

LANGUAGES & NATIONS
THE DRAVIDIAN PROOF IN COLONIAL MADRAS

THOMAS R. TRAUTMANN

© 2006 UC Regents
Buy this book

University of California Press
Berkeley and Los Angeles, California

University of California Press, Ltd.
London, England

© 2006 by Thomas R. Trautmann

Library of Congress Cataloging-in-Publication Data

Trautmann, Thomas R.

Languages and nations : conversations in colonial
South India / Thomas R. Trautmann.

p. cm.

Includes bibliographical references and index.

ISBN-13, 978-0-520-24455-9 (cloth : alk. paper),

ISBN-10, 0-520-24455-9 (cloth : alk. paper)

1. India—Study and teaching—India—History—
19th century. 2. Dravidian philology—History—
19th century. 3. Orientalism—History—19th
century. 4. Indologists—India—Madras—
History—19th century. 5. Ellis, Francis Whyte,
d. 1819. I. Title.

DS435.8.T73 2006

410—dc22

2005018465

Manufactured in the United States of America

13 12 11 10 09 08 07 06 05 04

10 9 8 7 6 5 4 3 2 1

This book is printed on Natures Book, containing
50% post-consumer waste and meets the minimum
requirements of ANSI/NISO Z39.48-1992 (R 1997)
(*Permanence of Paper*).

Explosion in the Grammar Factory

In the European thought of the eighteenth century, languages and nations were understood to be parallel, in that the histories of both were viewed as governed by genealogical relations and linked; therefore, the genealogical relations among languages could serve to extend the reach of historical memory concerning the relations among nations and to repair it where it was defective. Language history in this sense became a new tool for ethnology on a universal scale, producing original and unexpected groupings of kindred languages that have in many cases endured to the present. To supply this new ethnological project with the raw material on which it operated required the production of grammars and dictionaries virtually without limit and covering the entire world—an explosion in the grammar factory that continues to this day.

Several of these new and still-valid groupings are associated with British India: the Indo-European language family, which is the best known and the pattern for all the others; the Malayo-Polynesian language family; the Indo-Aryan origins of the Romani language; and the Dravidian language family. Of these four cases, the emergence of the concept of a Dravidian language family has the richest archive, available in hitherto unexamined colonial records of Madras. The first published proof of the Dravidian language family appeared in British-Indian Madras in 1816, the product of a circle of scholars associated with the College of Fort St. George. In this book I examine the languages-and-nations project that the British brought with them to India in the light of the Dravidian proof,

and vice versa; that is, this book moves on two planes, using each to illuminate the other. The premise of the book is that the exceptional productiveness of British India as a terrain for the languages-and-nations project of Europe had to do with the exceptional development of language analysis in India since the times of Pāṇini and earlier. The conjuncture of these two traditions of language analysis in British India can be examined at close quarters through the Dravidian proof.

The first two chapters concern, respectively, the European and Indian traditions of language analysis. In the remaining chapters the discussion turns to Madras and the Dravidian proof. In the present chapter I examine well-known material, including the formation of the Indo-European concept, but from a new direction, and interpret it in a way that departs considerably from the received view. I begin with the idea of what I call locational technologies, the more inclusive set that includes the languages-and-nations project and its genealogical scheme of locating particulars in relation to one another.

TECHNOLOGIES OF LOCATION

The propensity to exaggerate the originality of one's thoughts is a failing that is perhaps most acute among those who do not work in teams as the laborers in the vineyards of the natural and social sciences do. Humanities scholars, given to the solitary mode of production, most often work in caves or in studies—like St. Jerome who, as translator of the Bible into Latin, an immensely successful book, is something of a patron saint of scholars. We have two images of St. Jerome. In one he is shown living in a rough cave, with ink and paper at hand, smiting his breast with a stone, saying to himself, as I imagine it, “I *must* finish my book!” The scholar in agony is paired with another image, of which Albrecht Dürer has made so appealing a rendering, that of the great scholar in his study, a tame lion at his feet, sunlight streaming through the window: the scholar happy in his work. I think of these two opposed images, of scholarly agony and pleasure, as St. Jerome before and after tenure. In both, the scholar is—not counting the lion—utterly alone.

