age than the original group of children, they modified the language to produce a much more complex system known as *Idioma de Señas Nicaragüense* (ISN). It is the only known example of a complex language being invented by people who had no previous contact with an existing complex language on which they might model their creation.

Unfortunately, the process was far advanced before even the teaching staff realised what had happened, so the precise sequence of developments was not documented while it was happening. It is therefore not possible to say how far the steps by which the language was built fit my claim that incomplete systems are unstable and in practice self-completing – and it would be unethical to re-create experimentally the conditions in which NSL and ISL were invented!

One feature of this language creation event that requires no experimental confirmation, however, is its speed. Whatever sequence the language building process followed, it was certainly quick, and this reinforces the possibility that fundamental shifts in human linguistic evolution may have been similarly rapid. It also suggests that children are the main drivers of such shifts. The first version of the language (LSN) was created by children who were well past the age at which language acquisition normally begins. The children who came after them were much younger, some exposed to LSN (and later ISN) from birth. It was this second generation, having begun at a younger – and therefore more neurologically flexible – age, who, in the time it took them to reach adulthood, transformed the language to a level comparable with other "first-language" systems (spoken as well as signed), and were its first native speakers.

2. Leaps and pauses

The speed with which NSL and ISL were invented seems to support the idea that the transitional phases of language evolution could have been extremely swift, but the circumstances are not necessarily the same. For instance, LSN was created by humans who had an advanced language capacity but no comparably advanced language for them to acquire. How often can this situation have arisen historically? The example of the Nicaraguan children suggests that, once invented, human language would have been taken as far as the prevailing neurophysiological capacity allowed. Thus linguistic sophistication would have kept pace with neurological development, as long as there were children exposed to it (and to each other). That makes it unlikely that an entire hearing community would have found itself with an advanced capacity for language and no comparably advanced language for it to learn – except in the case of the first language.

Lenneberg (1966, p. 225) pointed out that the level of linguistic ability is uniformly high across the entire human population, and this suggests that at some point the lack of a minimum level of language capacity has been severely penalised. Although a wide range of linguistic ability may have existed in the pre-linguistic population, in the absence of a language to reveal it, once such a language was invented – and the case of the Nicaraguan children suggests that it could have happened very quickly – those lacking that basic level seem to have vanished from the ancestral gene pool.

There are a number of ways in which such linguistic incapacity may have manifested. Firstly, those members of the innovating group who lacked the neurological capacity for language will have failed to acquire it, even though exposed to it as children. Also, the parents of those who were able to master the new language may not have been able to master it themselves, having been exposed to it too late in life. The same is also true of adults in other groups. Consequently, even if some of their offspring were capable of learning the new language, they would not have done so because their parents were too old to learn it from their speaking neighbours, and the children had no one else to learn the language from.

In this way a divide between speakers and non-speakers could have appeared both within the group and between groups – but it may also not have appeared at all. As long as linguistic incapacity did not conflict with group cohesion, the first speaking group could have continued to include non-speakers who passed that incapacity on to succeeding generations. Also, the wider social network of the group may have involved sufficient contact with the children of other groups for them to learn what their parents could not, allowing the new language to spread throughout the species. Again, this would have required speakers and non-speakers to co-exist and interbreed, with non-speakers sharing whatever benefits the possession of language may have brought.

Thus we have two scenarios. In the first, speech is confined to part of the group in which it first appears, and the resulting social division is reinforced genetically as speakers breed only with other speakers and non-speakers only with non-speakers. We know that modern humans are sensitive to very small deviations from linguistic norms, and it is not impossible that the first speakers may have been repelled by those whose speech was clumsy or non-existent. In this model the first speaking population would have been very small, but the uniformly high level of language capacity left by the disappearance from the gene pool of heritable linguistic deficiency would have allowed them to avoid being penalised for their lack of numbers.

In the second scenario, the capacity for language spread slowly from group to group, with individuals who lacked the innate capacity for language still being born into the group and accepted socially. Thus the variation in language capability remained wide at first, with the present uniformly high level of ability emerging only gradually, and perhaps only after hundreds or thousands of years.

The critical factor is, what was transmitted from one generation to another? There are two agencies of transmission: learning and heredity. The first requires infants to be exposed to language, a cultural practice the efficacy of which may have depended upon the heritable characteristics of the children, but which, in the absence of any other physical constraint, could as easily have been continued as discontinued. Heredity, on the other hand, is entirely constrained, and the spread of heritable language ability (or inability) was entirely a matter of who bore children by whom. This is the key to what follows, because until any development has *reproductive* consequences its impact on evolution may be nil.

In the first instance this is determined by who survives to sexual maturity and who procreates. Whether or not language is a determining factor in this equation depends upon social attitudes to it, and on its practical applications. In a group where some can speak and some cannot, attitudes to language may have been harsh, with contempt on one side and resentment on the other. Even with patience and compassion on the part of the speakers, and equanimity among the non-speakers, any such division into the "cans" and "can-nots" must have placed a great strain on the group. There are many ways in which such attitudes may have affected relations between individuals, and the withdrawal of support from individual group members could as easily have hurt speakers as non-speakers. The group as a whole may have been penalised by its failure to cooperate, or may have split into distinct speaking and non-speaking groups, both of which may have survived – or neither. Non-linguistic considerations may have led speakers and non-speakers to continue interbreeding, in which case perhaps the heritable capacity for language became more prevalent – or was so diluted that it disappeared. The smaller the group, the greater the influence of individual members and their choices. Much may have depended upon the non-linguistic attributes of those individuals, whether they survived to adulthood

or not, whether they produced offspring or not, whether they encouraged the new language, or not.

Here too the experience of the Nicaraguan children is a reminder of how uncertain the process may be. Bear in mind that those children with only a homesign system and only (hearing) family members to use it with did not modify it. Indeed, when the development of NSL began in 1980, 50 of the 200 deaf children had already been at the school for three years without having made any attempt to create a common sign language. The reason was that they had not been in contact with one another during that time, suggesting that contact with children of the same age and in the same situation is necessary to the process. Presumably this is because only contact with others reveals the deficiencies of the language and creates the necessary pressure to innovate. Such pressure seems not to have been generated by contact with hearing individuals who, as adults, were presumably not sensitive to the inadequacies of the homesign systems they used with their deaf children. Thus peer interaction between children seems to be crucial in a way that interaction with adults is not.

It does, however, show how easily adults could have interfered with the early stages of language creation. Children are the main drivers of linguistic change, as the example of LSN and ISN shows, but they are also the most physically and psychologically vulnerable members of any group. It is just as conceivable that non-speaking adults were able to snuff out a newly-invented language, as that speaking adults marginalised non-speaking children to the extent that they never had the opportunity to reproduce. This latter alternative would of course have resulted in a group whose linguistic capacity was uniformly higher, and in which successive generations of children were more likely to push the language as far as their neurological capacity would allow.

What this flurry of contradictory alternatives is meant to convey is that the invention of language was not necessarily a simple or linear process, but may as easily have involved detours, dead ends and false starts.

The other factor in language evolution is its practical value. The possession of language has benefits that are obvious to us (group cohesion, co-operation, forward planning, resource management, and so on) but what we have in mind is the complex language we speak today. Instead of taking the advantages of language for granted, however, we need to ask exactly how complex did a language have to become in order for it to be of any use at all. Viewed from this angle, it is possible to see the earliest versions of human language as irrelevant to everyday survival. In evolutionary terms, the only important fact may have been that the possession of language was not penalised. Its role may have been entirely social, an interesting party piece, but of no more *evolutionary* significance than the ability to whistle a tune is today. In this model the crucial moment will have come with the invention of a language whose practical applications were unmistakable. The alternative scenarios outlined above would have applied in

this situation too, but on this occasion the practical benefits of a more complex language make the rapid and explosive outcomes more likely than the gradual.

3. Early languages

What form might early languages have taken in their stable states, and what shifts in complexity might have required a leap from one stable state to another? This is an easy question to answer, but whether those answers bear any relation to historical fact is another matter. Some degree of corroboration may be gained from comparisons with child language acquisition. As the De Villiers put it, "in the first two years alone, the child makes a leap that recapitulates evolutionary developments several aeons in the making – but it is a leap … composed of many little steps" (De Villiers & Devilliers 1979, p. 18). Of course, what the modern child experiences is not necessarily the same as what its hominid forebears experienced. For a start, children undertake language acquisition rather than language invention. What's more, it is a process undertaken by the neurologically immature. Though the pace at which pre-existing language is acquired is largely regulated by neurological development, the constraint is only temporary. Even when language acquisition has been interrupted, the child rapidly masters those things which it was too immature to master at the time of interruption. This suggests that there would be a significant difference between a modern child and an adult hominid, even though their brain size and level of linguistic accomplishment were measurably the same, and therefore the limitations apparent during child language acquisition are not necessarily the same as those experienced by the adult hominid. One fundamental difference is that the modern child is aware of linguistic usage that it cannot yet master itself. It is constantly being drawn on by the awareness of gaps in its own repertoire. For the adult hominid, that same repertoire may have constituted the whole of the language available to it. Thus, although the language available to each may have been the same, and delimited by the same neurological constraint, the adult hominid would not have been aware of that constraint to the same degree as the modern child – if at all. As long as its language was "complete" there would have been no obvious loose ends for the hominid infants to appy their greater linguistic flexibility to, and carry into their adult language as spurs to further development by future generations.

Nevertheless, the process of child language acquisition does give us an idea of how neurological constraints affect language capability, and these offer clues as to the stages at which human language development may have come to a temporary halt.

One such constraint involves phonology. It has often been observed that very young children have an ability to pronounce particular words that is then lost. Thus a child who pronounced the word "turtle" correctly at 15 months was pronouncing it "kurka" at 18 months. Another was saying "pretty" at 16 months, but "bidi" at

18 months (De Villiers & De Villiers 1979, p. 30). This seems to be the result of a memory constraint that is encountered when the child shifts from isolated imitations of words to treating the language as a systematic whole with a single phoneme system. It is as if, instead of simply making the sound, for which it had sufficient short-term memory, it now has to hold the rest of the phoneme system in its mind too, which leaves so little free memory that it can only reproduce a simplified version of the sound. It seems to be a universal feature that although infants are able to recognise phonetic contrasts such as /pa/ and /ba/ as early as four weeks old (Crystal 1987, p. 238) and produce a wide range of sounds when babbling, the first systematic use of words employs a phoneme inventory consisting of MA and TA, whatever the phoneme inventory of the language it is trying to acquire (Cruttenden 1979, p. 17). The extent to which this phoneme inventory is universal has been questioned, but the severe narrowing of the phonemic range early in language acquisition seems uncontroversial. What it suggests is that our hominid ancestors may have started by using single words in isolation, and that when this gave way to a phonemic system they were limited to a minimal phoneme inventory. This in turn would have restricted the available vocabulary, either qualitatively by requiring words to be longer, or quantitatively by limiting the total lexicon.

Another early manifestation of memory constraint is the use of "replacement sequences" in which a child produces a series of statements such as "daddy pat," "pat dog," "daddy dog," implying that it has formulated the sentence "daddy pat dog" but lacks the memory to reproduce it in its entirety. It may be that similar utterance patterns characterised one of the earlier stages of language. Although the modern child is attempting constructions that it has heard elsewhere and which the hominid child had not, the hominid may still have reproduced parts of the modern child's replacement sequences as statements that it was able to formulate in isolation but not combine in a single string. Whether the hominid would also have been able to formulate that longer sentence is another matter. It may be that the memory constraint placed a limit on sentence length that applied to conception as well as performance.

Each of these innovations (the use of isolated sound symbols, the development of a phonemically coded sound system, and the construction of sentences with syntax) had the potential to divide hominid populations into those who were capable and those who were not. Also, because the ability to conceptualise sentence structure and the ability to produce and interpret sound use different areas of the brain, it could have resulted in individuals who were capable in one area being penalised for weakness in others. The selective pressure of all three acting together would have had a particularly dramatic effect.

This ability to keep the entirety of the language accessible to short term memory also underpins the ways in which modern humans are able to make assumptions about the listener's ability to know what we are talking about. This affects not only the speed

at which information can be exchanged, since it permits us to avoid endless repetition, but also enhances the quality of expression up to and including the use of metaphor.

One further constraint that occurs in child language acquisition is the delayed ability to deal in hypotheticals. This ability only appears between the ages of four and five, even in languages where the required grammatical formatives are in place at an earlier age (De Villiers & De Villiers 1979, pp. 132–133). This too may involve a minimum memory requirement, which suggests that at some stage – possibly the last stage prior to our own? – human language was entirely imperative and indicative. Switching between the purely factual and the hypothetical is bound to result in misunderstanding *unless* there is a system in place that can encompass all the permutations. I would argue that here is one of those qualitative leaps that will either be achieved quickly or not at all, and will, if successful, result in a form of language that is dramatically more flexible than every other language spoken at that time.

These are just some of the stages of language acquisition that may have had counterparts in human language evolution. The modern child faces several such hurdles at any one time, and presumably our ancestors also had to master several innovations at once. However, as they lacked the modern child's awareness of the gaps in its linguistic competence, their linguistic evolution must have been subject to interruptions and pauses that the modern child does not experience.

4. The chronology of language development

The most obvious question is, when did the first language creation event occur? And, if the subsequent development of an originally crude language took place by a series of bursts, what form did these take, and when? These are things that by their nature leave no record of themselves, so the evidence we have for what happened is at best circumstantial. Nevertheless, it is possible to narrow down the options.

To begin with, our closest primate relatives are an indicator of the likely linguistic capacity of our own ancestors at the time the species separated. Assuming that hominid language ability did not subsequently lessen at any point, this can be taken as the absolute lower limit. A number of well-known experiments have been conducted in which some sign language was taught to chimpanzees. Whilst this implies that our ancestors must have had some linguistic ability, the limited progress made by the chimps seems to be entirely reliant upon human tutors who exposed them to a fully-formed language that the chimps showed no sign of being able to invent for themselves. Crucially the awareness of their own limitations that drives modern children was conspicuously absent. These experiments all used the common chimpanzee, *Pan troglodytes*. It has been suggested that their close relative the pygmy chimpanzee, *Pan paniscus*, is in many respects the more archaic and as such a better guide to what our own ancestors

were like, and indeed *P. paniscus* does seem to communicate its own desires in a more spontaneous way than *P. troglodytes* (Zihlman 1989, p. 100). Nevertheless, the difference is not enough to alter the conclusion that the first human language was invented after the split from the chimpanzee line, and that the capacity for language played no part in that split.

Where else might one look? Though archeology offers clues in the form of anatomy and artifacts, neither of them are particularly good indicators of linguistic ability. The significance of artifacts is easy to overstate. It is not difficult to see how advances such as the making of stone tools or the management of fire, or even rock art, might have been achieved solely by demonstration, with no linguistic component at all. The things for which language is indispensable are mostly abstracts that, by definition, leave no physical trace. Consequently language is almost the last thing one might expect to find in the archeological record, until the invention of writing.

The value of the anatomical evidence is also limited, because the vocal apparatus does not survive, and neither does the brain. All scientists have to go on is cranial morphology and patterns of muscle attachment and innervation that may be visible on bones, and the significance of these is also easy to overstate. Lieberman, for instance, has argued on the basis of basicranial and mandibular morphology that the larynx of European Neanderthals was not sufficiently descended for it to produce the speech of modern humans, though Neanderthals elsewhere were less handicapped in this way (Lieberman 1989, p. 412). According to Terrence Deacon this laryngeal descent began in *Homo erectus* about 1.75 million years ago (Deacon 1997, p. 409). Since this is before the separation of *H. erectus* and *H. sapiens*, it would suggest that laryngeal descent was not necessarily a factor in bringing about that separation. Equally, the fact that Neanderthals could not produce the sounds of modern languages does not mean they were incapable of developing a comparably complex language based on those sounds they could produce. After all, complexity is a function of brain activity, and the Neanderthal cranial capacity was little less than our own.

Similarly the expansion in brain size evident between the Australopithecines and *Homo habilis*, and again between *H. habilis* and *H. erectus*, predates the separation of *H. erectus* and *H. sapiens.* Both could therefore have spoken languages that shared a common (*H. erectus*) ancestor – or language may not yet have been invented at the time of separation. There are two reasons why brain size can not automatically be equated with language use. One is that it is the organisation of the brain that makes language possible as much as its size. As Lenneberg points out, "it is not so much one or the other specific aspect of the brain that must be held responsible for the capacity of the language acquisition but the way the many parts of the brain interact. Thus it is mode of function rather than specific structure that must be regarded as the proper neurological correlate of language" (Lenneberg 1966, p. 243). The second is that cranial enlargement was not necessarily selective for brain function.

The role of neoteny as a factor in human evolution is considerable, in particular the "maintenance of fetal growth rates [leading] to hypertrophy of organs developing with early positive allometry in uterine ontogeny (the brain)" (Gould 1977, p. 406). Thus the enlargement of our ancestors' brains was in the first instance the result of genetic mutation, and as with all mutations the question was, would it be penalised? This is particularly important in the case of cranial enlargement because the foetal skull can be lethal. Portman (cited in Gould 1977, p. 373) estimated that if human gestation was retarded in proportion to that of other pongoids it would last 21 months, implying that humans as a species are all premature. It is not difficult to see this as a consequence of those foetuses who left it much longer than 9 months having killed their mothers in childbirth!

Such a severely premature baby needs much more looking after, and one might expect that their low survival rates might lead to them being abandoned as a matter of course. Two factors may have countered this, however. One is a maternal attachment to all newborn regardless of their viability (common to modern humans); another is that if full-term foetuses are killing off the species and only the premature are left, these become the future of the group rather than a threat to the future of the group.

For this situation to arise, though, it needs the entire group to be affected, so the necessary mutation must have occurred in a population that was both isolated and very small. Before it became a threat, however, increasing cranial size may have brought sufficient benefits – not language related – for the *lack* of cranial size to have been progressively penalised. In that case the mutation responsible for this particular type of retardation could have become more widespread than the condition itself, depending upon how precisely it was coded genetically. Then, when cranial enlargement became lethal it would have happened over a wider area, so isolating affected groups from unaffected groups whose genes might otherwise have kept the problem in check. This in turn may have caused a temporary lowering of local population density, which left fewer rivals who might otherwise have penalised the cranially-enlarged population while it was still to discover what clever adults their premature children would grow into.

The situation is made much more complex by the fact that retardation had so many other consequences that might as easily have been burdens as blessings according to circumstance. From the point of view of language, though, it may be enough to know that *H. erectus* probably had sufficient cranial capacity for language, and that this capacity did not arise from selection on the basis of language use. Crucially it may have created one of the factors that made the invention of ISN so rapid: the capacity for language and no existing language to apply it to.

The question is: did *H. erectus* invent language? Two factors suggest that it did not – or, rather, that the *H. erectus* group that invented language evolved into *H. sapiens* as a result. Firstly, there are physical differences between *H. erectus* and

H. sapiens – thigh length, laryngeal descent, etc. – suggesting that they are separated by something more qualitatively significant than mere genetic drift. Secondly, there is the fact that the one species was everywhere replaced by the other. The size of the *H. erectus* brain implies a great capacity for language, such that if one *H. erectus* group had invented language its children could very quickly have developed it into something complex enough to be useful. It may not have been as complex as modern language, and may still have needed to pass through some of the stages suggested earlier, but it would from the outset have been capable of producing a division of speakers from non-speakers that became endogamously fixed. Once that happened, the lack of opportunity to acquire language would have been penalised as much as the lack of ability. However, the fact that one variant of *H. erectus* came to replace all the others suggests that the development of language arose from a heritable change in brain *organisation*, and that the Neanderthals, despite their large brains, were unable to master the new language regardless of opportunity.

In this model *H. sapiens* would by definition have had no pre-linguistic phase, and the evidence of archeology and mitochondrial DNA would locate the linguistic explosion from which they emerged in Africa, some time between 200 and 100 kyrs BP. The languages spoken today would all be versions of this first language, but their most recent modern ancestor would have been the one version of the many intermediate-stage languages in which the completeness that characterises all modern languages was finally achieved.

5. On the language of evolution

Throughout this paper the evolutionary implications of language development have been discussed in terms of what is penalised rather than what is favoured, and that is deliberate. I see our understanding of the evolutionary process as having been distorted by the language in which it is described, and my choice of words is an attempt to root out some of these linguistically-embedded errors.

At the heart of this distortion is that famous phrase "the survival of the fittest", which is as inaccurate as it is well-known. Natural selection is *not* the survival of the fittest, it is the elimination of the unfit – a fundamentally different process; more often than not, the Law of Nature is "if it ain't broke, don't fix it!" The shark, for instance, has not evolved fundamentally in 200 million years. Why not? Because it has never been penalised for being what it is. Our own evolutionary path can be traced right back to single-celled organisms. We are complex creatures, but some of our distant relatives are still single-celled organisms. Why? Because they have never been penalised for it – yet the comparison is always made in terms of humans having been favoured in ways that the single-celled organisms have not. This is a habit that leads inexorably towards

imputations of destiny, purpose, and design, and produces an account of evolution that is a history written by the victors. A first step towards a less partisan interpretation of the process would be if biologists were instead to describe the evolutionary process in terms of what is penalised.

A significant aid to such a reversal of descriptive practice would be a moratorium on the use of the question "why?". Scientists are forever asking "why?" when what they should really be asking is "why not?". What makes "why?" such a problem is that it conflates two separate questions: "as a consequence of what?" and "for what purpose?", and this carries within it the assumption that there must be a reason for everything, and everything must have a purpose. Moreover, this conflation of consequence and purpose into a single interrogative "why?" is found in nearly all the world's languages, which suggests that this underlying assumption has been programmed into us by our evolutionary history.

There is a strong possibility that human beings are the only creatures on earth who know far in advance that they personally are going to die. Certainly we are the only ones capable of putting it into words, which may amount to the same thing. What this also allows us to do is to conclude that ultimate death makes everything we do ultimately futile, and that discovery is potentially fatal. One of the iron laws of natural selection is that apathy is nearly always penalised, and I would argue that recognising the futility of their existance must have left many early humans apathetic enough to be penalised for it. I deliberately do not say "our ancestors", because we seem to be descended exclusively from those who were not penalised on account of their apathy, but were able – literally – to talk themselves into carrying on. The proof of this lies in the universal compulsion on the part of modern humans to find explanations for things, to find the universe meaningful, and see our lives as having some purpose. The accounts of concentration camp survivors bear this out. Again and again these recall how often a sense of purpose, a belief in some reason why they personally had to survive, made the difference between life and death. Those who could not somehow convince themselves that there was a point to carrying on did not survive. That gives us an idea of how our instinct to explain and to find meaning where there is none came to be so deeply engrained.

Where this becomes a problem is that we see purpose everywhere, and happily apply this lust to find the universe meaningful to things that may not be meaningful at all – such as the universe. What makes this dangerous is that we become susceptible to teleological arguments, and these lead us to conclusions that may be at best wrong, and at worst positively lethal. History leaves us in no doubt as to what humans are capable of when they believe themselves to be fulfilling some higher purpose. Confining ourselves to science and the factually wrong, I consider that replacing the question "why?" with the less charged question "why not?" as our primary avenue of enquiry might go a long way towards eliminating the anthropomorphism that otherwise contaminates so much intellectual activity, up to and including evolutionary biology.

Of course, anthropomorphism is only possible because humans are able to see things from the perspective of others. As with the foreknowledge of death, whilst this itself may not be unique, our ability to encode it linguistically is. One final attribute that may also be unique to humans is that we know where babies come from.

Biologists speak so routinely of the urge to reproduce that its existence is never questioned, but this too may be no more than anthropomorphism. On the contrary, I would argue that, humanity excepted, there is no urge to reproduce, only the urge to have sex. Here is the ultimate teleological argument, that because sex results in reproduction, reproduction is the purpose of sex. But how do we know that *any* other species has made the connection we have made? Might their reproductive behaviour amount to nothing more than a response to stimulus? After all, that is how a farmer persuades a bull to inseminate an artificial cow. And even where the choice of a mate appears to ensure the best possible offspring, is *that* any more than stimulus and response? The male bower bird builds a nest as part of its courtship ritual, and this is explained as the female choosing a mate on the basis of his ability to provide a secure environment for her young. But is the male really doing anything more than following an instinctual urge to build something that happens to serve as a nest, and is the female simply making herself available for sex in response to a visual trigger that happens to function as a nest? The reason animals always choose a mate with characteristics that make them more suitable parents may be nothing more than a coincidence, arising from the fact that those whose choice is triggered by characteristics that do not result in the production of viable offspring do not pass that choice on to a new generation. That which is inherently self-perpetuating perpetuates itself. That which is not inherently self-perpetuating does not perpetuate itself. This applies equally to behaviour and to physical form. Thus, for the world to have become full of creatures that perpetuate themselves requires neither design nor purpose, only that logical outcomes accumulate.

The irony of this is that, having removed purpose and design from the evolutionary process, human evolution obliges us to reintroduce it. There *is* an element of intelligent design in human evolution, and it is triggered every time someone chooses a partner on the basis of the offspring they may produce. This makes it possible for us to participate in our own evolution in a way that is denied to creatures that are not consciously aware of the reproductive consequences of their sexual acts. And what enables this is the possession of language.

One of the great imponderables is the extent to which humans are able to *know* anything they cannot put into words, and that is a morass I am happy to skirt around. So, whether our early ancestors were surprised to hear that sex produces babies and babies resemble both parents, or whether it was no more than a statement of something they already sort-of-knew, the fact remains that the possession of language enabled them to put the matter into words and develop strategies for dealing with it that were not available to their non-speaking predecessors. In this way the evolution of

language affected the evolutionary process itself, in that the invention of kinship terms allowed the imposition of a taboo against incest. It has been argued that this taboo not only guards against in-breeding, but also widens the evolutionary possibilities available to the species by allowing a greater range of genetic variation to be passed on from generation to generation. This is probably true, as far as the incest taboo is observed in practice. In the beginning, however, it may have had quite the opposite effect.

As we have seen in the case of the Nicaraguan children, once innate human language ability has a formal language to work on, it can develop that language to the limit of current neurological capacity within a generation, and if this process occurred among a population in which, for lack of a language to expose it, a great variety in linguistic ability existed, then the development of a language that a large part of the group simply could not master may have resulted in a sudden severe restriction of the range of genes that were passed on. And this new invention, by putting into words what had hitherto been no more than a vague sense of unease, and making explicit the connection between sex and procreation, could have caused that variation in linguistic capacity to become self-reinforcing as speakers and non-speakers separated into endogamous groups that no longer inter-bred. It may even be that one of the most significant milestones in the whole of human evolution was the first time a woman said, "I'm not having sex with him. He can't talk!"

Of course, as with all simple models, one must expect the process to have been much more complex in reality. Apart from anything else, there have always been (and sadly still are) numerous ways in which women come to bear children by someone else's choice of father.

One thing that is clear is that, because of language, the normal rules of evolution no longer apply to *homo sapiens* in significant respects. One is that language is a behavioural phenomenon rather than strictly anatomical, the other is that it permits conscious participation in the evolutionary process. A species that has four legs cannot evolve into a species that has six, except via a process that may take millions of years. Similarly a species in which sex is triggered by a nodding of the head can change only slowly into one in which sex is triggered by the swishing of the tail. Humans, however, could replace their spoken languages with languages based on tap-dancing, or playing cards if they wanted to badly enough, and do it instantly. Admittedly this is unlikely, but there is no *anatomical* barrier to such a change in behaviour. It does, though, illustrate that because our linguistic ability enables us to imagine such alternatives, communicate them to others, and agree to push the change through at all costs, it makes possible massive short-cuts in the evolution of human behaviour.

Nevertheless, even though we have a unique ability to decide whose sperm fertilises whose egg, our physical evolution has remained constrained by the structural limitations of sperm-to-egg fertilisation – until now. We await with excitement and trepidation the short cuts that may be offered by genetic engineering.

Acknowledgments

The writer is a member of the Evolution of Human Language Program organised by the Santa Fe Institute, whose support is gratefully acknowledged. I am also indebted to the Rosetta Project and the Max Planck Institute for much-needed assistance. Last but not least, though, thanks to Hal Fleming who has been a guide and an inspiration from the very beginning. Be of good cheer, Hal.

References

Cruttenden, Alan. 1979. *Language in Infancy and Childhood: A Linguistic Introduction to Language Acquisition.* New York.

Crystal, David. 1987. *The Cambridge Encyclopedia of Language*. Cambridge University Press.

Deacon, Terrence. 1997. *The Symbolic Species: The Co-evolution of Language and the Human Brain.* London.

De Villiers, Jill G. & Peter A. de Villiers. 1978. *Language Acquisition*. Harvard University Press.

De Villiers, Peter A. & Jill G. de Villiers. 1979. *Early Language*. Fontana/Open Books.

Gardner, Beatrice T. & Allen Gardner, R. 1975. "Evidence for Sentence Constituents in Early Utterances of Child and Chimpanzee." *Journal of Experimental Psychology: General* 104: 3. 244–267.

Gould, Stephen Jay. 1977. *Ontogeny and Phylogeny*. Cambridge, Mass: Harvard University Press.

Kegl, Judy, Ann Senghas & Marie Coppola. 1999. "Creation through Contact: Sign Language Emergence and Sign Language Change in Nicaragua." *Language Creation and Language Change: Creation, Diachrony, and Development,* ed. by Michael DeGraff, 179–237. Cambridge, Mass.: The M.I.T. Press.

Lenneberg, Eric, H. 1966. "The Natural History of Language." *The Genesis of Language: A Psycholinguistic Approach,* ed. by Frank Smith & George A. Miller, 219–252. Cambridge, Mass.: The M.I.T. Press.

Lieberman, Philip. 1989. "The Origins of some Aspect of Human Language and Cognition." *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans,* ed. by Paul Mellars & Chris Stringer, 391–414.

Morford, Jill P. & Judy A. Kegl. 2000. "Gestural Precursors to Linguistic Constructs; How Input Shapes the Form of Language." *Language and Gesture*, ed. by David McNeill, 358–387. Cambridge University Press.

Premack, David & Ann James Premack. 1983. *The Mind of an Ape*. New York.

Slobin, Dan I. 1966. "The Acquisition of Russian as a Native Language." *The Genesis of Language: A Psycholinguistic Approach*, ed. by Frank Smith & George A. Miller, 129–148. Cambridge, Mass.: The M.I.T. Press.

Zihlman, Adrienne, L. 1989. "Common Ancestors and Uncommon Apes." *Human Origins,* ed. by John R. Durant, 85–105. Clarendon Press, Oxford.

The age of Mama and Papa

Alain Matthey de l'Etang & Pierre J. Bancel
Association for the Study of Language in Prehistory

The global distribution of *mama/papa* kinship terms has been traditionally explained as the result of linguistic convergence, not of common origin. It is usually alleged that these terms are in no way resistant to linguistic change and that they are subject to constant modification, loss and replacement by other nursery-shaped kinship terms. A serious etymological survey shows that kinship nursery terms are, to the contrary, extraordinarily resistant to phonetic and semantic change, as the most ancient written data clearly prove. The cumulative evidence discards the traditional linguistic explanation for the global distribution of *mama/papa* words and advocates for their antiquity within the language families where they are found and, further, for their common descent from a language ancestral to all existing languages.

1. Presentation

A few years ago, the late Larry Trask launched an article on the web entitled "Where Do *Mama/Papa* Words Come from?" the purpose of which was, as Trask himself declared, to put "a final nail in the coffin of the Proto-World conjecture," i.e., the assumption that the worldwide distribution of kinship terms of the model *mama/papa* had to be explained by the existence of a language ancestral to all of the world's languages in which they appear. The topic is certainly not new, and like a vast majority of linguists before him, the author considered the global distribution of such words as the result of linguistic convergence.

Everyone, within the linguistic community, of course knows the arguments that were presented by Jakobson (1960) in his famous article "Why Mama and Papa?" to account for this global convergence. On the one hand, they were founded on the universal phonetic properties of the babbling stage.[1] On the other hand,

1. Jakobson (1960) offered several excellent articulatory, auditory, and acoustic explanations to the greater simplicity of vowel *a* and of stops *p, b, m, t, d, n* (as well as, to a lesser extent, *k, g, ŋ*). These sounds are the very first phonetic segments to be mastered, between 6–9 months of

they explained the particular phonetic shape of such words, most notably the universal association between nasal consonants *m*, *n*, *ŋ,* and the maternal entity like *mama* "mother," by the fact that extant nursery terms prompt the young children to convert the primary nasal babbling associated with feeding and sucking, into parental terms (Jakobson 1960: 130). It is also well known that what prompted Jakobson to write his article was the statistical confirmation made by the ethnologist G.P. Murdock (1959: 1) of "the universal tendency for languages, regardless of their historical relationships, to develop similar words for father and mother on the basis of nursery forms."

On the same page, Murdock developed his own theory about this universal convergence which is worth being fully quoted:

> As standard parental terms become phonetically and morphologically modified in consequence of the normal processes of linguistic change, forms develop which are difficult for very young children to pronounce. Under such circumstances, simpler nursery forms tend to appear – carved, so to speak, out of infant babblings under parental encouragement. From time to time, it is alleged, such nursery forms come to replace the traditional words in standard usage. Since their phonetic range is severely limited by the speech potentialities of young children, similar forms tend to crop up through convergence in historically unrelated languages.

This paragraph has constituted an undisputed postulate for almost each and every historical linguist until today.[2] It was beautiful in its simplicity, very convincing, apparently so well comforted by Jakobson's subsequent linguistic remarks, that everything was for the best in the best of all possible worlds. One thing just needed to be done: to demonstrate it practically, and it is precisely the task that Larry Trask undertook in his article.

---

age, by the child in his initial "babbling stage" of language acquisition. They may reasonably be regarded as the simplest speech sounds (for further phonetic justifications see Bancel & Matthey de l'Etang 2005.) With regard to syllable building, V (e.g., *a*) is of course the simplest of all syllable types, though it may not be considered phonetically articulated. In turn, CV (e.g., *pa*) is the simplest of articulated syllables; then, VCV (e.g., *apa*) is the simplest of all syllable sequences, while CVCV (e.g., *mama*) is the simplest sequence of complex syllables. Finally, reduplicating VCV and CVCV syllables (e.g., *ama or kaka*) are simpler than any other sequences of the same type using different vowels and/or consonants in each syllabic unit (e.g., *uta* or *biku*). It is thus utterly natural that babies begin speaking using sound sequences like *ama*, *papa*, *nana*, or *ata*: in all phonetic respects, these sequences are the easiest ones for them (as well as for everyone, though our adult speaking skills render it imperceptible to us).

2. Ruhlen (1994) openly questioned the convergence theory.

The idea that Trask repeats ad nauseam in the fifth and most interesting part of his essay, entitled "They just keep coming", is that linguistic change affects nursery-shaped kinship terms, just the way it does for any other word:

> In other words, the *mama/papa* words should be slowly but steadily disappearing from the world's languages. Are they? Well, [Trask goes on,] these words are assuredly not magical, and they are subject to the same linguistic processes as other words.

As a matter of fact, Trask thought that the history of languages abundantly and easily proves this assertion.

He uses as a model for all languages, what he believes happened to the Indo-European kinship terms **patēr* "father" and **matēr* "mother" since the PIE period. Trask states that **patēr* and **matēr* were originally *mama/papa* words which had acquired a kinship suffix *-ter*, and that "these original words for mother and father, where they have survived at all, have undergone the usual changes in the languages possessing them" to the point that some of them have become hardly recognizable. And he concludes that "the *mama/papa* words are in no way resistant to the process of linguistic change including regular changes in pronunciation. Nor are they resistant to loss" [our emphasis, AME & PJB].

So there is only one way, the author thinks, that languages can remedy this erosion or loss, and this is by constantly creating new *mama/papa* words. Trask believed he could illustrate this phenomenon by using two sets of examples: the first one is about the languages that have partially or totally lost their "inherited" kinship terms and have replaced them, as he says, by "newer" *mama/papa* words. The second set concerns languages that do still possess their "inherited" terms but have already adopted "newer" nursery-shaped terms. After browsing examples taken from various Indo-European languages, but also from the Turkic and Dravidian families, Trask thought the question was definitely settled:

> The conclusion is inescapable. The *mama/papa* words are not fossilized relics of some ancient ancestral language at all. Instead, they are being created all the time. New examples of *mama/papa* words are constantly being invented and passing into use. At first, these new words survive alongside the older ones as informal or intimate versions, but eventually they may take over completely and drive the older words out of the language. This process is self-renewing for ever. […] This endless re-creation and recycling of *mama/papa* words explains a great deal. It explains why we find these words so often, in so many languages.

The whole of Trask's demonstration is certainly of the utmost relevance. In effect, most of the examples that he chose belong to languages or language families where written documentation goes back into a deep past, allowing us to verify and actually refute his allegations. We will examine these examples following the exact order that Trask himself chose to follow, starting from the languages where the "inherited" terms

were lost and allegedly replaced by "newer" nursery-shaped kinship terms and continuing with languages where "inherited" formal kinship terms still exist but are, in one way or another, threatened by "newer" kinship terms. We will observe that none of Trask's allegations resists a serious historical analysis, and that there is no way that the vast majority of the examples of *mama/papa* words that he presented can be considered as newer than the inherited terms. We will then consider examples of historically well documented nursery-shaped kinship terms belonging to families that Trask did not envision, which also clearly demonstrate their extraordinary stability over considerable periods. We will then question the validity of kinship terms' linguistic reconstructions and finally propose our interpretation for the permanence of the *mama/papa* kin terms.

2. The examples of Larry Trask

2.1 Cases of replacement

2.1.1 *Welsh* tad *and* mam

Trask asserts that, in Welsh, the term inherited from Proto-Indo-European *pater* "father" (attested in other Celtic languages such as Gaulish *atér* "father" or Old Irish *athair* "father") disappeared and was replaced by a "new word" *tad*.

In the first place, the Welsh term *tad* "father" is anything but new in the Brythonic branch of Celtic. It is attested in XIIth–XIVth century Middle Welsh *tad* "father" (Charles-Edwards 1993: 169), in XIIth–XVIIth century Middle Breton *tat* "father" (Izard 1965: 93) and also in Old Cornish *tat* "father" (*Vocabularium Cornicum ca.* 1150). According to Charles-Edwards (1993: 169), this word must "go back at least to the Romano-British period."

In fact, *tad* must have belonged to the common Brythonic lexicon, and even to the Insular Celtic lexicon (comprising the Brythonic and Goidelic branches). The first reason is that the Old Irish (a Goidelic language) word *dait* ~ *data* "foster father" (Charles-Edwards 1993: 169; Izard 1965: 93) is evidently related to Brythonic *tad* ~ *tat*. The second reason is that neither Middle Briton nor Old Irish seem to have borrowed the word *dad* from English, a language with which they never were in close contact. On the contrary, the English word *dad,* an isolated form within the Germanic group, only attested in written English from the XVth century on, is likely to have been borrowed from Brythonic at the time of the Anglo-Saxon invasion or later.

Secondly, the Old Welsh derivative *edrydd* ~ *edryf* "paternal kin" is, according to Charles-Edwards (1993: 169), parallel to Old Irish *aithre*, and derives from Proto-Celtic *atrio* "paternal kin," itself cognate with Latin *patrius* "paternal kin." This form clearly shows that the Indo-European root *pater* "father" did not completely disappear from Welsh, either.

2.1.2 *Romanian mámă and tátă*

Trask alleges that in Romanian the Indo-European inherited term for "father" was lost and replaced by a "newer" term *tátă*. This allegation is certainly questionable. Indeed, one certainly cannot exclude the inheritance of *tátă* from the Latin address term *tata* "dad." The Romanian online dictionary (DEX) derives *tátă* (gen. *tati*) "father" from Latin *tátă*. (gen. *tatae*) "father". This hypothesis is fairly plausible, but one cannot exclude, either, that *tátă* was borrowed from the neighboring Slavic languages.

What makes us take the first hypothesis seriously is the fact that, although it is not well known, numerous Romance languages or dialects preserve this Latin term with its original meaning. From Switzerland around Lucerne to southern Italy, including regions like the Abbruzze, Trentino, Campania, Latium, Sardinia, Calabria and cities like Naples, numerous Italian dialects do refer to the father as *tata*. All these terms basically remain address terms as was the case in Latin. What happened to Spanish is also worth being mentioned. In 1492 the Spanish humanist Aelio Antonio de Nebrija (1492) published a Latin-Spanish dictionary entitled *Lexico o diccionario latino-español*, in which the Latin *tata* is translated by *taita* "padre de los niños"; more recently, the *Diccionario Crítico Etimologíco de la Lengua Castellana* (Corominas: 1954) reports the use of *tata* or *taita* in Castillan until the XVIIth century. Modern Castillan does not use *taita* anymore, but the form *tata* "father, child" is still commonly encountered in some Spanish-speaking Latin-American countries.[3] Finally, Catalan has also preserved *tata* until today in some local dialects. The *Diccionari català-valencià-balear*, a comprehensive dictionary based on various Catalan dialects, mentions *tata* "father" (*"pare" en llanguatge infantil o "pare del infants"*) and gives as a primary source the *Lexicon seu Dictionarium Aelii Antonii Nebrissensis*, a Catalan-Castillan-Latin adaptation of Nebrija's *Lexico*, published in 1585 in Barcelona.

Trask also curiously argues for a recent origin of the word *mámă* "mother," which, he says, replaced the "inherited" Latin word *mater*. There is no doubt, however, that *mámă* derives from the Classical Latin *mamma* and cannot be a Romanian innovation.

So Romanian lost its "inherited" reference terms for mother and father, and used instead the "inherited" address terms *mama* and the "inherited" or Slavic address term *tata*, to fullfill both the address and reference functions.[4] The reasons for which the inherited *pater* and *mater* disappeared from Romanian, whereas they are still in use in other Romance languages, constitute a subject for debate.

---

3. It is sometimes argued that this Latin-American form was borrowed from Amerind languages where *tata* is quite frequent.

4. The *Dictionnaire roumain-français* (1992) explicitly mentions this last usage for both terms.

2.1.3 *The terms* ata *and* aba ~ baba *in Turkic languages*

According to Larry Trask,

> … the inherited Turkic words for "mother" and "father" are *ana* and *ata*, respectively, and these words […] are still the everyday words in most Turkic languages. But, in the best-known Turkic language, [Modern] Turkish, the word *ata* has now become specialized. It is no longer the everyday word for "father," and instead it is an elevated word meaning "forefather, ancestor." […] But the everyday word for "father" is now *baba*. This, of course, is another *mama/papa* word, and it used to be the Turkish word for "daddy," but now it is the ordinary word for "father," and "daddy" must now be expressed by adding a diminutive suffix, producing *babacik* […].

This apparent succession of replacements and meaning shifts seems to illustrate Trask's thesis that nursery words undergo constant and quick replacements through time. But the illustration fails because his presentation is partial – in all the senses of the word.

Let us examine the *ata* forms in the first place. Proto-Turkic *ata* most certainly referred to the father, as it is reflected in many languages of this family: Old Turkic *ata* "father," Karakhanid *ata* "father," Middle Turkic *ata* "father," Azeri *ata* "father," Sary-Yughur *ata* "father," Tatar *ata* "father," Kazakh *ata* "father," Nogai *ata* "father," Balkar *ata* "father," Kumyk *ata* "father," Bashkir *ata* "father," Khakassian *ada* "father," and Tuva-Tofalar *a'da* "father" (EHL). All these terms are not only traces of the Proto-Turkic form; they also show that this word underwent no phonetic and semantic change at all in most of the Turkic languages through their entire history. The most ancient attested examples of *ata* are found in Uighur (EHL), and in Middle Turkic (Krader 1963: 391–392) with the meaning of "father," some 1,300 and 1,000 years ago respectively. Isolating the case of Modern Turkish from other Turkic languages is exactly contrary to the comparative method.

Now, is the "specialization" of meaning of *ata* in Modern Turkish from "father" to "forefather, ancestor" a true change of meaning? In fact, a good number of modern Turkic languages use *ata* either as "ancestor" or "father" and "ancestor:" Modern Turkish *ata* "ancestor," Turkmen *ata* "paternal grandfather," Kirghiz *ata* "father, ancestor," Karakalpak *ata* "ancestor," Uighur *ata* "father, ancestor," Altai *ada* "father, ancestor." Consequently, it is highly likely that *ata* originally meant both "father" and "ancestor." The specialization that occurred in Modern Turkish (as well as in Karakalpak) is certainly not a semantic innovation at all.

As mentioned above, Larry Trask believes that *baba* "father" is an innovation in Turkish. Is it the case? In fact, one also finds *baaba* "grandfather" in Turkmen, a language belonging to the same southern branch of Turkic as Turkish, but also *baba* "grandfather", "elder" in Uzbek (Schurmann 1962: 200), a language belonging to the eastern branch of Turkic. The first possibility is that *baba* forms are derived

by reduplication from the Turkic word *apa ~ aba* "father, ancestor," attested by Old Turkic (Orkhon) *aba* "grandfather," Karakhanid *aba* "father, ancestor, bear," Turkish *aba* "father," Azeri (dial.) *aba* "father," Turkmen (dial.) *aba* "father," Salar *aba* "father," Sary-Yughur *awa* "father," Tatar (dial.) *aba* "father," Kirghiz *aba* "father," Balkar *appa ~ aba* "father," Bashkir (dial.) *apa* "father," Khakassian *aba* "father," Tuva-Tofalar *awa* "father," Altai *aba* "father, bear," Chuvash *oba* "bear" (EHL).[5] Moreover, Turkic is one of the three branches of the Altaic family, together with Mongolic and Tungusic, where the root *apa ~ aba* "father, ancestor" is abundantly attested.

The second hypothesis is that the Turkish, Turkmen, and Uzbek *baba* forms were borrowed from the neighboring Iranian languages (such as Farsi, Pashto or Tajik). Iranian languages are derived from Indo-European, where the root *papa ~ baba* "Father" is also present in the Anatolian, Indic, Nuristani, Italic, and Greek branches. The fact that Modern Turkish also uses another term *peder* "father," from Iranian origin, gives substance to this hypothesis. Of course, these two hypotheses are not contradictory. If, as it is probable, Turkish-speaking people when they reached Anatolia had the use of an *apa ~ aba* form in their language, it would have been only easier for them to adopt a *baba* form from Indo-Iranian speakers in the newly conquered regions.

Whether borrowed from Iranian or derived from Proto-Altaic (or both), Modern Turkish *baba* "father" is certainly not a newly created word. The highly significant point here is that an overwhelming majority of Turkic languages preserved the two Proto-Turkic forms *aba* and *ata* without modification.

2.2 Cases of coexistence between "inherited" terms and "newer" terms

2.2.1 *The Greek terms* babas/pateras *and* mama/miteras

In order to prove the recent origin of *babas*[6] "daddy" in modern Greek, Trask engaged in a short linguistic demonstration. *Babas,* he says,

> cannot be ancient in Greek [contrary to *pateras* "father"], because the consonant /b/ of Classical Greek changed in every case into /v/ in the postclassical period. For example, Classical Greek had the word *biblios* "book" […]. But the Modern Greek form of the word is *vivlio*, with the earlier /b/s changed into /v/s.

---

5. Forms meaning "bear" result from a linguistic taboo on the true name of the bear, a frequent phenomenon with dangerous animals. It is also attested in Slavic for the bear (e.g., Russian *med'ved'* "bear," literally "honeyeater"), in Eastern (savannah) Bantu languages for lion, and in Northwestern (rainforest) Bantu languages for gorilla. A hunter may not tell he goes hunting bear (nor lion or gorilla), lest he puts himself at risk of being killed or injured. Rather, he says "I go out to see Grandad."

6. Trask's *babbas* seems to result from a confusion of two different Modern Greek words: *pappas* "priest, pope" and *babas* "dad". We restore the correct transcription.