For historians and others writing in solitude (I include myself), it is all too easy to be seduced by the pleasing notion of one's own originality, lured by its inherent sweetness and egged on by romantic ideas of individual work of literary genius and the individual scientific breakthrough. Through the distorting optic of an exaggerated sense of individual originality, the social and the historical pass out of view, and a

single self occupies the center of the field of vision. It takes special effort to remember that every intellectual project derives its meaning in relation to larger, collective projects that long preceded and will long outlive the individual, and that the text written by an individual contains within it many voices of a continuing conversation. When we open out the field of vision to its widest extent, the individual work becomes a speck in a larger intellectual project that is the work of many hands across many nations and centuries, It is in relation to this wider field that the efforts of individuals are rendered significant and lasting.

One such project of the *longue durée* is the charting of the heavens, a project that has been underway since the times of the ancient Sumerians, perhaps longer, and which we have every reason to think will continue in process as long as there is a human race to carry it on. Astronomy above all seems to have a unitary history that combines the work of countless individuals of many nations over a very long time. John Playfair, speaking at the end of the eighteenth century, put it nicely when he said that the successive developments in the observation of the heavens and the reasoning about them comprise “an experiment on the human race, which has been made but once” (Playfair 1790:136). It is this compelling sense of singularity that puts astronomy at the heart of ideas about science as a progressive accumulation of knowledge that is universal.

Another such project is astronomy’s inverse, the mapping of the earth’s surface. Yet others are the construction of a unitary chronology of the past, and the classification of nations and languages—the topic of this book. We may call all of these locational projects because they define representational spaces and represent entities as locations within those spaces.

The space of each of these locational projects of the *longue durée* is defined by what, for want of a better word, I will call a technology of location. The star chart is a good example. It defines its space by dividing the sky as seen from earth with lines of declination and right ascension. Within that space one determines the placing of each heavenly body in terms of degrees (or hours), minutes, and seconds. Those units bear witness to the Mesopotamian origins of this locational project: 360 degrees in the whole circle of the sky (or 24 hours), 60 minutes in a degree, 60 seconds in a minute, in the base 60 numbering system of the Sumerians. Every star in the heavens has its position fixed by a pair of numbers. The star Aldebaran, for example, is at right ascension 4 h. 30 m., declination 16°19'. The relation between any pair of stars is a derivative of their positions within this space.

The star chart, like the other locational projects, arose in the deep past and is very much in use today. This long-term intellectual venture, carried forward by the incremental contributions of innumerable individuals over many centuries and across many different countries and cultures, this vast and largely anonymous collective effort, is, like the others, a part of the living core of modernity. Yet precisely because it is so very central it is practically invisible. No deep rupture, no Kuhnian paradigm shift, has cast it aside. Not even the famous shift from an earth-centered, Ptolemaic conception of the planetary system to a sun-centered, Copernican one has upset the structure of the star chart.

The space of the star chart first developed as a fiction that, though false, turns out to be highly useful: the useful fiction that the sky is the interior of a titanic sphere, on the surface of which the stars are hung. This imagined sphere was then marked off into sectors by a rectilinear grid as a locational technique to fix the heavenly bodies in place for study. Or perhaps, in the beautiful metaphor of an ancient Sumerian poem, the sky is the tablet of lapis lazuli upon which the goddess Nidaba inscribes cuneiform signs, the stars, which tell the destinies of human beings down below.¹

For the study of the earth, this imagined celestial sphere was projected back onto the real sphere, or spheroid, of the earth, for which it is both useful and a reasonably true representation. The earth, in turn, was divided into sectors by degrees of longitude and latitude, which define a related locational space. Thus the earthly space was theorized through astronomy. It was *astrology*, however, that was the connective tissue between the two projects of mapping the stars and mapping the earth, the belief, that is, in the influence of the heavenly bodies upon earthly destinies—for what the goddess Nidaba writes on the tablet of the sky is the destinies of humans, if only we can read it. The desire to read that sky tablet of our futures was a powerful motivation for the entire enterprise from the start, but it was the locational projects themselves that survived the ultimate casting out, from the table of recognized sciences, of the ages-old project of astrology that first set these sciences in motion.