This linguistic argument is irrelevant in the particular case of the modern term *babas* "dad" (vocative *baba*). This term is certainly not descended from Ancient Greek through regular phonetic evolution, and no one has ever suggested it does. To the contrary, *babas* is known to have been borrowed from Osmanli Turkish (Chantraine 1968) during the Ottoman domination over Greece. The ancient Greek form was the Homeric *pappas* "dad" (vocative *pappa*) (Liddell & Scott: *A Greek-English Lexicon*). This term still survives in Modern Greek to refer to the priests of the Orthodox Church and, most interestingly, this word in its vocative form *páppa* "dad" is still in use today in the Pontic dialect of Greek (Fauvin and Nikaki pers. com.).

Finally, Trask also claims that the modern Greek term *mama* "mum" is newer than the inherited term *miteras* "mother." Once again he forgets that *mama* has counterparts in Classical Greek such as *mammé* or *amma* (Liddell & Scott). No additional comment is needed.

2.2.2 *French* papa/pɛʁ *and* mamã/mɛʁ

Among the terms that Trask considers "newer" than those derived from PIE *pater* and *mater* are the French *papa* "dad" opposed to the formal "inherited" term *pɛrɛ* "father," and *mamã* opposed to the formal *mɛʁ* "mother".

The case of the address term *papa* "dad" is intriguing. This word is certainly not a new word in French. It is attested in XIIIth century Old French (*Dictionnaire historique de la langue française*). But its derivation from Classical Latin is certainly problematic: *pappa* is attested in Classical Latin as an onomatopeia used by infants to *"call for food"* (Lewis and Short) but one also finds *papas* in the nominative form in Juvenal's *Satires*, generally translated as "tutor" or "governor." Then, since the end of the IInd century AD, *papa* has been used by ecclesiastical writers such as Tertullian or Prudentius (Lewis and Short) with the meaning of "bishop." During the Middle Ages, *papa* has been specifically used to refer to the pope (Rome's bishop). Finally, no data at hand provides clear evidence for the use of *papa* in the sense of "dad" among Latin speakers. This is why some authors suggest that *papa* was borrowed from the Greek *pappas* "dad," a word that we just mentioned. This hypothesis is credible, and the fact that the liturgy of the Latin Church was in Greek during this period might explain why it has been adopted by the Latin Fathers of the Church.

Furthermore, the French address term *mamã* "mum" derives from the Latin term *māmma* like its Italian, Romanian, Spanish and Portuguese *mamma* counterparts do. Along with *papa*, *mamã* is attested in Old French from the XIIIth century.

2.2.3 *Italian* babbo/padre *and* mamma/madre

Are also thought newer than the "inherited terms" *madre* and *padre*, the Italian hypocoristic *mamma* "mum" and *babbo* "dad." The claim that the Italian address term *mamma* "mum" is newer than *madre* is properly extravagant. Trask just writes as if

Latin never existed. Strictly speaking, *madre* should be called "newer" than *mamma*, as *mamma* matches exactly its Latin ancestor *māmma*, while *madre* shows significant phonetic transformations in relation with its own ancestor *mater*.

All over Italy, *babbo* is used to address the father, but we also find for the same usage *papà*, like in French and, as we mentioned above, in some regions, *tata*. Even if its ultimate origin is not well determined, *babbo* is a pretty old word, as several occurences of it have been recorded in various Tuscan and Corsican texts, dated from the XIIIth and XIVth centuries (TLIO) and notably in Dante's *Divine Comedy* (Inferno XXXII, 9). A phonetically very close word meaning "father" is Sardinian *babbu*. TLIO claims that *babbo* comes from childish Latin **babbus*. To our knowledge, no Latin dictionary mentions such word.

2.2.4 *Bengali and Hindi* baba *and* pita

Trask chooses another example from Bengali and Hindi:

> In Bengali the formal word for "father" is the inherited *pita* but the informal one is *baba*…. In the closely related Hindi, the inherited *pita* is now strictly an honorific term, and the ordinary word for "father" is *baba* or *bap*.

Trask believed he captured live the process of replacement of a formal (ancient) term by an informal newly coined nursery word. Unfortunately, faithful as he is to his method, he just extrapolates from only one or two languages within a linguistic family. If he only had compared the languages of the Indic family, consulting Turner's *Comparative Dictionary of Indo-Aryan Languages* (1962–1966) or Karve's *Kinship Organization in India* (1953) he would have noted, on the one hand, that 48 out of the 56 languages or dialects belonging to the Indic branch of Indo-Iranian for which there is accurate data, display *bappa* or *bapa* words. A vast majority of these terms do refer to the father, a few of them however refer to the father's brother, the brother or even the grandfather. The vast distribution of this *bapa* form in modern languages proves beyond discussion that it can by no means be a term that has recently appeared in each and every Indic language, by some convergence process. It is a common form inherited from the ancestral language of the Indic branch and beyond from Proto-Indo-Iranian.[7] Lewis Morgan (1871: 405) reported its existence in Sanskrit, but we haven't been able to find this word in the dictionaries of Sanskrit that we have accessed. Turner (1962–1966: 520, 9209) mentions *bappa* in Prakrit with the meaning "father." Regular use of Prakrits is attested from the IIIrd century BC on.

7. The term *baba* "father," "grandfather," "brother" is also widespread in the Nuristani and Iranian branches of Indo-Iranian. For Nuristani see Strand (1997–2008).

On the other hand, he would have noticed that there are apparently only 3 descendants (Bengali, Hindi, Pahari: *pitā*) from the Sanskrit singular nominative *pitā* which originally meant "father, father's brother" or even "ancestors in the male line," when it was employed in the plural (Karve 1953: 38). The present meaning of these reflexes is almost exclusively restricted to "father." Finally, we found 14 languages with reflexes of Sanskrit *pitruya* and of Prakrit *pittiya*, two terms specifically referring to the father's brother (Karve 1953: 38; Turner: 8188).

But this survey would not be complete without mentioning the existence, in ancient Indo-Aryan languages, more than 2,000 years ago, of a third widely attested term, also referring to the "father," and it is of course *tata*. In Sanskrit, besides *pitā*, we very often find *tata* which, when used in reference, always stands for "father" and, when used in address, "may stand for any male relative or acquaintance" (Karve 1953: 38). In Prakrit, we have *piu(a)* "father" and *pittiya* "father's brother" that we just mentioned (Turner: 8188),[8] *tā* "father" or "son" (Turner: 5754), and finally *bappa* "father" (Turner: 9209). In Pali (*Pali-English Dictionary*: 458), the nominative form *pita* stands for the father, and *tata* is a vocative form to address the father, an elder, a younger, or a superior (*Pali-English Dictionary*: 299). In Ardhamagadi (Karve 1953: 88), both *piya* and *tāya* apparently refer to the father. So much for the ancient languages; if we turn to modern languages, we will notice that *tata* survives in 27 out of the 56 languages that we mentioned. In most Indic languages, it is found under a modified phonetic shape *dada*, and stands actually for the paternal grandfather (Turner: 6261). The *tata* form only survives in Romany *tatta* "father" and Kowar *tat* "father" (Turner: 5754).[9]

None of these kinship terms, as we can see, has disappeared from the modern Indic languages, and none can be argued to be older or newer than the others. Our comparative survey renders a slightly different conclusion as Trask's. Indisputably, simpler nursery-shaped terms – and *baba* even better than *tata* – have perfectly resisted linguistic change during more than two millennia, while "inherited" forms have resisted much less, to the point that, save very few exceptions, *pitā* "father" almost vanished from the Indic family.

The various senses of each of these terms in modern Indic languages have to be explained by the fact that languages (or even entire branches of Indic), taken individually, have only retained portions of the classificatory relationships that these terms originally referred to. As a final remark, let us point out that it would be risky to allege the anteriority of Sanskrit *tata* and *pitā* vs Prakrit *baba*. As we already mentioned, the

---

8. Prakrit *piu*(*a*) "father" has numerous counterparts in modern Indic languages.

9. In the Nuristani branch (Turner: 5754; Strand 1997–2004), it is the *tata* form which prevails, with its original meaning "father" or "father's brother" (Waigali *tata* "father," Vä *tâta* "father," Amês *tâta* "father, father's brother").

wide distribution of *baba* through the Indic group certainly testifies to its great antiquity. Moreover, Sanskrit, as a literary language, probably did not use this address term, even though it was in common use in the vernacular languages of this period. These three terms *baba*, *pitā* and *tata* belong to the common vocabulary of Indo-European, as it is shown in Matthey and Bancel (2005).

2.2.5 *Tamil* appa

The last of Trask's examples that we will consider is the Tamil "formal word" *takkappan* "father," opposed to the "informal" term *appaa* "father." According to Trask, the informal *appaa* is just another case of innovation.

A first remark is that Tamil as a whole has no less than 6 words referring to the father: *ayya*, *appan*, *appaci*, *appu*, *takappan*, *attan*, *ammān* (Karve 1953: 196). A second remark is that *takappan* is in no way "older" than *appan*, as it is obviously a compound word formed with *appan* "father" and *tak*, an adjective form of the verb *taku* "to be excellent" (Emeneau 1953: 342, 10). So *takappan* literally means "excellent father."

Now, what do we know about the origin of *appan* ? This question is partially answered using data from the Old Tamil literature and inscriptions. Tamil is one of the rare classical languages possessing an unbroken literary tradition. The first written transcriptions of Tamil go back to the IInd century BC and its rich literary tradition makes an etymological survey possible.

One of the earliest Tamil inscriptions (IInd century BC), found in the Mangulam cave (district of Madurai) and published by I. Mahadevan (2003: 317), displays the third person singular possessive form *tantai* to refer to the father. This form is also found in Old Tamil literature along with *antai* "my father", *nuntai*/*untai* "your father" (Emeneau 1953: 340). Apparently none of these forms survives today in modern Tamil, but similar terms are used in other Dravidian languages (Emeneau 1953: 340 and 350).

Mahadevan (2003: 609) observes that *appan* does not occur in the early Cankam literature, and it is not found either on the stone inscriptions of the earliest period (IInd century BC). Its first recorded occurrence is on an inscription from the IIIrd century AD, found in Edakal cave (Northern Kerala) in which it does not specifically stand for the father, but is used as a masculine honorific suffix. Quoting Burrow and Emeneau (1984: 15, 156a), Mahadevan (2003: 609) connects this particularity to Old Kannada onomastics and the frequent usage in Kannada to add *appa* to the proper names of men, as a mark of respect,[10] and he believes that the occurrence of *appan* in Edakal's inscription "reveals Kannada influence in the region." One of the earliest occurences of *appan* clearly referring to the father is found in the Tevaram hymns,

10. Burrow and Emeneau (1984: 16) mention the same usage of the affix *appa* in Tulu and Telugu.

composed at some point between the VIth and the VIIIth century AD (Mahadevan: 609). Even considering these rather recent dates, Mahadevan has no doubt about the antiquity of this form. His opinion is certainly supported by a comparison between the sister languages of the Dravidian family.

On the one hand, six languages from the southern branch of Dravidian do display *appa* words to refer to the "father": Tamil, Kodagu, Malayalam, Kannada, Tulu, Telugu. Three languages of the central branch (Hill Maria, Gondi, and Konda) display phonetically comparable terms which also refer to the father: *tappe*, *āpōral* and *aposi* respectively.[11] On the other hand, a few Dravidian languages use words like *aba* for the same usage. Besides *appan* or even *appā*, Tamil uses *abba* to refer to the father, a language of the central branch, Kui, uses *aba* for "father, grandfather, ancestor," and three languages from the northern branch (Kurukh, Malto, Brahui) respectively display *abbā*, *abba*, *aba* to refer to the father and also (Brahui) the grandfather.[12] The *Tamil Lexicon* makes a connexion between the first set of *appa* forms and Prakrit *appa* "father." Emeneau, for his part, (1953: 351) suggests a link between the second set of terms and the "Indo-Aryan material" and he mentions Baluchi *abba* "father," Sindhi *abo* "father."

The vast distribution of *appa*/*abba* forms throughout the Dravidian family makes very unlikely that *appan* emerged at some point in Tamil and progressively replaced the "inherited" *tantai*. The fact that we do not find it in the earliest literary data does not mean it did not exist as an address form in the vernacular usage at that time. We just similarly observed that, while it is apparently absent from Sanskrit (but present in the more recent Prakrit), *baba* is the most common word to refer to the father in the Indic family. Our opinion is that the near phonetic identity of all these Dravidian words indicates that Tamil *appan* is original to this language, and that EHL's proposed reconstruction **ap-* "father" for Southern Dravidian is correct.

3. Conclusion

Up to this point, the vast majority of Trask's alleged "newer" words have not provided support for his replacement thesis. All of them have been merely abstracted from

---

11. The use that each of these languages makes of the *appa*/*aba* terms certainly varies, but we have to confess that our information on this topic is rather scanty. Only a few authors mention other meanings than "father." Beck (1972: 289) clearly notes that Tamil *appan* "father" is a term of reference and so does Bossé (1983: 97). Grigson (1949: 308) reports the same for Hill Maria Gondi *tappe* "father." Emeneau (1953: 184) points out that Brahui uses *abbā* as a "respectful term of address" in addition to "father" and "grandfather," and finally Srinivas (1942: 204–205) indicates that Kannada uses *appa* as a vocative (address) term, while it uses *thande* and *thayi* as denotative (reference) terms.

12. This information has been collected from the *Tamil Lexicon*, Emeneau (1953), Trautmann (1981), Burrow & Emeneau (1984) and EHL.

their comparative and historical context. Once collated with the historical documentation, these words are easily shown to be the direct descendants – both phonetically and semantically – of forms that were used centuries and even millennia ago. If we compare them to data from closely related languages, they immediately appear as obvious cognates of words from sister languages, that must have been used through millennia with the same meaning.

So much for Trask's "innovations," then. But before we put a full stop to the story, let us now examine other examples of nursery-shaped kinship terms taken from families not envisioned by Trask and for which there are substantial written documents covering different historical periods.

4. Other ancient *mama/papa* words

4.1 Korean *appa* (c. 600 years)

In his article about Korean kinship terminology, Kwang-Kyu Lee (1973) reports several terms referring to the father, two formal forms *abŏji* or *abŏ-nim* and an address form *appa*, used by children. On its part, EHL mentions another more "vulgar" form *abi*. All these forms have a Middle Korean (ca. 600 years ago) precedent, *àpí* (EHL), that also referred to the father. The difference between these terms is not significant, as we can see.

4.2 Mongolic *aba* (c. 800 years)

The term *aba* is attested in early written Mongolian documents, some 800 years ago (EHL) with the meaning "father." We find numerous exact replicas of this term in modern Mongolian languages: Monguor *aaba* "father," Kalmuk *aaba* "father," Khalkha *aab* "father," Dongxian *aba* "father," Shary-Yoghur *aba* "father."

4.3 Tibetan *pa* (c. 1,300 years)

In most Tibeto-Burman languages, *pa* refers to the father (Limbu *pā*, Dumi *pa*, Kulung *pa*, Yamphu *pa*, Tibetan *pha*, etc.). Benedict (1941a: 438; 1972: 19) reconstructed the Proto-Tibeto-Burman form **p'a* "father." Moreover, direct evidence of this word is attested in Old Tibetan. One can find *pha* notably in the Zhol Inscriptions (Tibet, after 763 AD) and in the Lho-Brag inscriptions (Tibet), the date of which is unknown (Fang Kuei Li & Coblin 1987).[13]

---

13. Old Tibetan *yab* "father" is even more frequently attested than *pha* "father" (Fang Kuei Li & Coblin 1987: 143–144; Richardson 1985: 170, 175–176; Bacot, Thomas & Toussaint 1940: 39–40) and is still used today as a term of respect (Benedict 1941b: 329). The respectful term *yum* "mother" is also attested in Old Tibetan (Richardson 1985: 176; Bacot, Thomas & Toussaint 1940: 39–40) and in Modern Tibetan (Benedict 1941b: 329).

4.4 Chinese *baba* and *mama* (c. 3,300 years)

Mandarin and Cantonese address terms for father and mother are *papa* and *mama*. They are both represented by pictophonetic characters which combine a character indicating the meaning and another one indicating the sound. In order to represent the father, we have the paternal reference character *fu* combined with the sign symbolizing the syllable *ba*; for the mother we find a combination of a character referring to the mother and another referring to the horse, pronounced *ma*. These two pictophonetic characters with similar prononciations are attested since the Tang period (VIIth–IXth century AD).

But there is a more ancient and simpler way to symbolize father and mother in Chinese writing. Archaic inscriptions dated from 1,300 to 1,100 years BC (bronzes from the Shang and early Zhou periods) show that the character standing for the father was basically the same as the one that is in use today to express the referential part (*fu*) of the pictophonetic character that we just mentioned. Strong arguments have been put forward, pointing out that at this time (1,300 BC) this character was pronounced *ba* (EHL). Other archaic inscriptions from the same period display another character referring to the mother. This character, which according to EHL must have been pronounced *məʔ* at the time, is still in use today under a developed form, but is pronounced *mu* in most Chinese dialects. The written forms of these two last characters can be traced back to the Jiaguwen pictographic period, dated ca. 1,500 years BC.

We can thus state that the address terms used in modern Chinese dialects for "father" and "mother" are just the same as those that were used in Archaic Chinese at least 3,300 years ago.

4.5 Semitic *ʔab-* and *ʔumm-*(c. 4,500 years)

One of the earliest recorded kinship terms in the world's languages is the enclitic possessive form *abi* "my father." What is absolutely striking is that this form has been used in most Semitic languages up until Modern Standard Arabic: Old Akkadian (2,600 BC), Old Babylonian (1,800 BC), Middle Babylonian (1,400 BC), Old Aramaic (850 BC), Moabite (800 BC), Phoenician (800 BC), Classical Arabic (600 AD). Beside this perfectly preserved possessive form, old Semitic documents have also recorded the bound form *abu*, known to most of our contemporaries, because it is heard almost everyday on TV, and litterally meaning "father of." It is encountered first in Old Akkadian (2,100 BC), in the well known Old Babylonian Hammurabi's Codex (1,800 BC), and also in some Classical Arabic early stone inscriptions from the middle of the VIIth century AD.

This is not all, here is the impressive list of cognates collected from EHL and other sources for the form *umm* "mother (of)," including Akkadian *ummum* "mother," Eblaic *ʔummum* "mother," Amorite *ʔummum* "mother," Ugaritic *um* "mother," Hebrew *ʔēm*

"mother, father's wife, grandmother, ancestress," Judaic Aramaic *ʔimma* "mother," Epigraphic South Arabian *ʔmm* "mother," Geez *ʔəmm* "mother," VIIth century AD Arabic *ʔumm* "mother," Modern Arabic *ʔumm* "mother," Tigre *ʔəm* "mother," etc.

These Semitic examples certainly constitute the best documented evidence of the resistance of nursery-shaped kinship terms to linguistic change[14]. Table 1 presents in chronological order numerous cognates of **ʔab*-, taken from various languages within the Semitic family and showing this remarkable stability.

Table 1. *ʔab*- forms in ancient and modern Semitic languages

Ancient Semitic languages	**Dates**
Old Akkadian *abí* "my father" (Puabi tomb)	c. 4,500 BP
Old Akkadian *abí* "my father" (Sargonic, Gelb 1961: 136)	c. 4,300 BP
Old Akkadian *abí* "my father," Old Akkadian *a-bu* "father" (Ur III, Gelb 1961: 129 & 127)	c. 4,100 BP
Old Babylonian *abi* "my father" (legal documents), *abu* "father of" (Codex Hammurabi),	c. 3,800 BP
Ugaritic *ʔab* "father (of)" (The Epic of Kret tablet)	c. 3,400 BP
Middle Babylonian *abi* "my father" (Amarna tablet 17)	c. 3,400 BP
Old Aramaic *ʔbi* "my father" (Tell Dan stela)	c. 2,850 BP
Moabite *ʔbi* "my father" (Mesha stela)	c. 2,800 BP
Phoenician *ʔbi* "my father" (Kilamuwa inscription)	c. 2,800 BP
Pre-Exilic Hebrew *ʔāb*- "father" (EHL)	c. 2,700 BP
Epigraphic South Arabian (Sabean) *ʔb* "father" (EHL)	c. 2,400 BP
Hebrew *ʔb*- "father" (Qumran texts)	c. 2,100 BP
Geʾez *ʔab* "father" (EHL)	c. 2,000 BP
Mandaic Aramaic *ab* ~ *aba* "father" (EHL)	c. 1,700 BP
Judaic Aramaic *ʔabbā* "father" (EHL)	c. 1,600 BP
Arabic *ʔab* "father" (EHL), *abū* "father of" (Medina inscriptions c. 625 AD)	c. 1,400 BP

Modern Semitic Languages

Standard Arabic *ʔabu* "father (of)', *ʔabi* "my father"	Harsusi (South Arabian) *χayb* "father" (EHL)
Amharic *abbat* "father" (EHL)	Tigrinya *ʔabbo* "father" (EHL)
Soqotri (South Arabian) *ʔab* "father" (EHL)	Harari *āw* "father" (EHL)
Tigre *ʔab* "father" (EHL)	Mehri (South Arabian) *χáyb* "father" (EHL)
Gurage *ab* ~ *a<u>b</u>* ~ *a<u>b</u>i* ~ *aw* "father" (EHL)	

14. Beside *ʔab*, a number of Semitic languages also use the vocative form *baba*. Our information about *baba* is rather scanty but we know it is not a new word in Semitic, since EHL mentions its existence in Mandaic Aramaic (c. 300 AD).

5. The linguistic reconstructions

Most linguists feel very uncomfortable presenting reconstructions of nursery-shaped kinship terms. Most of the time, when they publish them, they also mention in a footnote, as a kind of excuse for their boldness, that these terms are "universal nursery words"! This feeling of general discomfort is notably perceptible in Guthrie's Bantu kinship terms reconstuctions. While he admitted **tàátá* "my father," he hesitated to acknowledge **bàábá* "father" because it originated in baby talk (Guthrie 1970, 4: CS 1686 & 3: CS 7). Meeussen (1969), for his part, ratified both reconstructions. Despite all these inconsistencies, numerous reconstructions can be found in the linguistic literature.

For Afroasiatic we have: Proto-Semitic **ʔab* "father" (EHL), Proto-Southern Cushitic **aba* "father" (Ehret 1980: 281), Proto-Eastern Cushitic **a(a)bb-* ~ **baabb-* "father" (Blažek 2002: 110), **ʔab* "father" (EHL), Proto-Central Cushitic **ʔab* "father" (EHL), Proto-Berber **V-bb-* "father" (Blažek 2002: 110), Proto-Western Chadic **ʔub-* ~ *ʔab-* "father" (EHL), Proto-Central Chadic **ʔab* "father" (EHL), Proto-Eastern Chadic **ʔab* "man" (EHL), without forgetting Ongotan *ʔabba* "father" (Fleming, pers. com.). All these roots are reflexes of Proto-Afroasiatic **ʔab-* "father" (EHL).

Let us also mention that Dolgopolsky (1998, quoted in Hage 2003: 319) proposed **ʔaba* or **apa* "father" for Proto-Nostratic. The Nostratic megaphylum, as defined by Dolgopolsky, comprises Indo-European, Afroasiatic, Kartvelian, Uralic, Altaic and Dravidian. Greenberg (2002: n° 142) proposed **apa* as one of the words referring to the father in Proto-Eurasiatic. Greenberg's Eurasiatic comprises Indo-European, Etruscan, Altaic, Eskimo-Aleut, Uralic-Yukaghir, Gilyak, and Chukchi-Kamchadal. Numerous reconstructions have been otherwise proposed in other linguistic families: Proto-North Caucasian **aba* "father" (EHL), Proto-Bantu **baba* "father" (Meeussen 1969), Proto-Tibeto-Burman **p'a* "father" (Benedict 1941: 438), etc. Reconstructed (*t*)*ata* forms have been also proposed: Proto-Austric**(*ta*)*ta* "grandfather" (Hayes 2002), Proto-Siouan *atí* ~ *tatí* "father, paternal uncle" (Matthews 1959: 255), Proto-Bantu **tàátá* "my father" (Guthrie 1970, 4: CS 1686), Proto-Athapascan **-ta?* "father" (Hoijer 1956: 325), etc. All these reconstructions seem perfectly valid.