Ptolemy of Alexandria, in Roman Egypt of the second century A.D., wrote works of astronomy, astrology, and geography that became canonical for later ages and had an immense influence in the Christian West, the Muslim Middle East, and beyond, into Central Asia and India.² Al-

1. I thank Piotr Michalowski for this information. See Michalowski 1992.

2. The main astronomical work of Ptolemy is the *Syntaxis*, called the *Mathematical collection* or *The great astronomer*. His *Geography* has already been mentioned. His main

though Ptolemy was not the inventor of the highly theorized space of the locational projects of astronomy and geography, he raised the science of Mesopotamia, Egypt, and Greece to a higher level, and it was through his writings that this tradition was transmitted to future ages. We know the names of some of his predecessors in the development of this space, such as Marinus, Hipparchus, and Poseidonius. But the names of the many Mesopotamians and Egyptians who had made celestial observations for two thousand years before Ptolemy are largely unknown. There is a certain justice in calling this astronomical and geographical space Ptolemaic, and I shall do so in this book, but it must be understood that it was not his invention; it was the culmination, in his works, of a long-continued effort by many, many people.

The Ptolemaic grid of declination and right ascension for the mapping of objects in space from the viewpoint of earth, and the grid of longitude and latitude for the mapping of objects on the earth's surface as seen from space, have become the taken-for-granted frameworks within which those mappings proceed as a steady accumulation of knowledge by increments—our surest example of progress in knowledge. Such locational devices are not mere metrics, like the meter stick that stands inertly in the corner until it is taken up to measure something. The grid defining Ptolemaic space is, rather, like a vessel that is meant to be filled; it has, as it were, the project of its filling engineered into it. There is nothing passive about the grid. It *asks* to be filled up with an infinite number of points of information, put into meaningful relation to one another through their locations in the grid. It is a project engineered into a tool for its accomplishment, a locational project embedded within the locational technology for carrying it out.

While the locational *project* of mapping the earth has an event-filled history that we can partially capture, the locational *technology* is the enduring, defining frame for the project. One might suppose that Ptolemaic space has had a continuous existence from antiquity to the present, but that is not the case. Most of the first printed atlases produced in Europe and many early Arab maps were Ptolemaic, in the double sense that the maps were framed by the Ptolemaic grid of longitude and latitude, and the maps themselves depicted the features of the earth's surface as rep-

work of astrology is called, simply, the *Tetrabiblos* (*Four books* or *Fourfold book*). Although histories of science usually deal only with his astronomical and geographical works, his astrological work was part and parcel of his life project, as was the case for most astronomers of the ancient world.

resented in the tables of places and their longitudes and latitudes given in the text of Ptolemy's *Geography* (see figure 1). But it is a question whether the maps attributed to Ptolemy and attached to his *Geography* are truly his or are of later design. Moreover, many maps of antiquity and of the European middle ages were clearly made in a quite different space that is not Ptolemaic; examples are pilgrimage maps and the well-known T-O maps, in which a circular earth is divided into three continents by a T-shaped watery body: Asia at the top, Europe to the left, and Africa to the right. For a very long time many, indeed most, maps were not constructed within the Ptolemaic space, especially in the middle ages.

Moreover, there were serious alternatives to the Ptolemaic space that had possibilities for cumulative, scientific mapping and that might have displaced it. A leading example comes from the *portolan*, or harbor-finding, sea charts for sailors developed in Catalonia and coastal northern Italy in the thirteenth century. These charts had wind roses, or compass roses, with lines (called rhumb lines or loxodromes) radiating outward from their centers. The four-sided figures formed by the intersection of lines from several compass roses at standard locations could define the position of any point in the open ocean and fix the locations of the coastline for purposes of navigation, just as latitudes and longitudes could. We have several surviving examples of the transferal of these principles to the mapping of the world as a whole, in competition with the Ptolemaic maps (see figure 2). The space of the *portolan* maps still survives today, marginally, in the navigational charts used at sea, but always in combination with the now standard Ptolemaic coordinates of longitude and latitude.

Again, the Ptolemaic map of the world, and Ptolemy's tables of locations which the map illustrates, contained many errors which were exposed and corrected by the new knowledge accumulating through European seagoing. Since the new knowledge contradicted Ptolemy's tables and his highly theorized world map, it is not surprising that we have early printed maps based on the new knowledge that cast aside the Ptolemaic grid of longitude and latitude along with the Ptolemaic map. The history of mapping in Europe since the first printed atlases of the Renaissance in brief is the story of how the map of Ptolemy was ushered out and replaced by the knowledge newly acquired on the ground, even as the Ptolemaic grid was made the standard frame for the space of that new, improved mapping. We may say that Ptolemaic space survived not only the overthrow of Ptolemy's planetary system by that of Copernicus but also the demise of his own map of the world. Everything became obsolete but