6. Conclusion: The age of *Mama* and *Papa*

The idea that *mama/papa* words "are in no way resistant to the process of linguistic change, including regular changes in pronunciation, [n]or are they resistant to loss," just like ordinary words, does not correspond to any truth. Quite the contrary, data from old written languages every time constitute direct evidence and proof that kin

nursery terms have perfectly resisted linguistic change phonetically and semantically, while the so-called formal "inherited" terms haven't. The *mama/papa* words have been passed (taught) from one generation to the other across immemorial ages, without much noticeable change, and are in general as old as the language family in which they are found (except in isolated cases of borrowing, as we saw it is likely for English *dad*, and perhaps for Modern Turkish *baba* and Greek *babas* – and even then, these words still may be traced back to the origin of the donor language family). The Turkish word *ata* has not varied a bit since the time of the Orkhon inscriptions; nor have Welsh *mam* and *tad* since the Romano-British period, nor Indic *baba* since Prakrits were first written, nor has Tibetan *pha* since the time of the Tibetan Annals, nor have Chinese *baba* and *mama* since the Archaic Chinese period, nor Semitic *ʔab* and *ʔumm* since the time of Akkad and Ebla. Table 2 and 3 sum up all the historical information we have been able to collect about *mama/papa* words in the Ancient World.

Beyond the overall stability of these words over time, another remark may be done about Tables 2 and 3. Old languages belonging to the same family tend to resemble each other more than to other ones. This idiosyncrasy can be illustrated for instance by the fact that Indo-European languages as Sanskrit, Greek and Latin all show a CVCV form, while Semitic and Altaic languages display a VCV form, and Dravidian a VCCV form. More precisely, Dravidian *appa* does not look exactly like Semitic *ʔab-*, Bantu *bààbá* is not exactly like Turkic *apa*, nor Bantu *tààtá* is exactly like Turkic *ata*, etc. This fact, too, may only result from inheritance from a common source word in the mother language of the respective families.

So this is one of the main pillars of the convergence theory that is falling apart. There is nothing like an "endless recreation and recycling" of *mama/papa* words. This fact is now well established and also discards the traditional explanation of convergence in terms of babbling. The nursery kinship terms, phonetically simple as they are, but also both phonetically and semantically idiosyncrasic to each language family, are taught by the adults to the children; they do not spontaneously crop up from babbling. They belong to the lexicon of each language, just like any other word, and they are not simply interchangeable. Moreover, as address terms among adults, they fulfill a crucial social function.

One of the consequences of all these observations is, as we just said, to validate the reconstructions obtained through linguistic comparison. Some of these reconstructions reach the stage of high level phyla (Austric, Eurasiatic, Afroasiatic), which means we are back in time to ca. 10,000–20,000 years ago, and still witnessing this nasal/mother ~ oral/father opposition. But what is left now is to explain why so many *papa/tata* shaped-terms are specifically attached to the paternal entity and why so many *mama/nana* shaped terms are attached to the maternal entity. It is generally argued that Roman Jakobson already answered this question fifty years ago. As a matter of

fact, he did not. He contented himself by proposing very interesting clues. He notably observed (French edition 1969: 127) and we already mentioned it in our introduction, that the existing nursery words prompt children to gradually convert the nasal interjection, spontaneously emitted in association with feeding, into a parental term, *ie*: *mama*/*nana*. And just after this statement, he emphasized the fact that the words *papa* and *mama* were "learned." Hovewer, as Merritt Ruhlen in his article "Why *Kaka* and *Aya*" (2000: 527) observed, Jakobson also suggested some link between the order of apparition of parental terms and the order of development of consonant in child language. Ruhlen, mentioning other nursery kin words of comparable distribution, like *kaka* and *aya*, said that

> [i]f we are to extend Jakobson's explanation to take care of these examples, we must hypothesize that older male relatives somehow appear on the scene at the moment the child is learning velar consonants, which are not particularly early in their development, while older female relatives somehow come to be associated with *aya*. It does not seem to me plausible that human society could be organized so neatly […].

And it does not seem to us, either.

So "the inescapable conclusion" must be historical, and the "Proto-World conjecture" is not, as Larry Trask wrote, "dead in the water." As we demonstrated in previous articles (Matthey & Bancel; Bancel & Matthey 2002, 2005), there is only one explanation for the amazing distribution of *mama* and *papa* words: that these words are as old as the Proto-Sapiens language to which they used to belong.

Table 2. (*t*)*ata* forms in ancient languages reflected in modern related languages

Languages	Archaic terms	Derived terms	Minimal time span
Latin	*tata* F	Italian, Catalan, Spanish, Romanian *tata* F	2,300 years
Sanskrit	*tāta* F	Romani *tatta* F, Waigali	3,000 years
Pali	*tāta* F	*tata* F, Kol *dādā* F	2,200 years
Avestan	*tā* F	Besud *atá* F, Jaghuri *atai* F	3,000 years
Middle Welsh, Middle Breton, Old Cornish	*tad* ~ *tat* F	Modern Breton, Modern Welsh *tad* F	800 years
Gotic	*atta* F	German (dial.) *ätte* F	1,600 years
Old Slavon	*teta* F	Bulgarian *tata* F	900 years
Uighur	*ata* F	Azeri *ata* F, Shary-Yughur	1,200 years
Middle Turkic	*ata* F	*ata* F, etc.	900 years
Old Basque	*ata* F	Modern Basque *aita* F	1,700 years

Table 3. (*P*)*APA* ~ (*B*)*ABA* forms in ancient languages reflected in modern related languages

Languages	Archaic terms	Derived terms	Minimal time span
Latin	*pappa* F	French *papa* F, Italian *papa* F, etc.	2,000 years
Homeric Greek	*pappa* F voc. (Homer)	Modern Pontic Greek voc. *pappa* F	2,800 years
Prakrit	*bappa* F	Uriya *bappa* F, Marathi *bāpa* F, etc.	2,200 years
Khotanese	*paba* F	Pashto *bâbâ* F, GdF	1,300 years
Eastern Pahlavi	*bâbâ* first part of masc. name	Farsi *bâbâ* F, GdF	1,700 years
Orkhon	*apa* ancestor, GdF	Balkar *appa* ~ *aba* F	1,300 years
Karakhanid (Middle Turkic)	*aba* F, "ancestor,"	Kirghiz *aba* F, etc.	900 years
Written Mongolian	*aba* F	Monguor *aba* F	800 years
Middle Korean	*àpí* F	Modern Korean *appa* F	600 years
Old Akkadian	*abi* "my father"	Standard Arabic *abi* "my father"	4,500 years
Old Tamil	*-appan* honorific masc. suffix	Modern Tamil *appan* F, Kannada *appa* F, honorific masc. suffix	1,700 years
Old Tamil	*appan* F	Modern Tamil *appan* F	1,400 years
Old Tibetan	*pha* F	Modern Tibetan *pha* F	1,300 years
Old Chinese	*ba* F	Modern Mandarin, Cantonese *baba* F	3,300 years

Table 3. (*M*)*AMA* forms in ancient languages reflected in modern related languages

Languages	Archaïc terms	Derived terms	Minimal time span
Latin	*mamma* M	Italian *mamma* M, Romanian *mama* M, etc.	2,300 years
Classical Greek	*mammê*, *ammia* M	Modern Greek *mamma*	2,500 years
Middle Welsh, Middle Breton, Old Cornish	*mam* M	Modern Breton, Modern Welsh *mam* M	800 years
Pali	*ammā* M	West. Pahari *ammā* M	2,200 years
Old Basque	*ama* M	Modern Basque *ama* M	1,700 years
Old Chinese	*mə?* M	Modern Mandarin *ma* M	3,300 years
Eblaic	*ummum* M	Modern Arabic *ʔumm* M	4,500 years

References

Bacot, J., Thomas, F.W. & Toussaint, C. 1940. *Documents de Touen-Houang relatifs à l'histoire du Tibet.* Paris: Paul Geuthner.

Bancel, P.J. & Matthey de l'Etang, A. 2002. Tracing the Ancestral Kinship System. The Global Etymon kaka. Part I: A Linguistic Study. *Mother Tongue* 7. 209–243.

Bancel, P.J. & Matthey de l'Etang, A. 2005. Kin Tongue. A Study of Kin Nursery Terms in Relation to Language Acquisition, With a Historical and Evolutionary Perspective. *Mother Tongue* 9. 171–190.

Beck, B.E.F. 1972. *Peasant Society in Konku: a Study of Right and Left Subcastes in South India.* Vancouver: University of British Columbia Press.

Benedict, P.K. 1941a. *Kinship in Southeastern Asia*. Dissertation (Anthropology), Cambridge: Harvard University.

Benedict, P.K. 1941b. Tibetan and Chinese Kinship Terms. *Harvard Journal of Asiatic Studies* 6. 313–337.

Benedict, P.K. 1972. *Sino-Tibetan: a Conspectus*. Cambridge, UK: Cambridge University Press.

Bengtson, J.D. & Ruhlen, M. 1994. Global Etymologies. Ruhlen 1994b. 277–336.

Blažek, V. 2002. Toward the Berber Kinship Terminology in the Afroasiatic Context. *Articles de linguistique berbère, réunis et édités par Kamal Naït-Zerrad, Mémorial Werner Vycichl*, 103–135. Paris, Budapest, Torino: L'Harmattan.

Bossé, O. 1983. Le vocabulaire de parenté telugu. *L'Homme* 23: 2. 97–108.

Burrow, T. & Emeneau, M.B. 1984. *A Dravidian Etymological Dictionary*. 2nd Ed. Oxford [Oxfordshire]: Clarendon Press. Online: Digital Dictionaries of South India: http://dsal.uchicago.edu/dictionaries/burrow/.

Charles-Edwards, T.M. 2003. *Early Irish and Welsh Kinship*. Oxford: Oxford University Press.

Chantraine, P. 1968. *Dictionnaire étymologique de la langue grecque*. Paris: Klincksieck.

Corominas, J. 1954. *Diccionario Crítico Etimológico de la Lengua Castellana.* Berna (Bern): Editorial Francke.

Diccionari català-valencià-balear. http://dcvb.iecat.net/.

Dictionare ale limbii române. http://dexonline.ro/.

Dictionnaire historique de la langue française. 1992. Paris: Le Robert.

Dictionnaire Roumain-Français. 1992. Paris, Bucuresti: L'Harmattan, Editura Babel.

Emeneau, M.B. 1953. Dravidian Kinship Terms. *Language* 29: 3. 339–353.

Emeneau, M.B. 1955. *Kolami: A Dravidian Language*. Berkeley, Calif.: University of California Publications in Linguistics Vol 12.

Ehret, C. 1980. *The Historical Reconstruction of Southern Cushitic Phonology and Vocabulary*. Berlin: Dietrich Reimer.

Evolution of Human Language Project (EHL), Santa Fe Institute. http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl.

Fang, Kuei Li & South Coblin, W. 1987. *A Study of the Old Tibetan Inscriptions.* Institute of History and Philology, Academia Sinica, Special Publications, 91. Taipei: ROC.

Gelb, I.J. 1961. *Old Akkadian Writing and Grammar.* Chicago: The University of Chicago Press.

Greenberg, J.H. 2000–2002. *Indo-European and Its Closest Relatives: The Eurasiatic Family, vol. I: Grammar; vol. II: Lexicon*. Stanford, Calif.: Stanford University Press.

Grigson, W.V. 1949. *The Maria Gonds of Bastar*. London: Oxford University Press.

Guthrie, M. 1967–1971. *Comparative Bantu.* 4 vol. Farnborough, Hants: Gregg International.

Hage, P. 2003. On the Reconstruction of the Proto-Nostratic Kinship System. *Zeitschrift für Ethnologie* 128. 311–325.

Hayes, A.V. 2002. *Austric Glossary*: http://home.att.net/~lvhayes/Langling/langpg3.htm

Hoijer, H. 1956. Athapaskan Kinship Systems. *American Anthropologist* 58. 309–334.

Izard, M. 1965. La terminologie de parenté bretonne. *L'Homme* 5: 3–4. 88–100.

Jakobson, R. 1960. Why Mama and Papa? *Perspectives in Psychological Theory, Essays in Honor of Heinz Werner*, Ed. by B. Kaplan & S. Warner, 124–134. New York. [Repr. in R. Jakobson. 1969. *Langage enfantin et aphasie,* 121–130. Paris: Les Editions de Minuit.]

Karve, I. 1953. *Kinship Organization in India*. Poona: Deccan College Monograph Series.

Kwang-Kyu, Lee & Youngsook K. H. 1973. Teknonymy and Geononymy in Korean Kinship Terminology. *Ethnology* 12: 1. 31–46.

Krader, L. 1963. *Social Organization of the Mongol-Turkic Pastoral Nomads.* Indiana University Publications, Uralic and Altaic Series, 20. The Hague: Mouton.

Lewis, C.T. & Short, C. *A Latin Dictionary, Perseus Project.* http://www.perseus.tufts.edu/cgi-bin/resolveform?lang=Latin.

Liddell, H.G. & Scott, R. *A Greek-English Lexicon, Perseus project*. http://www.perseus.tufts.edu/cgi-bin/enggreek?

Mahadevan, I. 2003. *Early Tamil Epigraphy from the Earliest Times to the Sixth Century AD*. Cre-A. Chennai, India and the Department of Sanskrit and Indian Studies: Harvard University USA.

Matthews, G.H. 1959. Proto-Siouan Kinship Terminology. *American Anthropologist* 61: 2. 252–279.

Matthey de l'Etang, A. & Bancel, P.J. 2002. Tracing the Ancestral Kinship System: The Global Etymon kaka. Part II: An Anthropological Study. *Mother Tongue* 7. 209–243.

Matthey de l'Etang, A. & Bancel, P.J. 2005. The Global Distribution of (P)APA and (T)ATA and Their Original Meaning. *Mother Tongue* 9. 133–169.

Meeussen, A E. 1969. *Bantu Lexical Reconstructions*. http://www.linguistics.berkeley.edu/CBOLD/Docs/Meeussen.html.

Morgan, LH. 1871. *Systems of Consanguinity and Affinity of the Human Family.* Smithsonian Contributions to Knowledge, vol. XVII. Washington: Smithsonian Institution.

Murdock, G.P. 1959. Cross-Language Parallels in Parental Kin Terms. *Anthropological Linguistics* 1: 9. 1–5.

Nebrija (de), E.A. 1492. *Lexicon o diccionario latino-espan~ol* (facsimile). Bibliotheca Virtual Miguel de Cervantes. http://www.cervantesvirtual.com/servlet/SirveObras/07032774389636239647857/ima0192.htm, fol. XCVIIr.

Pali-English dictionary. 1921–1925. The Pali Text Society, Chipstead. http://dsal.uchicago.edu/dictionaries/pali/.

Richardson, H.E. 1985. *A Corpus of Early Tibetan Inscriptions*. Hertford: Royal Asiatic Society.

Ruhlen, M. 1994a. *The Origin of Language. Tracing the Evolution of the Mother Tongue*. New York: John Wiley.

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

Ruhlen, M. 2000. Why Kaka and Aya? *Functional Approaches to Language, Culture and Cognition*, Ed. by D.G. Lockwood, P.H. Fries & J.E. Copeland. Amsterdam-Philadelphia: John Benjamins.

Schurmann, F. 1962. *The Mongols of Afghanistan. An Ethnography of the Moghôls and Related Peoples of Afghanistan*. The Hague: Mouton.

Srinivas, M.N. 1942. *Marriage and Family in Mysore*. Bombay: New Book.

Strand, R. 1997–2008. *Nuristân Hidden Land of the Hindu-Kush. Kinship Terminologies of the Hindu-Kush.* http://users.sedona.net/~strand/kinship.html.

Tamil Lexicon. [Madras], University of Madras, 1924–1936. http://dsal.uchicago.edu/dictionaries/tamil-lex/.

Tesoro della Lingua Italiana delle Origini (TLIO). CNR – Opera del Vocabolario Italiano. http://tlio.ovi.cnr.it/TLIO/.

Trautmann, T.R. 1981. *Dravidian Kinship*. Cambridge, UK, New York: Cambridge University Press.

Turner, R.L. 1962–1966. *A Comparative Dictionary of Indo-Aryan Languages.* London: Oxford University Press. http://dsal.uchicago.edu/dictionaries/soas/.

Vocabularium Cornicum. http://www.carlaz.com/cornish/voccorn.txt, http://wikisource.org/wiki/Vocabularium_Cornicum.

Wordick, F.J.F. 1970. *A Generative-Extensionist Analysis of the Proto-Indo-European Kinship System, With a Phonological and Semantic Reconstruction of the Terms*. Ph.D. thesis, Ann Arbor (Michigan).

The millennial persistence of Indo-European and Eurasiatic pronouns and the origin of nominals

Pierre J. Bancel & Alain Matthey de l'Etang
Association for the Study of Language in Prehistory (ASLIP)

To Hal Fleming, a true mind and friend

In further defense of the Proto-Sapiens antiquity of global kinship etymologies, we illustrate the long-lasting survival of personal pronouns first in the Indo-European family, then in the Eurasiatic macrofamily of languages. We then put forward a conjecture about how the category of 1st and 2nd person pronouns might have originated.

1. Presentation

The claim that *papa, mama* and *kaka* kinship terms have been inherited since the ±100,000 year-old Proto-Sapiens origin of all human languages (Bengtson & Ruhlen 1994; Ruhlen 1994a; Bancel & Matthey de l'Etang 2002; Matthey de l'Etang & Bancel 2005) flies in the face of the widespread belief that words are replaced at such quick pace that it would erase any trace of an original vocabulary in less than 10,000 years. This belief would imply that ancestral words disappear from the lexicon of all languages at a relatively regular pace.

Such is not the case, however. Some words last much longer than others. Dolgopolsky (1964) already found that the 1st and 2nd person pronouns are the first and third most resistant word meanings, respectively. More recently, Pagel (2000, quoted in Ruhlen 2007) found comparable results and calculated that 1st person singular pronouns have an average half-life exceeding 21,000 years.

We will flesh out these statistical claims with solid linguistic facts. In the Indo-European language family, 1st person *m-* and 2nd person *t-* pronoun roots have been preserved in all descendant languages over some 7,000 to 8,000 years. Moreover, in the still more ancient Eurasiatic family, 1st person *m-* and 2nd person

t- pronouns have been preserved in most groups, and in almost all their descendant languages for at least 10,000 years, as claimed by Greenberg (2000). We conclude that Dolgopolsky and Pagel are right, and that some words may last for almost indefinite timespans. Thus, the a priori objection against the Proto-Sapiens antiquity of kinship terms, based on any average word replacement rate, simply does not hold. Finally, we propose a conjecture about the fact that, like most other ancient pronominal roots in the world's languages (Ruhlen 1994b), the *m-* and *t-* Eurasiatic pronouns are built from the same consonants as Proto-Sapiens kinship terms. In our opinion, the ancestral kinship terms have been the precursors of personal pronouns.

2. The exceptional persistence of Indo-European pronouns

Indo-European languages give us a clear example of the pronouns' exceptional resistance. Indo-European languages descend from Proto-Indo-European, a language spoken some 7,000 to 8,000 years ago. Proto-Indo-European was a language with nominal declension, marking cases with suffixes. So were all its early written descendants such as Hittite, Sanskrit, Avestan, Tocharian, Classical Armenian, Classical Latin, Classical Greek, Old Church Slavic, or Gothic, and still are many modern descendant languages such as Hindi, Modern Armenian, Rumanian, Modern Greek, Russian, German, and so on. We will examine here in full detail the 1st person pronoun only.

2.1 Indo-European 1st person pronoun

Personal pronouns took part to the nominal declension system. Among them, the 1st person pronoun had the striking peculiarity to be built on two different roots. There was an isolated form *eg^{h}om* "I" for the nominative (subject) case, alternating with a root *m-* bearing various suffixes marking the accusative, genitive, dative, and ablative cases. This use of a special word for the nominative of the 1st person singular pronoun was isolated within the Proto-Indo-European declension system. All other nouns, pronouns, and adjectives used the same root for the nominative as for all other cases, each case being marked by a specific suffix. From a functional viewpoint, the word *eg^{h}om* was doubly weird: (1) within the 1st person singular pronoun's declension, it implied to remember two entirely different word roots; (2) within the whole declension system, it implied to remember that the 1st person pronoun behaves differently from all other nominals. In an evolutionary perspective, the *eg^{h}om ~ m-* alternance was a typical candidate for an analogical simplification which would have generalized one of the two forms. We will see that this theoretical hypothesis is amply verified in the Indo-European languages' history.

Table 1. Evolution of *eg^{h}om* "I" ~ *me* "me" in languages descended from Proto-Indo-European

Italic		**Slavic**		**Armenian**	
Class. Latin	*egŏ ~ mē*	O. Ch. Slavic	*jazŭ ~ mę*	Cl. Arm.	*es ~ im* (gen.)
French	*je ~ me*	Russian	*ja ~ me(-nja)*	Mod. Arm.	*jes ~ im* (gen.)
Portuguese	*eu ~ me*	Ukrainian	*ja ~ me(-ne)*		
Spanish	*yo ~ me*	Polish	*ja ~ m(-nie)*	**Iranian**	
Catalan	*jo ~ me*	Czech	*já ~ me*	Avestan	*azəm ~ ma(-m)*
Occitan	*ieu ~ me*	Slovak	*ja ~ ma*	Farsi	*ma(-n)*
Italian	*io ~ me*	Serbo-Croat.	*ja ~ me*	Baluchi	*ma(-n)*
Rumanian	*eu ~ me*	Slovenian	*jaz ~ me*	Kurdish	*ez ~ mi(-n)*
		Bulgarian	*az ~ me*	Ossetic	*æz ~ my(-n)*
Celtic				Pashto	*zə ~ maa*
Old Irish	*mē*	**Germanic**			
Middle Welsh	*mi*	Gothic	*ik ~ mi(-k)*	**Nuristani**	
Mod. Welsh	*mi*	English	*I ~ me*	Kâmviri	*ōtθ ~ ıa*
Cornish	*my*	Dutch	*ik ~ mi(-j)*	Kâtavari	*uze ~ ie*
Breton	*me*	German	*ich ~ mi(-ch)*	Sañuvîri	*âi ~ yũ*
		Danish	*jeg ~ mi(-g)*		
Greek		Swedish	*jag ~ mi(-g)*	**Indic**	
Class. Greek	*egō(n) ~ (e-)me*	Icelandic	*ég ~ mé(-r)*	Sanskrit	*ahám ~ mā́(-m)*
Mod. Greek	*égo ~ mu*			Pali	*aham ~ ma(-m)*
		Anatolian		Hindi	*mẽ ~ m(-ujh)*
Albanian		Hittite	*ūk ~ amm(-uk)*	Bengali	*am-*
Albanian	*unë ~ më*	Luvian	*? ~ mi/mu*	Punjabi	*me(-n)*
				Gujarati	*hu(-n) ~ ma(-ri)*
Baltic		**Tocharian**		Kashmiri	*bŏh ~ me*
Old Prussian	*aš ~ ma(-nę)*	Tocharian A	*ñ-uk* f. ~ *n-ä s* . m.	Marathi	*mi ~ ma(-dža)*
Latvian	*es ~ ma(-ni)*	Tocharian B	*ñ-aś*	Romany	*me ~ ma(-n)*

The first form is that of nominative case; the second form is that of accusative, except for Armenian, where it is replaced by that of genitive, and Tocharian A, where the alternance is to be observed between the two feminine and masculine forms of nominative. Underlined languages have generalized a single root for all cases and did not preserve any trace of the ancient alternance. Elements between parentheses are part of the word in the descendant language, but have historically been added to the root by analogy. After 7,000 to 8,000 years of evolution, all thirteen Indo-European branches but Celtic have preserved the alternance between nominative and the other cases, with two forms most of the time derived from *eg^{h}om* and *m-*. All Indo-European languages, without a single exception, have preserved the non-subject root *m-*, sometimes under an evolved form but mostly unchanged. Note that in several groups the original Proto-Indo-European genitive suffix *-n* (in the PIE genitive *me-ne* "of mine") has been generalized to a great part of the declension (Russian, Latvian, Romany), or even to the whole paradigm (Farsi, Baluchi, Tocharian), or survives in the possessive (German *mein,* English *mine*).