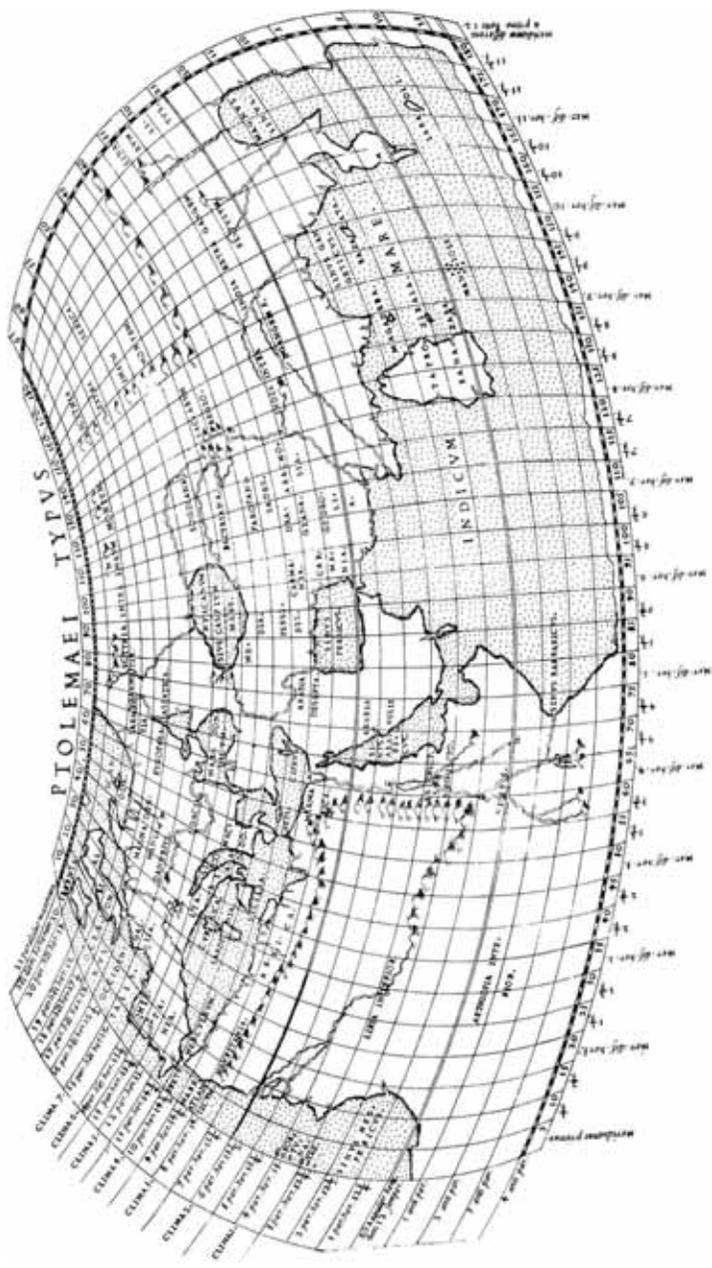


Figure 1. The world map of Ptolemy, according to a Venetian editor, 1561. Longitude is in fractions of hours east of the Fortunate Isles and latitudes in the number of hours in the longest day of the year. (From Brown 1949: 55.)