Albeit apparently weakly functional, the two alternating forms *eg^{h}om ~ m-* have survived until the present day in an amazing number of Indo-European languages. Table 1 illustrates its preservation with representative examples of the 1st person singular pronoun in all the Indo-European linguistic groups. In Table 1, in the box facing each language name, the nominative form comes first (in Tocharian A, the first form is the feminine variant of the nominative), separated by a tilde (~) from the accusative form (or the genitive, for Armenian, or the masculine form of nominative, for Tocharian A). Thus, in Classical Latin *egŏ ~ mē, egŏ* is the nominative form ("I") and *mē* the accusative ("me"). Languages for which a single root (as Tocharian B *ñ-*) or word (as Old Irish *mē*) is given use the same root or word in all grammatical cases.

The first language listed in each group is the ancient language closest to the origin, e.g., Classical Latin for the Italic group, or Avestan for Iranian. This immediately allows to note that nearly all the ancient Indo-European languages preserved both a derivative of *eg^{h}om* in the nominative and a derivative of *m-* in the other cases: Hittite *ūk ~ amm(-uk),* Vedic Sanskrit *ahám ~ ma(-m)* (with two variants [m$^{\frac{1}{2}}$] and [m$^{\frac{1}{2}}$-m]), Avestan *azəm ~ mą(-m),* Classical Greek *egō* (variant *egōn*) *~ me,* Classical Latin *egŏ ~ mē,* Gothic *ik ~ mi(-k),* Old Church Slavic *jazŭ ~ mę,* Old Prussian *aš ~ ma(-nę),* Tocharian A *(ñ-)uk* (feminine nominative) *~ ñ-* (masculine nominative *ñ-äṣ* and all other cases), Classical Armenian *es ~ im* (genitive), etc. Only Old Irish lost any trace of *eg^{h}om,* like all other, more recent Celtic languages, and generalized a derivative of *me* in the nominative. (In the examples above, the elements between parentheses linked to the rest of the word by a hyphen are part of the form cited, but have been added to the original form in the course of history by an analogical process. For instance, the Gothic accusative *mik* "me" evolved from an original *mi* to which a final *-k* was added by analogy with nominative *ik* "I;" Venetic, an ancient Italic language, did the same with *eχo* "I" ~ *meχo* "me.")

Among the thirteen Indo-European branches, only the Celtic group did not preserve any trace of *eg^{h}om.* In this group, the root *m-* was generalized to all cases, nominative included. This analogical replacement confirms the poor functionality of using two roots for the same meaning. Table 1 offers several additional instances of this replacement in the Tocharian, Iranian, and Indic groups. In these groups, whose respective proto-languages did preserve the *eghom ~ m-* opposition, some languages – e.g., Farsi, Baluchi, Hindi, Marathi, or Romany – later lost it and replaced the form derived from *eg^{h}om* by a form derived from *m-*. The fact that these languages belong to different groups whose respective proto-languages preserved a form derived from *eg^{h}om* implies that this analogical simplification occurred several times independently. To these languages may be added the case of French, a Romance language in which the descendant form *je* of Latin *egŏ* has

become close to a verbal prefix marking a 1st person subject, while the subject form of the independent pronoun is now *moi,* derived from a stressed variant of Latin *mē*.

It is thus a heavy trend from the part of the speakers of any inflected language to use a single root for all the uses of a word. However, the career of the two alternating forms of Proto-Indo-European 1st person pronoun did not stop with the ancient languages descended from Proto-Indo-European. Three millennia later, they continue to coexist in a very high number of modern languages. In spite of the restriction mentioned above, Modern French preserves both a derivative *je* from Latin nominative *egŏ* and derivatives *me, moi* from Latin accusative *mē*. Indeed, Modern French grammar abandoned the whole declension system inherited from Latin except in pronouns. The use of different forms according to whether the personal pronoun is subject (*Je vois l'homme* "I see the man") or object (*L'homme me voit* "The man sees me") practically constitutes in French the only trace (with the relative pronoun) of the declension system inherited by Latin from Proto-Indo-European.

The loss of declension in Modern French makes the preservation of the opposition between *je* and *me* ~ *moi* still more amazing. On the one hand, French preserves an atypical and – in terms of mnemonic efforts – costly opposition between two roots bearing the same meaning (which several other Indo-European languages have independently reduced to a single root); on the other hand, by doing so, French preserves a trace of a grammatical system having otherwise entirely disappeared. As paradoxically as French, all other Romance languages do preserve the two forms, together with their grammatical distribution between subject and non-subject, although almost all of these languages – to the single exception of Rumanian – otherwise lost, like French, the ancestral declension system.

Beyond Romance languages, the two terms inherited from Proto-Indo-European have been preserved, in their respective original grammatical functions, in all Germanic languages (even though Modern English lost, like French, any trace of declension other than in pronouns) as well as in all Baltic and Slavic languages, Modern Greek and Modern Armenian. Modern Albanian also maintains the opposition between subject and non-subject forms; however, while it preserved the original *m-* for the latter, it is unsure whether the subject form *unë* contains a derivative of *eg^{h}om*.

A number of modern Indic languages descended from Sanskrit did, like Celtic, generalize the non-subject *m-* to all cases: Hindi and Urdu *mẽ,* Punjabi *me(-n),* Marathi *mi,* etc. However, Gujarati as well as a Dardic language like Kashmiri still preserve the root alternation between subject and non-subject forms (but they replaced the derivative of the Sanskrit subject form *ahám* by new forms).

Nuristani languages (Strand 1997–2008) preserved the alternance between subject and non-subject forms. Kâtavari and Kâmviri subject forms *uze* and *ōtθ* "I" clearly descend from Proto-Indo-European *eg^{h}om.* Nuristani non-subject forms have undergone the strongest phonetic erosion, to the point of hardly preserving a trace of the original *m-* as a nasalization of the neighboring vowel, as in Kâmviri *īa* "me" (but Kâtavari *ia* "me" even lost this nasal feature); however, the original *m-* survives in the 1st person singular possessive pronoun (Kâtavari *iema,* Sañuvîri *imâ* "my") as well as in the 1st person plural personal pronoun (Kâmviri *imo,* Kâtavari *imu,* Sañuvîri *ima* "we"). In the Iranian group, Farsi and Baluchi *ma(-n)* only preserve the non-subject form, generalized to all cases. To the contrary, Kurdish *ez ~ mi(-n)* preserves the original opposition, and so do Ossetic *æz ~ my(-n)* and Pashto *zə ~ maa.*

Moreover, *m-* has also been preserved in numerous 1st person plural pronouns (e.g., Russian *my,* Lithuanian *mes,* Armenian *mek‘* "we"), 1st person singular possessives (e.g., French *mon,* English *my, mine,* German *mein,* Russian *moj,* Hindi *mera,* "my"), and 1st person verb suffixes (e.g., Latin *sum* "I am," *sumus* "we are," *amabam* "I loved," *fuimus* "we were," French *nous allons* "we go," *nous fûmes* "we were," Classical Greek *eimi* "I am," *pheromai* "I bear for myself," *pheromen* "we bear," Irish *bim* "I am," Albanian *unë jam* "I am," *ne jemu* "we are," Latvian *esmu* "I am," Russian *ja dam* "I will give," *my berjom* "we bear," Armenian *berem* "I bear," Tocharian A *nas-am* "I am," *nas-amäs* "we are," Avestan *barāmi* "I bear," Sanskrit *bhar-āmi* "I bear," etc.). Finally, let us observe that the *-n* suffix of Proto-Indo-European genitive *mene* "of mine" was extended to the accusative and other non-subject cases in several Indo-European groups (e.g., Russian *menja,* Latvian *mani,* Kurdish *min* "me") and even generalized to the nominative (Tocharian B *ñaś*, Farsi *man,* Punjabi *men* "I").

In sum, the reality appearing from Table 1 is that all modern Indo-European languages – from Icelandic to Russian and Bengali through Welsh, Armenian, Greek, Kâmviri and Pashto) – ALL, WITHOUT A SINGLE EXCEPTION, preserve a derivative from the original *m-*. In a small number of cases, the *m-* survives under a phonetically evolved form where it is impossible to recognize it at first glance, for instance in Tocharian B *ñ-* or still more in the Armenian accusative *is*, as well as in some Nuristani languages. All Indo-Europeanists, however, agree that these forms descend from the original Proto-Indo-European *m-*. In some languages, like Tocharian B, Farsi, Hindi and the whole Celtic group, the original *m-* was not only preserved but analogically extended to the nominative, where it replaced the derivative of *eg^{h}om.* But in a great majority of languages, Proto-Indo-European *m-* did not change at all – neither in its phonetic form, nor in its grammatical functions, nor in its "1st person" meaning.

As to the nominative form *eg^{h}om,* it survived in at least 11 groups out of 13 (to the exceptions of Celtic and perhaps Albanian). In 8 out of these 11 groups (namely, Italic, Greek, Germanic, Baltic, Slavic, Armenian, Nuristani and Anatolian), it even survived

in all their member languages; in the 3 other groups (Tocharian, Iranian, and Indic), only part of the descendant languages preserved a derivative of *eg^{h}om.* In all of the numerous languages where a derivative of *eg^{h}om* was preserved, it continues to coexist along with a derivative of *m-* (of course, since *m-* was preserved in all languages). Furthermore, the original distribution of *eg^{h}om* and *m-* between subject and non-subject cases has remained the same (or, in a few cases, almost the same). In 7,000 to 8,000 years, the replacement rate of 1st person pronoun *m-* amounts to exactly 0%, and its preservation to 100%. As to the form *eg^{h}om,* if we consider groups, the loss is of only 2/13, or 15.4%; if we consider individual languages, the fair number of Indic languages which lost it would perhaps raise this percentage to about 20% to 30%.

According to the average 14% replacement rate per millennium calculated by Swadesh (1952, 1954), pronouns should have been preserved in only 34.8% (after 7,000 years) to 29.9% (after 8,000 years) of languages descended from Proto-Indo-European. The 100% preservation of *m-* and the 70% to 84.6% preservation of *eg^{h}om* make clear that an average loss rate is meaningless with regard to the 1st person pronoun.

2.2 Indo-European 2nd person pronoun

This exceptional resistance of 1st person pronoun is also observed in the Proto-Indo-European 2nd person pronoun *t-*. This pronoun did not display any root alternation between the nominative and other cases. Table 2 displays a few representative modern forms derived from its subject form *tu* in each Indo-European group.

Table 2. Representative forms of 2nd person singular pronoun in the nominative case in languages descended from Proto-Indo-European

Groups	Ancient languages	Modern languages
Italic	Cl. Latin *tū*	Spanish *tu,* French *tu,* Italian *tu,* Rumanian *tu*
Celtic	Old Irish *tū*	Welsh *ti,* Breton *te,* Modern Irish *ti*
Greek	Cl. Greek *sú* < *tú*	Modern Greek *(és-)si* < *sú*
Baltic	Old Prussian *tū*	Lithuanian *tú,* Latvian *tu*
Slavic	Old Church Slavic *ty*	Russian *ty,* Slovak *ty,* Bulgarian *ti,* Slovenian *ti*
Germanic	Gothic *θu*	Dutch *du,* English *thou,* German *du,* Icelandic *ðú*
Albanian	—	Albanian *ti*
Armenian	Class. Armenian *dow*	Modern Armenian *tun*
Anatolian	Hittite *zik* < *ti-uk* (accusative *tuk*)	
Tocharian	Tocharian A *tu* ~ Tocharian B *tuwe*	
Iranian	Avestan *tū*	Farsi *to,* Baluchi *tao,* Ossetic *dy,* Kurdish *tu,* Pashto *tə*
Nuristani	—	Kâtavari *t'ü*, Kâmviri *t'ü,* Sañuviri *tū*
Indic	Sanskrit *tw(-əm)*	Hindi *tū,* Kashmiri *ts(-ah),* Gujarati *tu,* Oriya *tu(-me)*

All Indo-European languages preserved a 2nd person singular pronoun inherited from Proto-Indo-European *tu.*

Just like for 1st person *m-*, not a single Indo-European language did lose the Proto-Indo-European 2nd person *t-*. And most of them preserved, like for 1st person *m-*, the original consonant *t-* either unchanged or only slightly modified. Second person *t-* is also just as general as 1st person *m-* in possessives and verbal markers. Indeed, mentioning and analyzing all 1st person *m-* forms and all 2nd person *t-* forms in Indo-European languages would take a whole book – for each.

2.3 Indo-European numerals

A few other Proto-Indo-European words also survived in nearly all of the modern descendant languages, such as the basic numerals from "two" to "ten." Examples of derivatives of *dṷou* "two," taken from all thirteen Indo-European branches, are given in Table 3. As for 1st person *m-* and 2nd person *t-*, *dṷou* "two" survived in all and every descendant language of Proto-Indo-European – see the nice online compilation of the first ten numerals in over 4,500 languages by Mark Rosenfelder (no date).

Table 3. Representative forms of Proto-Indo-European *dṷou* "two" in descendant languages

Groups	Ancient languages	Modern languages
Italic	Cl. Latin *dúo*	Spanish *dos,* French *deux,* Italian *due,* Rumanian *doi*
Celtic	Old Irish *deu*	Breton *daou,* Modern Irish *dhaü*
Greek	Cl. Greek *dúo* ~ *dúō*	Modern Greek *ðío*
Baltic	Old Prussian *dwai*	Lithuanian *du,* Latvian *divi*
Slavic	Old Church Slavic *dŭva*	Russian *dva,* Polish *dwa,* Slovenian *dva*
Germanic	Gothic *twai*	Old Norse *tveir,* English *two,* German *zwo* ~ *zwei,* Danish *to*
Albanian	Old Albanian *do*	Tosk *dy*
Armenian	Class. Armenian *erk'u*	Modern Armenian *yergu*
Anatolian	Hittite *da*, Lycian *tuwa*, Luwian *duwa*	
Tocharian	Tocharian A *wu* ~ Tocharian B *wi*	
Iranian	Avestan *dva*	Farsi *do,* Ossetic *dyuö,* Kurdish *du,* Pashto *dva*
Nuristani	—	Kâtavari *d'ü,* Kâmviri *d'ü,* Wasiweri *lüü*
Indic	Sanskrit *dwau*	Hindi *do,* Kashmiri *zăh,* Marathi *don,* Oriya *dui*

All Indo-European languages preserved a form inherited from Proto-Indo-European *dṷou* "two."

In this series, the Armenian forms seem unrelated to the other words, but it is a well-known achievement of linguistic reconstruction to have shown that Classical Armenian *erk'u* "two" and Modern Armenian *yergu* result from a regular phonetic evolution of Proto-Indo-European *dṷou* "two."

2.4 Why persistent words, and what do they mean for language history?

A comparably high preservation rate of personal pronouns is encountered in most language families. This extraordinary persistence may be explained in a simple

word: FREQUENCY. Personal pronouns – and more generally person markers – are among the most frequent meaningful words in all languages: in the everyday discourse, most sentences include 1st and/or 2nd person markers.

This resistance to linguistic erosion gives pronouns a very high classificatory value, well known to long range comparatists. A recent attempt to invalidate this classificatory value of pronouns, explicitly directed against long-range comparison, claimed that there are languages where personal pronouns are "freely borrowed" from other languages (Thomason & Everett 2005). Their world tour in search of such cases resulted in less than two dozen languages – out of some 7,000 – in which they found examples apparently supporting this claim. Among them, about a half concern 3rd person pronouns – which are different in nature, as was shown long ago by Benveniste (1946). Finally, among their examples concerning 1st and 2nd person pronouns, several are more than dubious, e.g., when they ascribe the striking convergence between 1st person pronouns in Turkic, Mongolic and Tungusic languages (see sections 3.2 to 3.4 below) to borrowing rather than to common inheritance from Proto-Altaic – while they admit overtly that they "have not carried out any kind of systematic analysis of the Korean, Japanese, Tungusic, Mongolian or Turkic data."

On the basis of such scanty and uncertain evidence, the authors of this study go so far as concluding that "we can't know […] whether a given past culture was more like Indo-European languages, where most pronouns are in fact inherited, or more like those of languages whose speakers borrow pronouns freely." In fact, the weakness of their study *a contrario* demonstrates that pronoun borrowing is really exceptional, unintentionally confirming the classificatory value of 1st and 2nd person pronouns.

Less exceptional is pronoun replacement WITHIN a given language, due to a specific cultural attitude, leading for instance to replace the 2nd person singular pronoun by a plural or a 3rd person as a mark of respect (English lost its 2nd person singular by borrowing from Old French, if not the pronoun itself, the use of the 2nd plural when speaking to a single addressee), or tabooing the use of 1st person singular pronoun as an impolite expression of selfishness. It is worth noting, by the way, that such a taboo may have been at the origin of the Indo-European suppletion between the nominative and other cases in the 1st person pronoun: the taboo would have borne on the subject case, i.e., in sentences where the speaker directly refers to his own actions or feelings, but not on non-subject cases, where syntax makes the role of the speaker more incidental in the action or state described.

In contrast, most other Proto-Indo-European words known to us – even other items from the Swadesh's 200 basic words list – have been preserved in much less numerous modern languages, often in only two or three branches – and, within these branches, in only some languages. Contrary to numerals and personal pronouns, such words underwent a much higher replacement rate than 14% per millennium.

Consequently, it cannot be assumed that an average 14% replacement rate per millennium does lead, after 1,000 years, to the loss of any one word out of seven or so from the ancestral vocabulary.

2.5 Personal pronouns in Eurasiatic

It is thus an established fact that Indo-European personal pronouns *m-* and *t-* have been preserved for some 8,000 years in all the numerous descendant languages of the family. If these forms have been preserved in several hundreds of languages over such a long timespan, it might well happen that they had not been new in Proto-Indo-European, but had been inherited from a remoter ancestor language. This ancestral language, of course, could have other descendants than Indo-European. And, given the pronouns' extraordinary persistence, it seems worth trying and finding their trace in the possible sister languages of Indo-European as well.

The Eurasiatic hypothesis (Greenberg 2000, 2002) links Indo-European to six other language families, namely Uralic-Yukaghir, Altaic, Nippo-Koreo-Ainu, Gilyak, Chukchi-Kamchadal, and Eskimo-Aleut. (To these seven families, Greenberg tentatively adds Etruscan.) Sixty-nine common grammatical forms (Greenberg 2000) and some 430 common lexical roots (Greenberg 2002) support the Eurasiatic unity. Among them, 1st person *m-* and 2nd person *t-* are among the most widely represented grammatical roots within the Eurasiatic family. Most of the data mentioned below about these two pronouns in the various Eurasiatic groups are given by Greenberg (2000) in his Eurasiatic grammatical etymologies (n° 1, 1st person *m*; n° 4, 2nd person *t*; n° 7, pronoun base *ge*; n° 17, plural *r(i)*; and n° 25, genitive *n*). Additional data have been taken from Starostin *et al.* (2003; no date), Bomhard (2007), and a range of descriptive works. Most analyses presented here are due to Greenberg – our contribution is essentially to synthesize his observations on Eurasiatic pronouns, which appear in his book dispersed over several sections.

All these families have a 1st person pronoun *m-*, and in four of them (Uralic-Yukaghir, Altaic, Chukchi-Kamchadal, and Eskimo-Aleut) it is not only present but also a canonical marker for 1st person independent pronouns, possessives pronouns, and verb markers, just like in Indo-European. Moreover, in six families (the same four plus Nippo-Koreo-Ainu and Etruscan), we find both a root alternation between the nominative and other cases, and an extension *-n* which was more or less generalized to non-subject cases (and sometimes also to the nominative), but which initially marked the genitive case, as appears from the recurrence, throughout these families, of a *mVn-* 1st person possessive (ultimately derived from the genitive of the independent pronoun, like English *(of) mine* or German *mein* derive from the Proto-Indo-European genitive *mene* "of me").

As to the Indo-European nominative suppletive form *eg^{h}om,* it is less widespread, since several groups alternating two forms of the 1st person pronoun have another root in the nominative case. Clear cognates of *eg^{h}om* are however found throughout Chukchi-Kamchadal (with the most salient attestations in Chukchi), together with somewhat less close parallels in several Uralic languages.

The only language in which the Eurasiatic 1st person marker *m-* seems completely lacking is Ainu. It is not a surprise, since Ainu is certainly the most diverging language within Eurasiatic. Even within the Nippo-Koreo-Ainu family, Ainu is so divergent from both Japanese and Korean that most comparatists – and even many long-rangers – do not agree that it is related to either of these two languages, nor to anything resembling the Eurasiatic macrofamily.

The 2nd person marker *t-* is almost as widespread in Eurasiatic as is the 1st person *m-*, being a canonical marker in Uralic-Yukaghir, the Mongolic branch of Altaic (plus clear traces in Tungusic), Chukchi-Kamchadal, and Eskimo-Aleut. It however seems to be lacking in the Turkic branch of Altaic, the whole Nippo-Koreo-Ainu family, and Gilyak. Once again, the specific Indo-European form *tu* in the nominative case finds a clear parallel in Chukchi-Kamchadal, with the pervasive *turi, tuze, tuza* "you" plural forms in both the Chukchi and Kamchadal branches, and most strikingly, in the latter branch with three different attestations *tu ~ tue ~ tua* "thou" from various extinct Itelmen dialects independently collected in the course of the XIXth century by American and Russian explorers.

Table 4. 1st person *m-* and 2nd person *t-* in Uralic-Yukaghir personal pronouns

Language	1st SG.	1st PL.	2nd SG.	2nd PL.
Proto-Uralic	*mɤ*	*me*	*tɤ*	*te*
Estonian	*ma ~ mina min-*	*me*	*sa* (< *ta*)	*te*
Finnish	*mä ~ minä min-*	*me*	*sinä* (< *tinä*)	*te*
Sami	*mun-*	*mí*	*dun*	*dí*
Komi	*me menõ*	*mi*	*te tenõ*	*ti*
Udmurt	*mon-*	*mi*	*ton*	*ti*
Hungarian	*én* (< *émn*) *engem(-et)*	*mi*	*te teged(-et)*	*ti*
Enets	*mod'i*	*mod'iʔ*	*tod'i*	*tod'iʔ*
Nganasan	*mannaŋ*	*meeŋ*	*tannaŋ*	*teeŋ*
Yukaghir	*met*	*mit*	*tet*	*tit*

2.6 Personal pronouns in Uralic-Yukaghir

In all Uralic languages as well as in Yukaghir, the 1st person pronoun is built on the root *m-* and the 2nd person pronoun is built on the root *t-* (Table 4). For the 1st person,

the non-subject suffix *-n* is present in most languages from all the branches of Uralic, where it displays a strong tendency to generalization to the nominative (e.g., Finnish, Ostyak), and sometimes to the 2nd person as well (e.g., Sami, Udmurt, or Nganasan). The Proto-Uralic 1st person singular subject marker on verbs is *-m* (plural *-mek*), and the 2nd person is *-t* (plural *-tek*); the 1st and 2nd person singular possessors are marked by the same forms suffixed to nouns.

Moreover, Hungarian offers an accusative form strongly reminiscent of Proto-Indo-European *eg^{h}om* "I," analyzed by Greenberg (2000) as a compound *e-g^{h}o-m* "it is me" in which the final *-m* is the 1st person marker. In Hungarian, the form would have been turned into a redundant *en-ge-m(-et)* "me that's me" (the final *-et* marks the accusative case), an analysis confirmed by the Hungarian 2nd person pronoun accusative *te-ge-d(-et),* where both initial *te-* and final *-d* clearly derive from 2nd person *t-*. Other Uralic forms such as Vogul *am-kke-m* "I alone," the Selkup accusative *šim* (< *kim*) "me" and maybe the Kamassian suppletive form of the verb "to be" in the present tense (1st person *i-gä-m* "I am") may belong to the series, which finds a clear parallel in Chukchi-Kamchadal (see Table 10 below).