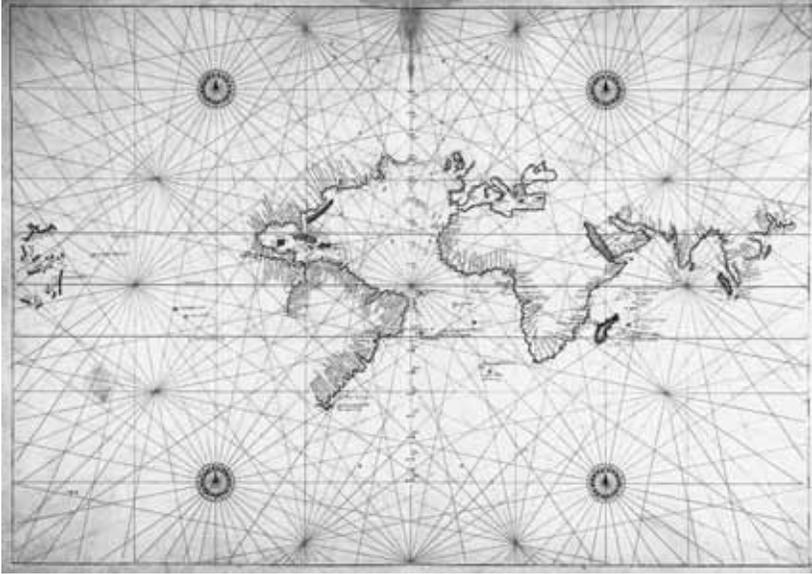


Figure 2. A *portolan* map of the world by the Genoese mapmaker Battista Agnese, c. 1536. (From Whitefield 1994: 59.)

the framework, while the framework itself has proven exceptionally enduring and, indeed, indispensable. Cheap hand-held Global Positioning System devices, drawing upon hugely expensive geosynchronous satellites circling the earth many times a day, now routinely locate objects of all kinds in Ptolemaic space.

We are all familiar with the graticule of longitude and latitude because it appears on maps and globes and is taught in school. The comparable technology for the location of historical events, however, is not well known, being a more dispersed entity with a more backstage existence. Yet all written history depends upon the chronologies that ancient nations worked out for themselves and the synchronisms among the national histories established by the Christian chronologers of the early centuries after Christ. It was Eusebius, in the fourth century A.D., who constructed a chronological canon or table that synthesized national chronologies into a single whole for the writing of a history of Christianity. This table is the foundation of world-historical chronologies to this day.³

3. I discuss the Eusebian grid of chronology more fully in Trautmann 1987:206–9.

The Eusebian chronological canon is a locational technology for time. It is a simple device in which events in the chronologies of the various ancient nations are put side by side in columns, such that each horizontal line represents a year of synchronic time. Thus, reading across the columns one finds the synchronisms between events in the various national histories. It was by this means that biblical history was synthesized with the Greek chronological system of Olympiads, and these plus the national histories of the Romans, Chaldeans, Egyptians, and others were brought into a synoptic table that rendered possible the writing of a history across nations—a universal history.

The familiarity of the Ptolemaic grid of longitude and latitude contrasts with the near-total obscurity of the Eusebian chronology. Very few historians who are not historians of the ancient world, let alone members of the general public, will even have heard of the chronological canon of Eusebius, and yet it is the source of all current dating of historical events. What is more, we have no complete surviving example of the chronological canon in the original Greek and know of it only through secondary sources and translations. It is a paradox of history that the locational technology upon which all dating in universal-historical time depends has been handed down only through fragments preserved by the later Greek chronologers, such as Gregory Syncellus, or in translations, especially the Syriac one edited by Josef Karst and the Latin one by St. Jerome, which supplied the chronological grid for historians in western Europe from very early on. Yet despite its dispersed, virtual state, the Eusebian grid for historical chronology has grown through the ages. We continue to lengthen its columns by the addition of earlier and later events, and to widen its reach by integrating into it the chronologies of more different national histories, even if we do not know we are doing so.

It is well to stress the theoretical, highly constructed character of the Eusebian chronology and the concept of historical synchronicity embedded in the structure of the table. Without that locational technology there is no basis in synchronism for the construction of histories across nations. It was the transnational spread of Christianity that occasioned the need to map the histories of the nations upon a unitary plan defined by the horizontal lines of synchronism and the before-and-after logic of the columns.

We come now to the third locational technology of this discussion, the one which especially concerns the topic of this investigation into lan-