2.7 Personal pronouns in the Turkic branch of Altaic

The Turkic branch of Altaic is divided into two main groups, Chuvash and non-Chuvash. Chuvash offers an alternation between nominative *e-pə* and the non-subject cases built on the root *m-* plus the non-subject suffix *-n*, giving in Chuvash *man-* (Table 5). Chuvash, the most divergent Turkic language, thus parallels exactly the general state of things for the 1st person pronoun within the two other Altaic branches, Mongolic (see Table 6) and Tungusic (see Table 7).

Table 5. 1st person *m-* in Turkic personal pronouns (no 2nd person *t-*)

Language	1st SG.	1st PL.	Language	1st SG.	1st PL.
Chuvash	*e-pə man-*	*e-pir pir*	**Azeri**	*män-*	*biz-*
Old Turkish	*ben- ~ men-*	*biz-*	**Mod. Turkish**	*ben-*	*biz-*
Uzbek	*men-*	*biz-*	**Gagauz**	*ben-*	*bis-*
Yakut	*min-*	*bihigi*	**Kazakh**	*min-*	*biz-*
Shor	*men-*	*pis-*	**Kirghiz**	*men-*	*miz-*
Oyrot	*men-*	*mis-*	**Tatar**	*min-*	*bez-*

In the other Turkic group, the Old Turkish free alternation *ben- ~ men-* in the 1st person singular illustrates the first step in the reinterpretation of the formerly grammatically distributed variants, together with the generalization of the non-subject suffix

-n to the whole singular paradigm. In present-day non-Chuvash Turkic languages, the two variants *b-* and *m-* are variously distributed between singular and plural, according to languages: Modern Turkish generalized *b-* to both singular and plural (where suffix *-n* is replaced by the plural suffix *-iz* > *-ri*, an alternation general in Turkic); conversely, Oyrot, Kirghiz and Balkar generalized *m-* to both singular and plural; many languages, such as Tatar, Azeri, Khalaj, Yakut, or Nogai, generalized *m-* in the singular pronoun and *b-* in the plural. As to the former non-subject suffix *-n,* it was already generalized to all singular cases in Old Turkish and is retained until today by all non-Chuvash languages as the mark of the singular in the 1st person pronoun. 2nd person marker *t-* seems to be lacking in Turkic, as the canonical *sen* "thou" ~ *siz* "you" are not thought to have evolved from *t-* forms. The non-subject suffix *-n,* which was generalized in the 1st person singular, appears in the 2nd person singular pronoun as well, like in many Uralic languages.

2.8 Personal pronouns in the Mongolic branch of Altaic

In Mongolic, the alternation between the suppletive 1st person *b-* in the nominative and *mVn-* in other cases was extended to the plural (Table 6). Once again, we retrieve the same configuration as for the Indo-European 1st person, with a root alternation opposing the nominative to other cases, with a root *m-* and a suffix *-n* in the genitive that was generalized in Mongolic to both the singular and plural of the 1st and 2nd persons. The only difference is that, instead of *eg^{h}om* for the suppletive nominative form, we have a nominative root in *b-*, like in Chuvash. It is certainly interesting to observe that, phonologically, this *b-* is the non-nasal counterpart of the non-subject root *m-*.

Table 6. 1st person *m-* and 2nd person *t-* in Mongolic personal pronouns

Language	1st SG.		1st PL. EXCL.		2nd SG.	2nd PL.		1nd PL. INCL.	
Mid. Mongolian	*bi*	*mino*	*ba*	*mano*	*či čino*	*ta*	*tan-*	*bida*	*bidan-*
Class. Mong.	*bi*	*minu*	*ba*	*man-*	*či činü*	*ta*	*tan-*	*bida*	*bidan-*
Khalkha	*bi*	*minii*	*biddə*	*man-*	*či činii*	*tā*	*tan-*	*biddə*	*biddn-*
Ordos	*bi*	*mini*	*bida*	*man-*	*či čini*	*ta*	*tan-*	*bidə*	*bidn-*
Kalmuk	*bi*	*minē*	*bidə*	*man-*	*či činə*	*ta*	*tan-*	*bidʌ́*	*bidn-*
Buriat	*bi*	*menī*	*bɛde*	*man-*	*ši šenī*	*tā*	*tan-*	*bɛde*	*man-*
Dagur	*bī*	*minī*	*bā*	*mān-*	*ši šinī*	*tā*	*tan-*	*beda*	*bedańī*
Monguor	*bu*	*muni*	*buda*	*ndāni*	*ći ćini*	*ta*	*tan-*	*buda*	*budasgi-*
Paoan	*be*	*mene*	*mange*	*mane*	*čə čənə*	*ta*	*tan-*	*bede*	*bedan-*
Dongxiang	*bi*	*mini*	*bidžien*	*ma-*	*čə čəni*	*ta*	*tan-*	*matan*	*matan-*
Moghol	*bi*	*mini*	*(-u̯)*	*mōn-*	*či činɛi*	*to*	*ton-*	*bidä*	*mōn-*

The form on the left is the nominative, that on the right is the genitive.

As to the 2nd person, the alternation opposing *č* in the singular to *t-* in the plural is obviously a consequence of the following vowel, the singular *-i* having transformed a former *t-* into *č* – a well-known and widespread effect of closed front vowels like *i*. Finally, the 2nd person *t-* is certainly present in the Mongolic 1st person plural inclusive – i.e., encompassing the speaker and the hearer(s) – *bi-da* "we [= I + thou/you]", a clear compound of the 1st person singular *bi-* "I" and a 2nd person plural pronoun *-ta*.

2.9 Personal pronouns in the Tungusic branch of Altaic

In Tungusic (Table 7), the 1st person pronoun follows the same pattern already observed in Chuvash and the whole Mongolic branch, which also left clear traces in the non-Chuvash group of Turkic. The nominative root consonant is *b-*, and the non-subject cases are built on *mV-n-*. Like in Mongolic, this pattern was generalized to the 1st person plural pronoun.

Table 7. 1st person *m-* and 2nd person *t-* in Tungusic personal pronouns

Language	1st SG.		1st PL.		2nd SG.	2nd PL.	1st PL. INCL.
Even	*bi*	*min-*	*bu*	*mun-*	—	—	*mut*
Evenki	*bi*	*min-*	*bu*	*mun-*	—	—	*mit ~ bit*
Negidal	*bi*	*min-*	*bu*	*mun-*	—	—	*bitta*
Solon	*bi*	*min-*	*bū*	*mun-*	—	—	*miti*
Nanai	*bi ~ mi*	*min-*	*bū ~ mu*	*mun-*	—	—	*bue*
Udehe	*bi*	*min-*	*bu*	*mun-*	—	—	*minti*
Jürchen		*min-*		—	—	—	—
Manchu	*bi*	*min-*	*be*	*men-*	—	—	*muse*

The root consonant *t-* does not appear in the 2nd person pronouns, but, as in Mongolic, it is clearly seen in the compounded 1st person inclusive pronoun.

2.10 Personal pronouns in Nippo-Koreo-Ainu

The Nippo-Koreo-Ainu taxon, posited by Greenberg as one of the seven primary Eurasiatic groups, is certainly the most disputed of all seven (Table 8). The relation between Korean and Japanese is popular among long-rangers, who often posit a relation of the two with Altaic – an idea dismissed by Greenberg as a geographical artifact. Ainu is still considered as an isolate by traditional linguists, and several long-rangers link it to Austric rather than to Eurasiatic. Pronouns illustrate these divergences, showing that these three languages, and particularly Ainu, are highly divergent within Eurasiatic.

Table 8. 1st person *m-* in Nippo-Koreo-Ainu personal pronouns

Language	1st SG.		1st PL.
Modern Japanese	*wa*		—
Archaic Japanese	*wa*	*wanu*	—
Ryukyuan	*baa*	*banu*	—
Korean	—		? *uli* (if *uli* < *wuli* < *buli*)
Ainu	—		—

Only Japanese, and particularly the quite specific Ryukyuan dialects, display a 1st person pronoun reminding a Eurasiatic pattern, with its root *b-* in the nominative to which adds an extension *-n* in the accusative. It is closely parallel to the Altaic pattern, except that the root alternation *b-* ~ *m-* is not present, a fact which could easily be explained by an analogical simplification of the kind observed in several Indo-European branches and in non-Chuvash Turkic. The same simplification also occurred in Uralic – except that, in this latter family, it is the *m-* which was generalized to the whole paradigm. A single attestation *wanu* "me," from an undetermined Archaic Japanese dialect, parallels this pattern with *wa* (subject) ~ *wanu* (non-subject), though its root consonant must have undergone a classical phonetic evolution *b* (> *v*) > *w.*

With *uli* "we," possibly descended from an earlier and unattested *wuli* > *buli,* Korean also presents a possible trace of a Eurasiatic 1st person plural pronoun. Ainu seems to lack any trace of a pronoun relatable to the 1st person pronoun *m-* or *b-* series, nor does any language in the Nippo-Korean-Ainu group display anything like a 2nd person pronoun *t-*.

2.11 Personal pronouns in Gilyak

Gilyak, traditionally considered as a language isolate, displays a clear trace of 1st person *m-* only in the dual and plural forms (Table 9). In the dual *megi* "the two of us," *-gi* is the canonical Gilyak dual suffix, which reflects a Eurasiatic dual marker *ki* (Greenberg 2000, etymology 14), also attested in Uralic (and perhaps Yukaghir), Turkic, Ainu, Chukchi-Kamchadal and Eskimo, as well as possibly in Indo-European.

Table 9. 1st person *m-* and 2nd person *t-* in Gilyak personal pronouns

Language	1st SG.	1st dual	1st PL.	2nd SG.	2nd PL.
Gilyak	—	*megi*	*mer*	*či* ~ (*ti*)	—

In the plural *me-r* appears the Eurasiatic plural marker *-ri,* also present in Indo-European (e.g., Latin *ama-mu-r* "we are loved") and in the plural of Turkic pronouns, as well as in Chukchi-Kamchadal plural pronouns, e.g., Chukchi *mu-ri* "we," *tu-ri*

"you;" possible parallels appear in Tungusic, Korean, and Ryukyuan (Greenberg 2000, etymology 17). The 2nd person *t-* appears in a single attestation from a XIXth century description giving *ti* "thou" instead of the usual Gilyak pronoun *či.*

2.12 Personal pronouns in Chukchi-Kamchadal

In the Chukchi-Kamchadal family, 1st person *m-* and 2nd person *t-* are present in both the singular and plural pronouns, with very different respective patterns in the singular and the plural (Table 10). In the singular, they are preceded by a prefix *gə-,* giving e.g., Chukchi *ɣə-m* "I," *ɣə-t* "thou," or Koryak Palana *ɣə-mme* "I," *ɣə-tte* "thou." In Western Itelmen *kə-mma* "I," *kə-zza* "thou," the 2nd person marker *t-* changed to *z-,* through phonetic evolution or under influence of the plurals *mu-za* "we," *tu-za* "you."

Table 10. 1st person *m-* and 2nd person *t-* in Chukchi-Kamchadal personal pronouns

Language	1st SG.		1st PL.		2nd SG.		2nd PL.	
Chukchi	*(e)ɣəm*	*ɣəmnin*	*muri*	*muryin*	*(e)ɣət*	*ɣənin*	*turi*	*turyin*
Palana	*ɣəmme*	*ɣəmnan*	*muri*	*moryinan*	*ɣətte*	*ynanna*	*turi*	*toryinan*
W. Itelmen	*kəmma*	*kəmma'n*	*muza*	*muza'n*	*kəzza*	*kəzza'n*	*tuza*	*tuza'n*
Sth. Itelmen	*kim*	*ma*	*muš*	*burin*				
Nth. Itelmen			*buže*					
N.E. Itelmen					*tu*			
E. Itelmen			*buze*		*tue*		*suze*	
Itelmen X					*tua*			

This pattern parallels Indo-European *eg^{h}om* "I," and Chukchi offers an exact correspondent of Indo-European *eg^{h}om* in the suffixed predicative form of singular pronouns: X-*eyəm* "I am X," X-*eyət* "thou art X" (litterally, "X-I" and "X-thou"), suffixed to both nouns (e.g., *ənpənačy-iyəm* "I am an old man" [litt. "old man-I"]) and stative verbal forms (e.g., *yə-čejv-iyəm* "I am gone" [litt. "gone-I"]; the variant *-iyəm* results from vowel harmony). Greenberg analyzes these forms (and those already seen in Uralic) as compounds of a demonstrative *e-* plus a "pronoun base" *-ge-* used as copula plus 1st person *- m.* In our opinion, the pronoun base *-ge-* is linked to the Eurasiatic comitative *ko(n) ~ ko(m)* (Greenberg's etymology 34, strikingly represented essentially in Indo-European and Chukchi-Kamchadal).

The plural pronouns are built without prefix, but receive a suffixed plural marker, witness the Chukchi opposition *mu-ri* "we," *tu-ri* "you," paralleled by Western Itelmen *mu-za* "we," *tu-za* "you." Greenberg quotes XVIIIth and XIXth centuries sources of particular interest, which describe three different extinct Itelmen dialects with an *m- ~ b-* alternation (strongly reminding the Altaic alternation) distributed over the subject and non-subject forms of the 1st person plural, e.g., Southern Itelmen *mu-š*

"we (subj.)," *bu-ri-n* "our" (literally "of us"), where appear both the *-ri* plural suffix already found in Latin, Turkic, Korean and Gilyak, and the now familiar non-subject suffix *-n*. This non-subject suffix *-n* is general in all the singular and plural non-subject forms of Chukchi-Kamchadal pronouns.

As observed by Greenberg, Chukchi-Kamchadal is with Indo-European the only Eurasiatic group attesting a vowel *u* for the 2nd person pronoun, which was apparently generalized to the 1st person plural. In all surviving Chukchi-Kamchadal languages, vowel *u* was lost in the 2nd person singular, probably as a result of the generalization of the *yə-t* "thou" pattern. However, ancient independent sources once again give 2nd person singular forms *tu-* "thou" for three extinct Itelmen dialects, thus establishing another close parallel between Chukchi-Kamchadal and Indo-European. The Chukchi 2nd person plural subject marker is *-təkˌ* just like in Uralic. Finally, non-subject forms of both singular and plural 1st and 2nd person pronouns regularly display a suffix *-nˌ* which also appears in all Chukchi-Kamchadal possessives, strongly contributing to establish its original genitive function.

2.13 Personal pronouns in Eskimo-Aleut

In the Eskimo branch, the 1st person singular and plural pronouns display an inter-dialectal variation between a form in *m-* in Sirenik, an Eskimo language spoken in Siberia and probably constituting a primary branch of Eskimo, together with Yupik and Inupik (unless it is a primary branch of Yupik), and a *w-* ~ *v-* form in Yupik and Inupik (Table 11). This alternation reminds the *m-* ~ *b-* alternation found throughout Altaic and in some Itelmen dialects.

Table 11. 1st person *m-* and 2nd person *t-* in Eskimo-Aleut personal pronouns

Language	1st SG.	1st PL.	2nd SG.	2nd PL.
Aleut	—	*-man/-mas*	—	—
Sirenik	*məŋa*	*məkəta*	*əɬpi*	*ɬpəsi*
Chugach	*xwi*	*xwaŋkuta*	*äɬpäk*	*äɬpici*
Central Alaskan Yupik	*wīī wīīŋa*	*waŋkuta*	*əɬpət*	*əɬpəci*
North Alaskan Inupik	*uvaŋa*	*uvaɣt*	*ilvic*	*iɬifsi*
East Canadian Inupik	*uvaŋa*	*uvaɣt*	*ivvit*	*ilitsi*
Greenlandic	*uaŋa*	*uaɣt*	*iɬɬit*	*iɬɬissi*

In different Yupik dialects an alternation between a naked subject form and a non-subject form with a nasal suffix is observed, e.g., Central Alaskan Yupik *wī* ~ *wīŋa*. Although this nasal is *-ŋ* instead of the expected *-nˌ* the alternation is quite close to the usual Eurasiatic pattern. In Aleut, the 1st person *m-* is present only in the suffixed plural forms *-man* (Eastern and Western Aleut) or *-mas* (Central Aleut).

The 2nd person *t-* is not as obviously represented in Eskimo personal pronouns, but specialists in Eskimo count the final consonant *-t ~ -n* of *ǝɬpǝt~ ilvin* as the root of this pronoun. The interdialectal oral ~ nasal alternation in word-final stops is a well-known feature of Eskimo languages, extending well beyond the scope of personal pronouns (and occurring among labial and velar stops as well). More direct is the 2nd person singular possessive suffix *-t* (in western Eskimo languages *-n*); the 2nd person singular subject marker of intransitive verbs is *-tit,* dual *-tik* (with the dual suffix *-ki* already found in Gilyak *me-gi* "the two of us," and observed with a plural value in the 2nd person plural verbal markers of Uralic *-tek* and Chukchi *-tək*). In Aleut, the 2nd person singular possessive *-n* regularly corresponds to Eskimo *-t,* and the 2nd person plural verbal marker is *-ðix* (< *tik* ?).

2.14 Personal pronouns in Etruscan

Among Etruscan pronouns, only the 1st person singular is known (Table 12). It displays an alternation between a nominative form *mi* and a non-subject form *mini.* The presence of a root *m-* plus a non-subject suffix *-n* is a strong indication that Etruscan belongs to the Eurasiatic macrofamily. According to Greenberg, it might even indicate that Etruscan is in fact Indo-European, although one might note that the particular form of the Etruscan alternation is more like that canonical in Uralic, e.g., Finnish *mä ~ minä* "I" opposed to *minä* "me."

Table 12. 1st person *m-* in Etruscan personal pronouns

Language	**1st sg.**		**1st pl.**	**2nd sg.**	**2nd pl.**
Etruscan	*mi*	*mini*	**?**	**?**	**?**

Conclusion

As in Indo-European, a 1st person pronoun *m-* is attested in all member languages of the Uralic-Yukaghir, Altaic, Chukchi-Kamchadal and Eskimo-Aleut families (plus Etruscan). Moreover, all these six language families attest to a root alternation between the nominative and non-subject forms, together with a suffix *-n* in the non-subject form. In Nippo-Koreo-Ainu and Gilyak, its presence is less certain, although the Ryukyuan *baa ~ banu* alternation, though deprived of *m-,* closely resembles the Altaic pattern, while Gilyak dual *megi* "the two of us" and plural *mer* "we," with their clearly Eurasiatic dual *-ki* and plural *-ri* suffixes, also seem pretty secure.

Thus, it is not only a common word root that these language families share with the 1st person *m-,* but a complex grammatical pattern involving a frequent root suppletion in the nominative and a suffix *-n* in other cases, certainly having originated as a genitive. Is it possible to imagine that such a complex pattern might have arisen six times by chance (if we leave aside the less obvious Nippo-Koreo-Ainu and Gilyak families)? It seems highly unlikely. As to the 2nd person pronoun *t-*, it is somewhat

less widely represented in Eurasiatic languages. Nevertheless, it is preserved in all languages of Indo-European, Uralic-Yukaghir, the Mongolic branch of Altaic, Chukchi-Kamchadal, and Eskimo-Aleut. It also left strong traces in the Tungusic branch of Altaic and perhaps in Gilyak, while it completely lacks in the Turkic branch of Altaic and the whole Nippo-Koreo-Ainu family – as well as in Etruscan because of our poor knowledge of this language. Once we add this 2nd person *t-*, plus the different suffixes briefly mentioned above, it seems that there is no more room for doubt – except for Nippo-Koreo-Ainu and perhaps Gilyak: the 1st and 2nd person pronouns we have studied (and, consequently, the languages to which they belong) must descend from a common origin.

Among the different suppletive nominative roots of 1st person pronoun, the close resemblance between Indo-European *eg^{h}om* "I" and Chukchi-Kamchadal *(e-)ɣəm* "I" is striking. Together with the fact that these two families are the only ones to attest *tu* for the 2nd person, it might suggest a special historical relationship between Indo-European and Chukchi-Kamchadal.

A geographic observation may find its proper place here. The language families concerned are not randomly distributed over the Americas, Australia, and Africa. Rather, they constitute a geographic continuum, exactly as if they had resulted from an expansion over northern Eurasia. This expansion, we might add, must have started from a homeland which is most probably to be sought somewhere near the northern Pacific coast of Asia, according to the law of maximal diversity area. This law, long-known in biology, was discovered in linguistics by Greenberg (1972) in his controversy with Guthrie (1970) about the Bantu homeland.

The age of the Eurasiatic macrofamily is impossible to assess precisely, except that it is of course older than Indo-European, so that the 1st person pronoun *m-* was preserved in the overwhelming majority of the descendant Eurasiatic languages through at least some 10,000 years or more. It makes definitely clear that (1) calculations based on an average word replacement rate are highly uncertain, and that (2) particular words may be preserved through almost indefinite timespans. Regarding our claim that the globally distributed *papa, tata, mama, nana* or *kaka* kinship terms have been inherited from an original Proto-Sapiens language, a consequence of these two conclusions is to invalidate the objections levelled at their common ancestry on the ground of word replacement rates.

3. Kinship terms and the origin of 1st and 2nd person pronouns

Eurasiatic pronouns *m-* and *t-* are built from consonants that are also the root consonants of the Proto-Sapiens kinship terms *mama* and *tata.* It appears that the case is far from being isolated in the world's languages. A cursory count through Ruhlen's (1994b) global survey of 1st and 2nd person pronouns reveals that about three quar-

ters of them are based on one of the consonants *t, k, m, n, j, ŋ*, and *s*. The first five are the root consonants of the globally-distributed kinship terms *tata, kaka, mama, nana* and *jaja*. As to the last two, namely *ŋ* and *s,* they are also the base of widely, though not fully globally distributed kinship terms. In particular, the *s* is the root consonant of the widespread *ise* "father, father's brother, brother," present at least in the Eurasiatic, Amerind and Niger-Congo language families. No other consonantal type is widespread among 1st and 2nd person pronouns.

Viewed from the kinship terms' side, only the *p ~ b* oral labial stop of the highly widespread *papa ~ baba* is extremely rare among the world's 1st and 2nd person pronouns. (Among the few exceptions are the suppletive *b-* forms in Altaic 1st person pronouns, which are evidently derived from an original *m-* root.) The near absence of *p ~ b* among the thousands of these pronouns in the world's languages, while almost no language lacks an oral labial stop, is typologically highly unexpected. Given that kinship terms and personal pronouns are the two most conservative word categories, and consequently are in general extremely ancient in each language family, two questions arise. The first question is whether pronouns could be traced back to Proto-Sapiens through a global comparison, as was already done for globally-distributed kinship terms. Answering this question would be much more difficult than for kinship terms, for at first glance personal pronouns in the world's languages look like a horrible mess of *m* "I" and *m* "thou," *n* "thou" and *n* "we," and so on. We will thus leave the question open.

Answering the second question will perhaps help to answer the first one – one day. This second question is whether there could or must be a historical link between the category of personal pronouns and that of kinship terms. And our answer is yes.

Nursery kinship terms, as we argued elsewhere (Bancel & Matthey de l'Etang 2004, 2005; Matthey de l'Etang, Bancel & Bengtson 2006), are definitely of Proto-Sapiens ancestry – because of their ubiquity and the impossibility that they had resulted from convergent innovations. Contrary to a widespread belief, such innovations are unattested in the low-level language families with a written history (Matthey de l'Etang & Bancel this volume). Nursery kinship terms must even be much more ancient than Proto-Sapiens, and certainly played a major role in the apparition of Proto-Human – language with a simple, not double, articulation, i.e., with words built from phonemes but without sentences built from words (Bancel, Matthey de l'Etang & Ruhlen 2006). The first phonetically articulated words, uttered by mouths and tongues that had not been designed for speech by evolution, must have been built from the simplest consonants cast into the simplest syllable structures. They also must have carried highly useful meanings derived from pre-language social functions. They also must have been easy to transmit from generation to generation through mouths, brains and ears lacking specialisation for language, so that this invention did not get lost. All these conditions are fullfilled by nursery kinship terms

(Lieberman *et al.* 1972; Lieberman 1992; Bancel & Matthey de l'Etang 2005), and by them only.

Personal pronouns, in turn, must be of much more recent origin, for at least two compulsory reasons. The first reason is that, before the apparition of articulated sentences, they were simply useless. As long as articulate language merely consisted in isolated, holophrastic words, or, in a later stage, in parasyntactic sequences comprising a small number of juxtaposed words, to design and use specific words to refer to the speaker and the hearer in relation with verbs and actions was not only aimless, it was plainly unconceivable. The development of personal pronouns from the already existing vocabulary may only have taken place in a much later stage, once speakers had begun to gather words into sentences and the most frequent combinations had begun to crystallize into syntax.

The second reason is that personal pronouns are conceptually the most difficult words – much more than *quantum* or *phenomenology.* Not that we mean the *ego* is such an unfathomable mystery that only an Austrian genius could have invented it in the beginning of the XXth century (while the *tu* still keeps waiting). Rather, the special difficulty of personal pronouns resides in their WAY OF MEANING. In all languages, they are not the only words whose meaning depends on the speaker – after all, the English words *here* and *now* respectively mean "(close to) the place where I am" and "at the moment when I am speaking." But personal pronouns are the only words whose reference EXCLUSIVELY CONSISTS OF the speaker (for the 1st person) and the hearer (for the 2nd), so that their reference automatically switches as the speech turn passes in the course of conversation.

Does this specificity really make pronouns difficult words? The fact that children learn to use them in a relatively late stage of language acquisition (around 3 years), after they have learned most of the other basic syntactic structures, indicates that such is indeed the case. Even older children of 4 to 5 years, apparently mastering the specific referential mode of personal pronouns, are easy to confuse when an adult playfully acts as if pronouns were normal words, i.e., as if he himself was "an I" and the child was "a you," and "corrects" the child each time he uses the pronouns the right way.

For other words, the reference remains the same, whoever is speaking. If I tell you about the colorless dog furiously sleeping on the green carpet near the door, and you answer me this sleeping dog on the carpet is a degenerate subject, in both sentences – yours and mine – the words *dog* and *carpet* refer to the same beings, and the word *sleeping* refers to the same action. Most other words, and in particular all common and proper nouns, function the same way. And among the thousands of common nouns, it is also true of *quantum* and *phenomenology.* If the semantic content of these two words is certainly difficult, their semantic container is strictly trivial.

Kinship terms, in turn, refer in a specific way. In English, for instance, *father* is a common noun, but *dad* is and is not. If I ask you about *"your dad,"* it is a common

noun, but if I inform you that *"Dad is out fishing,"* it is a proper noun referring to a single person – a specificity aptly rendered, in the usual typography, by the use of an initial capital. You understand that *Dad* is in fact my father, and if you reply *"Oh! Dad went out early, they must have gone together,"* I understand in turn that you are speaking of your own father. Thus, kinship terms intrinsically share properties with all three nominal categories of proper nouns, common nouns, and pronouns. Like proper nouns, they can refer to a specific individual. Like common nouns, they also can refer to a class of beings, defined by common properties of these beings. And, like pronouns, they can switch reference from a given individual to another as the speech turn passes.

Now, at the global level, we find the following picture. All languages have kinship terms, common nouns, proper nouns, and personal pronouns. Among the four categories, kinship terms certainly are the oldest ones. Common and proper nouns are certainly very ancient, and have appeared in the course of the Proto-Human language evolution as natural developments from the primeval kinship terms. Finally, personal pronouns may not have appeared before a much later stage, when language verged on double articulation – i.e., on language in the modern sense, with sentences made of syntactically articulated words, as are all known languages. Facing this general picture, it appears that kinship terms and 1st and 2nd person pronouns share most of their root consonants at the global level. A straightforward, commonsense explanation would certainly be that both categories are based on the most frequent consonant types, and that their convergence is thus purely coincidental. This explanation, however, faces two problems. The first one is that both kinship terms and pronouns are extremely resistant to change, as we have seen with Indo-European and Eurasiatic pronouns, and thus have been, in most cases, inherited from ancient ancestral languages – a fact in and by itself hardly compatible with the idea of a random distribution of their constituent phonemes. The second one is the near absence, among the world's 1st and 2nd person pronouns, of oral labial stops *p* ~ *b.* A random distribution of pronouns could hardly have missed to use these very basic sounds. Conversely, the absence of *p* ~ *b,* if it could be regarded as a probabilistic artifact in a single language, would be very unlikely if pronouns originated in several distinct ancestor languages.

Another explanation is that there is indeed a relation between the two categories. The Proto-Human language certainly made do without pronouns for a long time. Even a fully (i.e., syntactically) articulate language can function without personal pronouns, for instance using either common or proper nouns to refer to the speaker and the hearer, like in sentences such as *"One-eyed hunter kill big mammoth at black cliffs for clan"* or *"Sweet Rob bring newspaper from office for Darling Janie."* However, all known languages have personal pronouns. At some point in the evolution of Proto-Human to modern languages, speakers must have evolved personal pronouns from already existing words. Which words? These precursors of pronouns must have been nominals, which in the preceding stage were the most used words designating the speaker and the hearer – words of the kind of *one-eyed hunter, clan, Rob,* or *Janie.*

However, if either common or proper nouns had given rise to pronouns, the global comparative picture of present-day pronouns would be very difficult to explain. If all modern pronouns shared a common origin, and had descended from a subset of common and/or proper nouns in a single ancestor language, how could one explain that it seems impossible to assign any of the modern pronominal root consonants to a global origin? It seems at odds with the exceptional preservation of pronouns we have dealt with above. And if present-day pronouns descended from a subset of proper or common nouns in several different original languages, how could one explain that they converge so massively towards a handful of root consonants, whatever the language family they belong to?

It happens that a third nominal category may serve to the same purpose – kinship terms. And they are by far the most likely category to have given rise to personal pronouns. In prehistoric conditions, all humans lived in small groups of kindred individuals – as is the case of all historically known hunting-gathering peoples as well as of our closest cousins, chimpanzees and bonobos. In the stage of Proto-Human language that preceded the apparition of pronouns, kinship terms such as *mama, tata, nana* or *kaka* may have been the most frequent way to address people, so that one of them might easily have given rise to a 2nd person pronoun. It may seem less straightforward for the 1st person pronoun, since by definition there is no kinship term referring to oneself. However, just like for 2nd person, the 1st person pronoun must have emerged from an earlier nominal used by the speaker to refer to himself, and no other nominal category possesses such a word. It is perfectly conceivable that, in the stage before the emergence of personal pronouns, speakers referred to themselves by the kinship term used towards them by the addressee. In modern languages with personal pronouns, such practice may seem weird, but is occasionally used when speaking with children who do not master the use of personal pronouns, as in *"Mum wants Sonny to eat up those peas."* It would explain why so many present-day languages have pronouns built from the same consonants as the most widespread kinship terms, and in our opinion it is the only way to explain this striking convergence consistently.

However, several crucial questions remain to be answered. How was performed the transition from the masculine~feminine meanings of kinship terms to the 1st ~ 2nd person meanings of pronouns? Why, contrary to kinship terms, do the pronominal root consonants of present-day languages not display any evident semantic convergence? And why are *p* ~ *b* consonants almost lacking as pronominal roots, while *papa* ~ *baba* kinship terms are extremely widespread?

To the first question, we do not have the faintest beginning of an answer. With regard to the two others, we only may make speculative suggestions. The apparent lack of semantic convergence for pronominal root consonants at the world level might be due to the fact that, at the time of the initial dispersal of our Sapiens ancestors, their language only had begun to develop the syntactic articulation – which gave them a fantastic technical impulsion, as suggested for the first time

by Ruhlen (1997). The category of pronouns was not yet definitely fixed, so that not exactly the same set of kinship terms crystallized into pronouns in the different language families.

Regarding the lack of *p* ~ *b* consonants in 1st and 2nd person pronouns, one of our present puzzles regarding global kinship terms might perhaps give us a solution. Apparently, different Proto-Sapiens etymologies referred to the same kinship classes. We are able to reach a meaning like "male elder on the father's side" for both *papa* and *tata* in Proto-Sapiens, on the basis of more specialized meanings in modern languages, such as "father," "father's brother," "brother" and "grandfather." Taken together, these four meanings represent more than 70% of languages where *papa* (or *tata*) forms are attested, while the remaining 30% of languages use their *papa* or *tata* forms with many different meanings with no statistical significance (Bancel & Matthey de l'Etang 2004; Matthey de l'Etang & Bancel 2005). Thus, *papa* and *tata* appear as synonymous, and so do the three globally-distributed *mama, nana* and *jaja* "female elder on the mother's side." But, even though the difference between them escapes reconstruction at present, they certainly were not true synonyms – otherwise they would not have all survived. The widespread distinction in modern languages between "my father" and "your father," and between "my mother" and "your mother," sometimes with affixes but also often with two distinct terms, could be a trace of this putative original difference. Whatever this difference may have been, *papa* may have had a specificity that led to exclude it when referring to both the speaker and the hearer, so that it would not have given rise to a 1st nor 2nd person pronoun. This would explain the strange scarcity of oral labial consonants among the variety of pronominal roots in modern languages.

Finally, our conjecture may be condensed as follows:

Given that

(1) kinship terms must be of the highest antiquity in the development of articulate language, while pronouns may have appeared only at a much later stage;

(2) pronouns must have emerged from nominals frequently used to refer to the speaker and the hearer; and

(3) the pronominal root consonants at the global level are a subset of those of kinship terms,

It is difficult to see how personal pronouns could have emerged otherwise than from simplified kinship terms formerly used to represent the speaker and the hearer in discourse.

Conclusion

We have shown that resistant words such as the Indo-European and Eurasiatic pronouns may last for almost indefinite timespans. Their persistence, paralleled in the overwhelming majority of language families, gives credibility to the possibility that kinship terms might have lasted a hundred millennia, or more. We have then put forward the conjecture that there could be a relationship between kinship terms and 1st and 2nd person pronouns. Even if this conjecture should remain unproven and unrefuted for a while, it shows that we have linguistic means to investigate the evolution of human language from its earliest stages.

References

Bancel, P.J. & Matthey de l'Etang, A. 2002. "Tracing the Ancestral Kinship System: The Global Etymon KAKA. Part I: A Linguistic Study." *Mother Tongue* **7:** 209–243.

Bancel, P.J. & Matthey de l'Etang, A. 2004. "Kin Terms *(P)APA, (T)ATA, (M)AMA, (K)AKA* and the Origin of Articulate Language." Paper presented at the Language Origins Society 18th Annual Meeting, Oxford (Eng.), 9–10 July 2004. Appeared in 2006 in a special issue of *Cognitive Systems*, Groningen.

Bancel, P.J. & Matthey de l'Etang, A. 2005. "Kin Tongue. A Study of Kin Nursery Terms in Relation to Language Acquisition, with a Historical and Evolutionary Perspective." *Mother Tongue* **9:** 171–190.

Bancel, P.J., Matthey de l'Etang, A. & Ruhlen, M. 2006. "The Proto-Human Words PAPA and MAMA and the Origin of Articulate Language." Paper presented at the 105th American Anthropological Association Annual Meeting, to appear in the proceedings of the Per Hage Memorial Session.

Bengtson, J.D. & Ruhlen, M. 1994. "Global Etymologies." In Ruhlen 1994b: 277–336.

Benveniste, E. 1946. "Structure des relations de personne dans le verbe." *Bulletin de la Société de Linguistique de Paris* **43**–1. Repr. in *Problèmes de linguistique générale* 1: 225–236, Paris: Gallimard.

Bomhard, A.R. 2007. *Reconstructing Proto-Nostratic. Comparative Phonology, Morphology, and Vocabulary*. Charleston: Signum.

Dolgopolsky, A. 1964. "Gipoteza drevnejšego rodstva jazykovyx semej severnoj Evrazii s verojatnostnoj točki zrenija." *Voprosy Jazykoznanija* **2:** 53–63. [Translated into English as "A Probabilistic Hypothesis Concerning the Oldest Relationships among Language Families of Northern Eurasia." *Typology, Relationship and Time*, Ed. by V.V. Shevoroshkin and T.L. Markey, 27–50. Ann Arbor: Karoma. 1986.]

Greenberg, J.H. 1972. "Linguistic Evidence Regarding Bantu Origins." *The Journal of African History* **13**–2: 189–216.

Greenberg, J.H. 2000. *Indo-European and Its Closest Relatives. The Eurasiatic Language Family.* Vol. I: *Grammar*. Stanford, Calif.: Stanford University Press.

Greenberg, J.H. 2002. Indo-European and Its Closest Relatives. *The Eurasiatic Language Family.* Vol. II: *Lexicon.* Stanford, Calif.: Stanford University Press.

Guthrie, Malcolm. 1970. "Contribution from Comparative Bantu Studies to the Prehistory of Africa." *Language and History in Africa*, Ed. by D.E. Dalby, 20–49. London: Frank Cass.

Lieberman, P. 1992. "On the Evolution of Human Language." *The Evolution of Human Languages,* Ed. by J.A. Hawkins & M. Gell-Mann, 21–47. Redwood City, Calif.: Addison-Wesley.

Lieberman, P., Crelin, E.S. & Klatt, D. 1972. "Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man and Chimpanzee." *American Anthropologist* **84:** 287-307.

Matthey de l'Etang, A. & Bancel, P.J. 2005. "The Global Distribution of (P)APA and (T)ATA and their Original Meaning." *Mother Tongue* **9:** 133–169.

Matthey de l'Etang, A. & Bancel, P.J. 2008. "The Age of MAMA and PAPA." This volume.

Matthey de l'Etang, A., Bancel, P.J. & Bengtson, J.D. 2006. "The Proto-Sapiens Words PAPA, MAMA and KAKA and the Origin of Kinship Systems." Paper presented at the 105th American Anthropological Association Annual Meeting, to appear in the proceedings of the Per Hage Memorial Session.

Pagel, M. 2000. "Maximum Likelihood Models for Glottochronology and for Reconstructing Linguistic Phylogenies." C. Renfrew, A. McMahon & Larry Trask (Eds), *Time Depth in Historical Linguistics*, Cambridge (Eng.): McDonald Institute for Archaeological Research I: 189–207.

Rosenfelder, M. No date. "Numerals in 5,000 languages." http://www.zompist.com/numbers.shtml.

Ruhlen, M. 1991. *A Guide to the World's Languages,* vol. I: *Classification*. Stanford, Calif.: Stanford University Press.

Ruhlen, M. 1994a. *The Origin of Language. Tracing the Evolution of the Mother Tongue*. New York: John Wiley.

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

Ruhlen, M. 1997. Public conference given in Lyons for the appearance of the French translation of Ruhlen (1994a).

Ruhlen M. 2007. Post-Script to the 2nd edition of *L'Origine des langues. Sur les traces de la langue mère* [Fr. tr. of Ruhlen (1994*a*)]. Paris: Gallimard.

Starostin, S.A., Dybo, A.V. & Mudrak, O.A. 2003. *Etymological Dictionary of the Altaic Languages.* Leiden: Brill.

Starostin, S.A., Starostin, G.S., Nikolayev, S.F., Dybo, A.V., Mudrak, O.A. & Golovastikov, A. No date. "Nostratic Etymological Database (including Indo-European, Altaic, Uralic, Kartvelian, Dravidian, Chukchee-Kamtchatkan, and Eskimo Etymological Databases). Online at: http://starling.rinet.ru/index2.php?lan=en.

Strand, R. 1997–2008. *Nuristân. Hidden Land of the Hindu-Kush.* http://users.sedona.net/~strand/.

Swadesh, M. 1952. "Lexico-Statistic Dating of Prehistoric Ethnic Contacts, with Special Reference to North American Indians and Eskimos." *Proceedings of the American Philosophical Society* **96** 4: 452–463.

Swadesh, M. 1954. "Time Depths of American Linguistic Groupings." *American Anthropologist* **56** 3: 361–377.

Thomason, S.G. & Everett, D.L. 2005. "Pronoun Borrowing." *Proceedings of the Berkeley Linguistics Society 27th Annual Meeting* 301–315. Berkeley: Berkeley Linguistics Society. Online version http://www-personal.umich.edu/~thomason/papers/pronborr.pdf.

General index

A
Africa
  Black Africa 35
  East Africa 10, 13, 149
  Ethiopia 11, 49, 138, 141, 169, 353, 377
  Guinea Coast 25, 28–29, 34
  Nile Valley 9–13, 19–21
  North Africa 9, 17–22, 26, 28, 33–35
  Sahara 11–12, 19–20, 28, 35
  Sahel 28
  South Africa 353
  sub-Saharan Africa 266, 282
  Sudan 161, 170, 190
  supra-Saharan Africa 3, 4, 11, 13
  West Africa 19, 21, 25, 28, 33
Americas 22, 211, 233, 245, 263, 266, 268, 282, 309–316, 321–331, 351, 457
Andamans 201, 287, 289, 292, 307
anthropology 3, 17, 20, 25, 288, 343, 373
  arch(a)eology 8, 233, 237, 282, 327, 343, 373–378, 410, 412
  diachronic anthropology 17, 20, 25
  ethnography 17, 324, 343
  holistic anthropology (four field approach) 3, 13, 17, 25, 322, 343
  linguistics (see linguistics)
  paleoanthropology 343, 347
  paleontology 223, 227–228, 282, 343, 346
  physical (biological) anthropology 3, 17, 288, 324, 343, 381, 386–389
arch(a)eology (see anthropology)
archetype (see comparative mythology)
astrologers, astrology 26–29, 31
Australia 49, 258, 266, 282, 287, 289, 291–292, 300, 302–303, 306, 347, 354, 387, 457

B
babbling 408, 417–418, 433
Babylonia 25–26, 28
basic lexicon, basic vocabulary 57, 67, 153, 224–226, 257–258
Berbers 9, 13
brain (human) 344–345, 350–353, 381–382, 389–390, 392–396, 407–408, 410–412, 458
  Broca's area 345, 353
  Wernicke's area 353

C
chimpanzee, chimp 343–348, 374, 409–410, 461
Christian 19–21, 27–28, 272, 276
cognation 153, 173, 178, 180
common origin (linguistic) 223, 245, 417, 457, 461
comparative linguistics, comparative method (see linguistics)
comparative mythology 263, 265, 282–283
  archetype 264, 266
consonant(s) 63–64, 149–158, 154, 155–159, 163, 166, 178–179, 205, 218, 247–248, 288, 314–315, 317, 363–364, 367, 418, 423, 434, 440, 446, 452–453, 456–458, 460–462
convergence, convergent 189–190, 200, 202, 234, 363, 375, 417–418, 425, 433, 447, 458, 460–461
culture
  Acheulian 374
  biocultural change 12–13, 18, 20, 352
  culture-complex 25–26, 28, 33–34
  Levallois 374–377
  Natufian 19–22
  Paleo-Indian 334–336

D
dental morphology 17, 19–22
diffusion 12, 18, 26–28, 190, 204, 264–266, 332, 373
dragon, dragon-slayer (see myth)

E
East Asia 292, 348, 374
Egypt, Egyptians 8–10, 12, 33, 153, 266, 276
EHL (Evolution of Human Language Project) 244, 246, 255, 422–423, 428–432
elbow (in African languages) 99, 113, 169–178, 180
Ethiopia 11, 49, 138, 141, 169, 353, 377
ethnogenesis 17–18, 22, 245, 255, 347–348
  African Hybridization and Replacement Model 347
  African Replacement Model 347–348
  Great Leap Forward Model 354
  Assimilation Model 347
  Multiregional Evolution Model 347

Out of Africa, Out-of-Africa Model 13, 264–265, 289, 347–348, 373, 376
Eurasia, Eurasian 11–13, 19–20, 189–190, 201, 204, 211, 223, 228, 230–231, 234–236, 241–243, 245, 247–248, 255–258, 263, 265–267, 271, 273–274, 276, 282, 343, 347, 349, 351–352, 373, 375–377, 457
evolution (cultural) 3, 12, 19–20, 223, 231–232, 236, 330, 332, 335, 354, 387–392
evolution (linguistic) 149–150, 152, 157, 177, 223, 225, 231–232, 243, 259, 282, 288, 381–384, 386–387, 389–392, 401–406, 409, 411–416, 424, 441, 446, 453–454, 458, 460, 463
evolution (physical) 3, 18, 20, 343, 345–347, 360, 362–363, 368–369

F

Far East 266
fossil(s) (biological), fossil record 346–347, 353, 359, 368–369, 374, 377
fossil(s), fossilized (linguistic) 249, 289, 419

G

gender (grammatical) 39–43, 45–47, 248–249, 253, 290
  animate 39, 41, 47, 249, 252–253
  inanimate 39, 41, 43–44, 249, 252–253
gene, genes, gene flow 4, 6, 13, 18, 21, 233, 235, 237, 323, 343–352, 377, 388–389, 392, 411, 415
genetics 12, 20, 25, 233, 265, 282, 288, 343, 346–349, 376
  ancient DNA 348, 350
  ASPM locus 352
  autosomal, autosomal systems 4, 344, 348–349
  beta-catenin 352
  biallelic markers 5
chromosome 3–6, 10–14, 344, 346, 348–349
  chromosomal rearrangements 344
  X-chromosome 4, 344, 348
  Y-chromosome 3–6, 10–14, 344, 348–349
coalescent times 348
deletion 344
DNA 4, 6, 11–13, 323–324, 344–345, 347–350, 353, 388, 390–391, 412
effective population size 348
European starlings 354
FOX P2 352
genetic discontinuity 20, 22
Genographic Project 344, 349, 354
genome 3, 343–344, 346, 348–350
haplogroup A 7, 14
haplogroup B 7, 14
haplogroups 5, 9–10, 14
haploid systems 344, 48–349
HapMap 344, 350
hybridization 343, 345–347, 349
Hypervariable Region I 348
insertion, insertion/deletion events 344
International Haplotype Map 344
LMO4 gene 352
M60 7–10, 14
M35/78 3, 10–12
M35/81 3
M81 8–10, 12–13
M89 3, 7–11, 14
M91 7–10, 14
MCPH1 locus 351, 354
Mean Measure of Divergence 20–21
melanosome 351
microcephalin 351
mitochondrial DNA, mtDNA 11–13, 41, 344, 347–349, 353
non-recombining 4, 348
non-Yap 3
NRY 4–5
nucleotide substitutions 344
PN2 3, 7–10, 13–14
PN2/215/M35 3
pseudo-autosomal 4
RRM2P4 348–349
selective sweep 6, 350–351
SVT 360–361, 363–369
time to the most recent common ancestor, TMRCA 348, 353
UEP 4–5
underived 5, 7–10, 12–13
unique event polymorphisms 4
Yap 3, 14
geological epochs
  Holocene 4–5, 9–13, 19–23, 383
  Pleistocene 11–12, 17–22, 369, 373–375
gestures 353, 363, 366–367, 373–374
Ghana 28–29, 34–36
global etymology (transphylic root, transphylic lexeme, world etymology) 173–174, 180, 190–191, 194–195, 204, 417–435
glottochronology 57, 65, 223–224, 226–227
gold 29–30, 34–35
Gondwana (mythology) 264–265, 282
grammar (see morphology)
grammatical gender (see gender)

H

historical comparative mythology (see comparative mythology)
historical linguistics (see comparative linguistics)
hominid 288, 347, 363, 367–369, 381–382, 386, 393–396, 407–409
  hominid language ability 409
hominins 343–349
  Ardipithecus kadabba 346
  Orrorin tugunensis 346
  Sahelanthropus tchadensis 346

Homo 344, 347, 349–350, 369, 401, 410, 415
Homo erectus 349, 410
Homo ergaster 369
Homo habilis 410
Homo neanderthalensis 348, 359, 368–369, 410, 412
Homo sapiens 344, 347, 349–350, 415
humans 263, 265, 267, 277, 282, 288, 344–351, 354, 359, 367–369, 373–376, 389, 391, 394, 401, 404–405, 408, 410–415, 461
fully (modern) (human) 359–360, 369, 395
human/chimpanzee genomic differences 345
human genome 3, 350
human tongue 361, 363, 368–369
Humpty Dumpty 395–396
hyoid bone 369