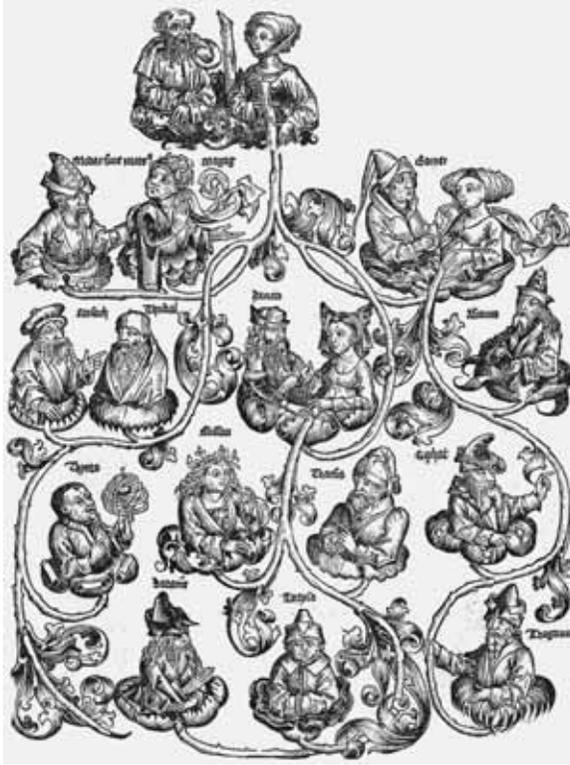


Figure 3. The Mosaic ethnology. The descendants of Japhet, son of Noah, as depicted in the *Nuremberg Chronicle*. (From Schedel 1493: fol. 16.)

guages and nations. It is the Tree of Nations, or, as I like to call it (since it comes from the Book of Genesis in the Bible, attributed to Moses), the Mosaic ethnology.⁴

In Genesis, the ten patriarchs from Adam to Noah are succeeded by a branching tree of Noah's sons, Shem, Ham, and Japhet, followed by their sons, and so forth, comprising a large family tree of patriarchs whose progeny are the nations of the earth. The names of the patriarchs are the names of the nations. Thus, for example, the patriarch Eber is the father of the Hebrew people, and the patriarch Javan gives his name to the Greeks, that is, the Ionians ("Yauna" in Persian, "Yavana" in Sanskrit) (see figure 3).

4. For more detailed discussion of the Mosaic ethnology, see Trautmann 1997, ch. 2, "The Mosaic ethnology of Asiatick Jones."

The underlying treelike branching structure is built up from relations of kinship. Not all kinship relations are used, however, but only patrilineal ones, relations of descent through males only. Thus relationships of kinship are sculpted by abstracting only the patrilineal ones to make a tree that ramifies endlessly and does not return into itself. This rami-fying figure is maintained by suppressing the representation of descen-dants through females and the marriages that would otherwise inter-twine the branches of the Tree of Nations. Such a structure is what the anthropologist E. E. Evans-Pritchard (1940) called a *segmentary line-age*, and it is of a kind widely used in North Africa and the Middle East, including the biblical lands, to represent and regulate relations among lineages.

The Mosaic ethnology is a simple locational technology for deter-mining the relations among peoples, conceived as branching lineages of the human family tree, as relations of near and far. It is quite capable of worldwide extension and has been the basis of ethnological classifications for a very long period of history. It is quite different from the Hegelian psychohistorical calculus of Self and Other that is so important a con-struct for theorizing today, since in the Mosaic ethnology every human being is related to every other, but in varying degrees of nearness. This device was widely used by all the Peoples of the Book—Jews, Christians, and Muslims. It was *the* framework within which newly discovered peoples were fitted, to give order and meaning to the rush of new eth-nological knowledge that came with the expansion of Islam, and, later, with the European expansion from the Renaissance onward. Thus Mus-lim scholars found it a matter of importance to determine whether the Chi-nese are descended from Ham, Shem, or Japhet, and Columbus brought with him to the New World a Morisco who knew Arabic and Hebrew to determine whether the American Indians were descended from the Lost Tribes of Israel. Because they are all based on the Mosaic ethnology, uni-versal histories in the Islamic Middle East and India, on the one hand, and in Christian Europe, on the other, have a family likeness at a fun-damental level. However fraught their relations became, Christians and Muslims in the middle ages and the early modern period shared an intel-lectual culture built upon common structuring principles that they did not share with peoples of nonbiblical religions. This commonality in-cluded the Mosaic ethnology as well as a master narrative of universal history, according to which the world was created but a few thousand years ago and peopled from a single central point, and the nations mul-tiplied following the confusion of tongues by God at the Tower of Ba-

bel. The fanning out of lineages from a point of origin indexes the figure of the patrilineal Tree of Nations.

The Mosaic ethnology is the main technology of location to be found in universal histories of the Peoples of the Book from the Bible itself down to the eighteenth century. It survives, and indeed flourishes, to this day, but in a transmuted, secularized, and scientized form, as the structuring principle of historical linguistics (while ethnological analysis has turned to the Hegelian frame of Self versus Other). This adaptation of the Mosaic ethnology to linguistic history deserves our attention.

It is something of a scandal for linguists that the roots of their science are planted in the Bible, and this fact has been the object of repeated attempts at erasure and willed forgetting. Histories of linguistic science written within the discipline have represented its rise as a rupture from the past, a breakthrough moment when science emerged, miraculously, from nonscience. Carryovers from the prescientific age into current usages, such as the names “Semitic” and “Hamitic”—taken from names of Noah’s sons—for linguistic groups are interpreted as nothing more than vestiges of older, nonscientific conceptions now scientifically reconceptualized. But in spite of its collective suppression in the charter myth of the discipline of linguistics, the Mosaic ethnology continues to shape the work of historical linguistics to this day. The radiating structure of the Tree of Nations, now under the new-old scientific name of “cladistics,” underwrites the grand projects of linguistics, such as George Grierson’s *Linguistic survey of India* (1903–27) and Antoine Meillet’s *Les langues du monde* (1952). The attempt to unify languages into a finite set of families through the conception of a radiating movement in time, and the ambition to unify all these families into a gigantic superfamily by discovering connections among them at ever-deepening levels of past time, ran strong in the project of Joseph Greenberg, whose books on the languages of Africa (1955) and the Americas (1987) were complemented by his work on Eurasian languages (2000), completed just before his death, as well as in Illich-Svitic’s work on Nostratic (Manaster Ramer 1993).

There is a mystery here that needs exploring, one that we will not get at through the “history of linguistics” narrative. It is the mystery of how a framework for the classification of *peoples* becomes transformed into one for the classification of *languages*. Solving this problem is crucial to understanding why languages and nations came to be twinned in European thought, and I will shortly propose a solution. Before turning to that matter, however, a few final comments are needed about the three technologies of location we have been examining: the Ptolemaic grid of

space, the Eusebian grid of chronology, and the Mosaic ethnological tree. They appear to be concerned with, respectively, space, time, and ethnology. But, in fact, all three are concerned with locations in both space and time at once, or, shall we say, in space-time. They differ not in their media but in their objects. The Ptolemaic grid may seem to be purely about space, but in fact space and time are mutually convertible at fixed rates of exchange, so to say, such that 15 degrees of longitude equals one hour's difference of time. The borders of the Ptolemaic world maps regularly defined longitude by both degrees of arc and hours of distance from a meridian, and the various latitudes by both degrees of arc from the poles and hours of daylight at midsummer. Indeed, the interconvertibility of time and space measures is essential to this locational technology. In the Eusebian chronology, time is the vertical axis, and space—that is, the geographically dispersed nations—is the horizontal one. And in the Mosaic ethnological tree, the vertical axis again is of time, while the horizontal one tracks the spatial dispersion of the nations. The three locational technologies are directed at different kinds of objects in unitary space-time: geographical objects, historical events, and human communities or nations.

LANGUAGES AND NATIONS

We come, then, to the main business of this chapter, which is to excavate the meaning of the twinning of languages and nations in European thought. We do so by attempting to solve the mystery of how the Mosaic ethnology, a classification of *peoples*, was transformed into a framework for the classification of *languages*. This is not, however, a mystery of which anyone has written before, and so I must first persuade the reader that there is indeed a mystery here to be solved, and explain why the mystery itself has been hidden.

As we have seen, one explanation for why the mystery has not hitherto been addressed is historical linguistics' nonrecognition of the Mosaic ethnology, the biblical Tree of Nations, as the source of the genealogical trees of language. A collective unwillingness to find the roots of linguistic science in religion, as well as to acknowledge the origin of the discipline in what is now the separate discipline of ethnology, is the reason the problem is not posed in autohistories of linguistics.

From that beginning, I direct the reader's attention to Sir William Jones's proposal of the idea of the Indo-European language family in British-Indian Calcutta, in order to examine closely how it came about

and why it came about at a particular point in history and not earlier. I do so for a couple of reasons. In the first place, after Jones, the study of Indo-European languages quickly became the most intensely cultivated field in comparative philology, supplying models and standards for other fields. In the second place, Jones provided the breakthrough moment for the disciplinary autonarrative of a bounded linguistic science miraculously emerging from the nonscience that preceded it, a narrative that exerts a strong pull upon any investigation of the event in question. I have already suggested that this narrative is defective, and we need to explore the matter more deeply.