I
India 263, 266–268, 271, 274, 276–278, 280, 282, 289, 291, 302, 425
Indra 267, 269–272, 276–278, 280–282
inheritance (genetic) 13, 20, 344
inheritance (linguistic) 41, 57, 118, 138, 151, 154, 257–258, 386, 388, 419–426, 428, 433, 439, 446–448, 457, 460
innovation (linguistic) 40, 159, 213, 223–225, 228–232, 236, 249, 332, 335, 401–402, 408–409, 421–422, 427, 429, 458
Islam, Islamic 8, 11–12, 28, 34

J
Jupiter 26–27, 31–33

K
kinship, kinship term(inology) 227, 290, 344, 415, 417, 419–420, 425–426, 429–433, 439–440, 457–463

L
language diversification 189–190, 241, 245, 255–256
Laurasian (mythology) 263–266, 282
Levant 9–10, 12, 18–22, 349, 375–377
lexical evidence, lexical reconstruction 228–229, 448
address form, address term 421, 424, 427–429
basic lexicon, basic vocabulary 57, 67, 153, 224–226, 257
cardinal numbers 213–219
chance resemblance 195, 258, 314
global etymology (transphylic root, transphylic lexeme, world etymology) 173–174, 180, 190–191, 194–195, 204, 417–435
kinship term(inology) 227, 290, 344, 415, 417, 419–420, 425–426, 429–433, 439–440, 457–463
lexeme 163, 190, 194, 200, 204
mama/papa words 293, 297, 417–424, 429–430, 432–435, 439, 458, 461–462
nursery forms, nursery words 417–420, 422, 426, 429, 431–434
phonaesthemes 189–190, 204
reference terms 421
linguistic change 223–25, 227, 233, 406, 418–419, 426, 431–433
apocopation 161, 166
attrition 161, 164, 393
clipping (linguistic) 163
deletion 153, 155
erosion 161, 163–166, 419, 444, 447
Klingenheben's law 163
loan (linguistic), loanwords 171–172, 199, 217–218, 309, 311–312, 329, 339
phonological evolution 149
syncopation 161, 164, 166
word replacement 440, 457
linguistics (comparative, historical), comparative method 12, 155, 166, 190, 204, 223, 226, 231, 234, 236, 255, 258, 265, 310, 381, 422
linguistic history 161, 164, 166
linguistic paleontology 223, 227–228, 343
linguistic prehistory 233

M
macrofamily, macrophylum, megaphylum (linguistic) 57, 94, 139, 189–190, 241–245, 251, 255–256, 258, 432, 439, 449, 456–457
mama/papa words 293, 297, 417–424, 429–430, 432–435, 439, 458, 461–462
marriage of sun and moon (myth) 263, 277
Mars 26–27, 32
Mediterranean Basin 25–26, 33
meme (isofunctional neuroanatomical unit) 382, 387, 389–392
Mercury (god, planet) 26–27, 31–33
Middle East 21, 352
mime (unit of imitation) 382, 390–391
Moon 25–27, 32, 263, 267–282, 296
morphology (linguistic), grammar (see also gender)
absolutive 135, 250
accusative 42, 440–445, 450, 453
agreement-marking 43
copula 39–48, 454
ergative 290
focus marker 39–40, 46–47, 49

nominative 42, 216, 424, 426, 440–445, 447–453, 456–457
shared irregularities 257
subject-agreement 39, 43
vocative 424, 426, 428, 431
Mother Tongue (original human language) 381
Mother Tongue (journal, newsletter) 190, 236, 243, 265, 287, 321, 353
mutation 3–4, 6–7, 12–14, 323, 344, 351–352, 363, 368, 373, 411
myth, mythology 263–283
Ama.no Kami 267
Amaterasu 280, 282
Apollo 273, 278
Asuras 267
Beowulf 267, 272
cow-herd 279
Dawn (Ushas) 268, 277–278, 280–282
dragon, dragon-slayer 263, 265–278, 282
Geryoneus 277–278
Herakles 267, 277–278
Kərəsāspa 270
Kuni.no Kami 267
Marduk 268, 273, 276, 278
moʻo (Hawai'i) 268, 276
Moon 263, 267–282, 296
Naga (serpent) 266, 271–272, 277–278
Nüwa (Nügua) 274
Pururavas 277, 281–282
python 276
reptile 257, 263, 267, 269–271
Siegfried / Sigurd 267, 272
summer solstice 263–264, 277–278, 280–282
Sun 263, 267–268, 273, 275, 277–282
Thraētaona 267, 269–270, 277
Tiamat 268, 273–274, 276
Titans 267
Urvasi 277, 281
Ushas (Dawn) 280–281
Vṛtra / Vərəθra 269–271
weaver girl 279–282
Yamata.no orochi 267–268, 272

N

narration, narrative, story line 265, 391
Native Americans 20–22, 266, 309–318, 321–339, 421, 458
natural selection 5–6, 20, 343–345, 350–351, 353, 384, 387–389, 391, 412–413
Neandert(h)al 348, 359, 368–369, 410, 412
Near East 3–4, 7–12, 17–18, 20, 266–267, 329
New Guinea 287, 289, 292, 293, 306
North America 235, 242–243, 266, 309, 324, 331, 334–336, 338, 376
Northern Eurasia 228, 231, 235–236, 241–243, 245, 247, 255–258, 457
nursery forms, nursery words (see lexical evidence)

O

Oceania 211, 289
Out of Africa, Out-of-Africa (Model) 13, 264–265, 289, 347–348, 373, 376

P

paleoanthropology (see anthropology)
Pan-African (lexical roots) 189–190, 192, 200, 203–204
crab 200–203
dog 199–200, 204
fly (verb) 200, 202–203
skin (Pan-African root) 190, 196–197, 203
tortoise (Pan-African root) 190, 195–196

P

papa (see mama/papa words)
Papua 287, 289, 291
phonemes 149, 246, 353, 367, 391–392, 458, 460
phonology 246, 253, 256–258, 359, 407
planetary week 25–36
populations 3–13, 17–22, 231, 233, 235, 292, 303, 316, 321–339, 343–354, 373–378, 390, 408
population movement, migration 4–5, 8–9, 11–12, 17–21, 234, 291, 302,327, 329, 334–336, 338, 349
population replacement 17, 22
prehistory, prehistoric periods
Bronze Age 269
Iron Age 223, 229, 231
Late Paleolithic 263, 282
Mesolithic 20, 228, 230–232, 237, 291, 302
Middle Paleolithic 354, 369, 374–375, 377
Neolithic 8, 11–12, 18–19, 223, 228, 230–232, 235, 237, 280, 291, 302, 309, 332, 336, 350, 374
Paleolithic 12–13, 20, 228, 232–233, 237, 263, 265, 282, 351, 354, 359, 369, 373–375, 377
Upper Paleolithic 13, 351, 354, 359, 369, 373–375, 377
primates 343, 345, 352, 360–362, 367–368, 395, 409
non-human primates 345, 360, 362, 368
chimpanzee, chimp 343–348, 374, 409–410, 461
gorilla 345, 423
Pan paniscus 409
Pan troglodytes 348, 409
pronouns 39, 42, 44, 48, 138, 242, 249–251, 253, 257, 288, 290–291, 402, 439–440, 443–463
"proof" (scientific) 241, 244, 255
proto-language 149–150, 191, 223–224, 226, 228–230, 232–233, 237, 249, 253, 258, 309, 312–313, 323–324, 330, 442

R

riddles 49–51
Roman Empire, Rome 26–29, 33, 267, 276, 424

roots (lexical) 152–153, 156, 159, 163, 165, 169, 171–173, 175–177, 189–192, 198, 200–201, 203–204, 228, 232–233, 288, 385, 432, 439–440, 442–443, 448, 457, 461–462

S
Saturn 26–27, 31–33, 35
science, scientific method, scientist 17, 29, 191, 376, 384, 410, 413
  abduction, abductive 255, 257
  absolutist 244
  best explanation 241, 244–245, 255–258
  fruitful (hypothesis) 139, 241, 245, 253, 257, 259
  negative evidence 245
  "proof" (scientific) 232, 241, 244, 255, 417
  relativist 244
skin color, pigmentation 350–351
South America 266, 282, 321, 324, 326–28, 330, 336
South Asia 289
South India 292, 302
speciation, species 343–347, 366
speech 8, 13, 45, 227–228, 234–235, 309, 345, 352–353, 359–369, 405, 410, 418, 458–460
  acoustic frequencies 361
  babbling 408, 417–418, 433
  formant frequency 360–361, 363, 365–67
  fundamental phonation 360
  laryngeal descent, larynx 353, 359–360, 366–369, 410, 412
  neural mechanisms 366
  neural substrate 359
  pharynx 361, 368–369
  phonation 359–360
  phonetic ability 359
  quantal vowels 361–363, 367, 369
  speech anatomy 359, 369
  speech community 13, 227–228, 235, 309
  speech perception 363–364
  speech physiology 360
  speech production 360, 363
stone tools 373–374, 377, 410
  Acheulian 374
  core and flake 374–375
  Levallois 374–377
suppletion, suppletive 172, 250–251, 257, 447, 449–451, 456, 458
Symbiotic Theory 381–382, 384, 386, 394
  symbiome, symbiomism 381, 395–396
  symbiosism, symbiotic 381–382, 384–386, 394–396
syntax 236, 264, 353–354, 359, 382, 391, 395, 408, 447, 459–461
  cleft construction 47
  factual statement 46
  focus 39–40, 46–47, 49
  identificational clause 42

T
taboo 276, 415, 423, 447
Tasmania 258, 287, 289, 306
tongue, tongue position 360–363, 366–369, 381
Tower of Babel (EHL) 244, 246
transphylic root, transphylic lexeme (see global etymology)

V
Venus 26, 31–33, 52
vocal cords, vocal tract 353, 359–360, 365–366, 369
  supralaryngeal vocal tract 360
vocalization 362, 366
vowel(s) 40–41, 50, 150, 158, 163, 177, 198, 203, 206, 215, 246–247, 288, 313, 317, 359–367, 369, 417–418, 444, 452, 454–455

W
week 25–36
Western Eurasia 373, 375–376

Index of languages and language families

A

Afroasiatic (Afro-Asiatic, Afrasian) 3–5, 10–13, 17, 18, 22, 57, 94, 138–139, 149, 161, 190, 200, 204, 241–243, 247–249, 256, 258, 290–291, 302, 432–433
  non-African origin of 18
Ainu 231, 236, 242, 259, 448–449, 452–453, 456–457
Almosan 321, 324, 338–339
Almosan-Keresiouan 324, 338–339
Altaic 173, 201, 223, 232, 236, 241–242, 249, 423, 432–433, 447–458
Amerind 22, 245, 316–317, 321–324, 330, 332, 334, 336–338, 421, 458
Andamanese 201, 287–293, 299, 302–303
Andamanese-Papuan-Australian 289–290, 292–293, 303
Andean 321, 324, 326, 328–331
Arabic 5, 9, 18, 34–35, 234, 329, 430–431, 435
Armenian 214, 234, 248, 272, 440–446
Australian 22, 258, 289–293, 300, 302–303
Austronesian 201, 231, 292

B

Bantu 7, 108, 137, 191, 194–196, 200, 205, 291, 302, 423, 432–433, 457
Bantu-Sudanese 291
Basque 201, 241–243, 246–249, 251–256, 259–260, 290, 434–435
Beja 5, 151, 153–155, 157, 159, 196–197, 200, 202
Bengali 425–426, 441, 444
Berber 3, 5–6, 9–13, 27, 34, 193, 196, 203, 248, 432
Berta 169, 171–177, 180, 193, 194, 202
Birale (see Ongota)
Boreafrasian 11
Boreal, Borean 246, 256
Burushaski 242–244, 246–248, 250–255, 259–260

C

Caucasian 203, 213, 215–216, 232, 241–259, 432
Celtic 231, 420, 441–446
Central Amerind 316–317, 321, 324, 330, 332, 334, 336, 338
Chadic 3, 5, 11, 161–166, 193, 195–201, 203, 205, 247–248, 432
Chibchan 321, 324, 326, 328, 330, 334
Chibchan-Paezan 324, 326, 328, 330
Chinese 203, 229–231, 242, 247, 251, 253, 255, 256, 264, 267–268, 274–275, 279–280, 430, 433, 435
Chukchee-Kamchatkan, Chukchi-Kamchatkan 241–242, 256
Coptic 5, 137, 139
Cushitic 3, 5–8, 10–11, 39–40, 49, 57, 149–153, 155, 157–159, 193–194, 196–200, 202–203, 205, 247, 432

D

Dahalo 149–153, 159, 193–194
Dene-Caucasian 241–248, 250–53, 255–259
Dravidian 174, 196, 201, 203, 241–242, 287, 291–292, 302–303, 307, 419, 427–428, 432–433

E

Egyptian 3, 5–6, 8, 10, 11–131, 247–248, 268, 278
English 27, 31–33, 179, 194, 203, 225, 229, 323, 365–366, 376, 420, 433, 441, 443–448, 459
Equatorial 321, 324, 326–331
Equatorial-Tucanoan 324, 326–28, 330
Eskimo-Aleut 173, 228, 242, 256, 338, 432, 448–449, 455–57
Eurasiatic, Eurasian 173, 236, 241–243, 245–253, 255–259, 288, 432–433, 439–440, 448–449, 452–458, 460, 463

F

French 27, 164, 254, 384, 424–425, 434–435, 441–447

G

Ge-Pano 321, 324, 327–328, 330
Ge-Pano-Carib 324, 327–328, 330
German 26, 31–32, 164, 214–215, 217, 242, 252, 257, 287–288, 306, 384–386, 388, 434, 440–441, 444–446, 448
Germanic 27, 31–33, 164, 217, 224–225, 231, 257, 267, 272, 420, 441, 443–446
gesture language, gestures (see also sign language(s)) 353, 363, 366–367, 373–374
Gidole 49–51
Greek 25–27, 33, 139, 201–202, 214–215, 217–218, 234, 247–248, 252, 267, 272–273, 278, 385, 423–424, 433, 435, 440–446

H
Hamitosemitic, Hamito-Semitic (see Afroasiatic)
Hausa 162–166, 196, 201, 203–204
Hindi 425–426, 440–446
Hittite 213–215, 218, 229, 231, 247–248, 252, 440–442, 445–446
Hokan 321, 324, 330, 334, 335–337
hominid language 381–396, 407–409

I
Indic 123, 214, 423, 425–428, 433, 441–443, 445–446
Indo-Aryan 231, 271, 425–426, 428
Indo-Chinese (see Sino-Tibetan)
Indo-European 123, 173, 201–203, 213–218, 223, 229–231, 233–234, 237, 241, 249, 252, 257–258, 265, 267, 269, 272, 288, 290, 383, 419–421, 423, 427, 432–433, 439–451, 453–457, 460, 463
Indo-Hittite 215, 229, 231
Indo-Iranian 218, 270–272, 423, 425
Indo-Pacific 243, 258, 287, 291–292
Iranian 203, 218, 231, 234, 267, 269–272, 423, 425, 441–442, 444–446
Italian 164, 287, 421, 424, 434–435, 441, 445–446
Italic 231, 423, 441–442, 444–446

J
Japanese 173, 201, 225, 229–230, 234–235, 242, 247–248, 252, 264, 266–269, 272, 277, 280, 390, 447, 449, 452–453

K
Kado, Kadu 93, 169–170, 172, 179–180, 194–195, 198–199, 206
Keresiouan 321, 324, 338–339
Ket 131, 248, 250–253
Khoisan 7, 13, 98, 190, 199, 288, 353–354
Kiowa-Tanoan 316–317
Korean 201, 230–232, 242, 247, 252, 429, 435, 447, 449, 452–455
Kunama 131, 169–177, 180, 194, 196
Kusunda 201, 243, 258–259, 293

L
Latin 18, 27, 34–35, 164, 202, 213–214, 218, 231, 247, 252, 290, 420–421, 424, 425, 433–435, 440–446, 453, 455
Lithuanian 247, 252, 267, 444–446

M
Macro-Carib 321, 326–328, 331
Malayo-Polynesian 289–290, 292–293
Manchu-Tungus (see Tungusic)
Mayan 318, 338
Mixe-Zoquean 317–318
Mongolic 224–225, 231, 423, 429, 447, 449–452, 457
Mon-Khmer 201, 203, 289, 292
Mosan 321, 337–338
Mother Tongue 381
Munda 291, 304–305
Munda-Polynesian 291

N
Na-Dene 22, 242–243, 246, 248–249, 251–255, 268, 331–332, 337–338
Nicaraguan sign languages 403
Niger-Congo 13, 189–191, 197–198, 201–202, 204–206, 257–258, 291, 387, 458
Niger-Saharan 189–191
Nilo-Saharan 13, 57, 109, 112, 131–133, 135, 138–139, 155, 169–170, 180, 189, 194, 197, 203, 205–206
Nippo-Koreo-Ainu 448–449, 452–453, 456–457
Nostratic 174, 232, 236, 241–243, 247, 249, 256, 432

O
Oceanic 289
Omotic 3, 5, 11, 39–40, 47, 57–139, 141, 193–194, 197–198, 200, 205, 248
Ongota 5, 115, 125, 127, 129, 132
Oromo 7, 50, 139, 153, 198
Otomanguean 309, 311–318

P
Paezan 321, 324, 326, 328–332
Palaeo-Siberian 223
Paleo-African 13
Pali 426, 434–435, 441
Penutian 317, 321, 324, 330–331, 334, 336–339
Persian 203, 214, 248
Prakrit 425–426, 428, 435
"primitive" language 288, 359, 395, 401
proto-chimpanzee 345, 347
proto-hominin 345
Proto-Human 458, 460–461
Proto-Sapiens 341, 434, 439–440, 457–458, 462
Proto-World 417, 434

R
Romance 27, 224–225, 231, 421, 442–443
Romanian, Rumanian 242, 421, 424, 434–435, 440–441, 443, 445–446
Russian 17, 242–243, 247, 386, 423, 440–441, 444–446, 449

S
Saami 247, 258
Sanskrit 213–214, 216, 252, 254, 425–428, 433–434, 440–446
"Scythian" (macrofamily) 241
Semitic 3–12, 18, 22, 195, 197–199, 218, 231, 247–248, 288, 290–291, 302, 430–433
Shabo 169–170, 178–180, 194
sign language(s) (see also gesture language) 401, 403, 406, 409
Sino-Caucasian 232, 241–242
Sino-Dene 242–243, 253
Sino-Japanese 225

Sino-Tibetan 201, 203, 231, 242–243, 248–249, 251–253, 255
Slavic 99, 123, 214, 217, 231, 267, 421, 423, 440–446
Somali 7
Songay, Songhay 127, 169, 170–172, 174–177, 180–181, 193, 195, 197–199, 202, 205

T
Tamil 196, 203, 291, 302–304, 427–428, 435
Tasmanian 289–290, 292
Tibetan 201, 203, 231, 242–243, 248–249, 251–253, 255–256, 429, 433, 435
Tibeto-Burman 201, 251, 429, 432
Tocharian, Tokharian 231, 252, 440–446
Tucanoan 321, 324, 326–328, 330
Tungusic 231, 241, 250, 423, 447, 449–450, 452, 454, 457
Turkic 229, 231, 241, 248, 250, 419, 422–423, 433–435, 447, 449–453, 455, 457
Turkish 234, 247–248, 252, 422–424, 433, 450–451

U
Ural-Altaic 223, 236
Uralic 173–174, 203, 223, 228–233, 237, 241–242, 249, 258, 288, 432, 448–451, 453–457
Uralic-Yukaghir 432, 448–449, 456–457
Uto-Aztecan 309–317, 334

W
Welsh 27, 420, 433–435, 441, 444–445
West Sudanic 190, 206

Y
Yeniseian 242–243, 248–253
Yukaghir 231, 242, 247, 252, 256, 432, 448–449, 453, 456–457
Yupik 247, 455

Z
Zargulla, Zergula 39–48, 59

Index of scholars discussed

B
Bellwood, Peter 309, 310, 373
Bender, M[arvin] L[ionel] 58, 60, 63, 94, 123, 172–173, 175, 177–178, 180, 191, 193–195, 197–200, 202–203, 205
Benedict, Paul K. 429, 432
Bengtson, John D. 173–174, 190, 195, 203, 218, 439, 458
Berger, Hermann 254
Bomhard, Allan R. 174, 241–242, 249, 252, 259, 381, 448
Bopp, Franz 288
Bouda, Karl 242

C
von Carolsfeld, Schnorr 289, 291
Christaller, J.G. 29, 33–34, 205
Collinder, Björn 241
Curr, Edward M. 287, 292, 300, 302, 306

D
Danquah, Joseph Boakye 29
Dolgopolsky, Aharon B. 174, 236, 241–242, 257, 259, 432, 439–440

E
Ehret, Christopher 11–13, 60, 64, 170–178, 180–181, 191, 194, 203–205, 432
Emeneau, Murray B. 201, 427–428

F
Fleming, Harold C. 25, 39, 47, 58, 109, 242, 256, 258, 381
Fowler, Catherine S. 309–310, 315

G
Gatti, Riccardo 287, 289–290, 292–293, 302
Greenberg, Joseph H. 18, 22, 170–171, 173, 181, 191, 197, 200, 205, 241–242, 244–245, 249, 252, 255, 257–258, 287–288, 316–317, 322–324, 363, 381, 432, 440, 448, 450, 452–457
Guthrie, Malcolm 191, 205, 432, 457

H
Haeckel, Ernst 288, 383, 387–388
Hale, Kenneth L. 315–317
Hayward, Richard J. 40, 58, 60
Heath, Jeffrey 181, 202
Helimski, Eugen 181, 235
Hill, Jane H. 310, 317
Hoijer, Harry 224, 432
Hopkins, Nicholas A. 302, 311

I
Illich-Svitych [Illič-Svityč], V.M. 241, 249, 252

J
Jakobson, Roman 417–418, 433–434
Janhunen, Juha 224, 228–230, 236
Jungraithmayr, Herrmann 163, 165, 196, 205

K
Kaufman, Terrence 225, 311–316, 318
Koeppelmann, Heinrich 241
Kroeber, A.L. 25–26, 28, 253
Kuhn, Thomas S. 289, 384

L
Lamberti, Marcello 57, 60, 194, 205
Lieberman, Philip 360–363, 366, 368–369, 410, 459

M
Manaster Ramer, Alexis 310
Müller, Max 288
Murdock, George Peter 418

N
Newman, Paul 12, 165, 245, 257
Nikola(y)ev, Sergei L. 242–243
Nkruma(h), Kwame 30–31

P
Pagel, M. 439
Pedersen, Holger 241
Peirce, Charles S.S. 255–256
Portman, M.V. 201, 287, 307, 411

R
Rask, Rasmus 241
Ray, Sidney H. 287, 293, 296, 306
Reh, Mechthild 172, 195
Renfrew, Colin 242–243, 373
Rensch, Calvin R. 312–316
Rousseau, Jean Jacques 288
Ruhlen, Merritt 18, 173–174, 181, 195, 203, 241–243, 245, 247, 253, 256–258, 322–324, 337, 339, 418, 434, 439, 440, 457–458, 462

S
Sagart, Laurent 251, 381
Sapir, Edward 242–243, 253, 255–256
Saussure, Ferdinand de 288, 393

Schadeberg, Thilo C. 39, 198
Schmidt, Hanns-Peter 271
Schmidt, P.W. 287, 293, 306
Shafer, Robert 242
Shevoroshkin, Vitaly 243
Sottile, Roberto 57, 60, 194, 205
Starostin, Sergei A. 57, 65, 138, 226, 232, 242–244, 246–248, 250, 252, 259, 339, 381, 448
Swadesh, Morris 203, 256, 445

T
Taube, Karl 312
Tikkanen, Bertil 252
Trask, R.L. [Larry] 201, 417–429, 434
Trombetti, Alfredo 256, 287–293, 302, 305, 381, 392
Turner, Christy G. 19–22
Turner, R.L. 425–426

U
Usher, Timothy 292, 293

V
Vajda, Edward J. 244, 257

W
Wescott, Roger W. 244
Whitehouse, Paul 243, 245, 248, 256–259, 293
Wichmann, Søren 317–318
Witzel, Michael 265–266, 268–272, 277–280, 282–283

Z
Zaborski, Andrzej 57
Zegura, Stephen 345, 347, 351–353