Accounts of the history of linguistics almost invariably quote a certain passage of Sir William Jones from the “Third anniversary discourse,” on the Hindus, which Jones delivered to the Asiatic Society at Calcutta, as its president and founder, in 1784. The overall scheme of these annual discourses marking the anniversary of the Society’s founding was to examine the five “stock nations” of Asia, namely, the Hindus or Indians, Arabs, Persians, Chinese, and Tartars, one by one, and determine their relations to one another. The examination focused on four features: language and letters, religion and philosophy, architecture and sculpture, and arts and manufactures. This is the famous passage:

The *Sanscrit* language, whatever be its antiquity, is of a wonderful structure; more perfect than the *Greek*, more copious than the *Latin*, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the roots of verbs and in the forms of grammar, than could possibly have been produced by accident; so strong indeed, that no philologist could examine them all three, without believing them to have sprung from some common source, which, perhaps, no longer exists; there is a similar reason, though not quite so forcible, for supposing that both the *Gothick* and the *Celtick*, though blended with a very different idiom, had the same origin with the *Sanscrit*; and the old Persian might be added to the same family, if this were the place for discussing any question concerning the antiquities of Persia. (Jones 1788d:422–23)

The conception expressed here is astonishingly modern. It specifies five ancient languages and posits that they have “sprung from some common source, which, perhaps, no longer exists,” and so constitute a family of languages. This is, of course, the Indo-European family, although the name had not yet been invented, and the inferred lost common source is now called Proto-Indo-European. This passage *does* have a breakthrough quality to it. The grouping of languages it propounds was without precedent, both in Europe and in India, and the idea that languages of Iran and In-

dia are closely related to European ones created an entirely new sense of deep history. The literatures of these languages preserved no recollection of a common origin and the migrations it implies, but the language themselves preserved proof of it where historical memory had left no trace. It was just this ability of philology to restore a lost history through comparison of languages that was the most spectacular of its powers. Moreover, the historical relationship Jones here proposes has proven durable, and the idea remains valid today, more than two centuries hence.

When this famous passage is abstracted from one narrative, the “Anniversary discourses,” and inserted into the narrative of the rise of linguistics as a self-contained field of scientific study, in the process its reading is changed. When we examine it in its original context, the breakthrough quality is not entirely lost but it is considerably qualified. The main thing to grasp about the “Anniversary discourses” is that they were an *ethnological* and *historical* study, not a linguistic one as such; thus the language data function in the argument as evidence for propositions about historical relations among nations or races, not for propositions about historical relations among languages as an end in itself.

Far from constituting the study of languages as a self-contained discipline, Jones treats languages as a means, and just one of many means, to disentangle ethnological relationships. It is a paradox that someone so gifted in languages regarded languages as mere instruments of knowledge and not objects of knowledge in themselves, and used the word “linguist” to mean simply a knower of languages, not a scholar of a self-contained object of knowledge. We see at work here the common presumption of his age: that languages and nations are inextricably connected, so much so that relations among languages index the relations among nations, and historical relations among nations can be inferred from relations among languages. Even when, in later years, comparative philology acquired a body of works all its own, the connectedness of language with issues of nation and race remained so deeply presumed that it was not available for discussion and debate. It was not until more than a century after Jones that Europeans declared, as a surprising new discovery, that the connection between languages and nations is contingent, not necessary.

As we zoom out from Jones’s passage on Indo-European and look at the “Anniversary discourses” as a whole, the picture changes considerably.⁵ This early formulation of the Indo-European language family has

5. The ten anniversary discourses were published in the first four volumes of *Asiatic researches*. They may also be found, collected together, in the *Works* of Jones, of which

several inclusions which are by no means acceptable today: the ancient Egyptians, for example, and the Chinese, and even the Incas and Aztecs of the New World. On the other hand, Jones does not include some languages universally acknowledged to be Indo-European today, most notably the Slavic, which he classifies with the Tartar language. Eventually one becomes aware that the scheme has an underlying structure, as follows:

Shemites	Religion
Hamites	Arts and sciences; civilization
Japhetites	Nomadism

In the conclusion of the “Anniversary discourses” Jones finds that the five principal nations of Asia are reducible to three: the Indians (which includes the Persians and the Chinese), the Arabs, and the Tartars—corresponding to Noah’s sons, Ham, Shem, and Japhet, respectively.

This structure of associations was taken over from Jacob Bryant’s *Analysis of antient mythology* (1744–76; see Trautmann 1997:41–46). It is somewhat eccentric, in that it identifies the Indians, and hence the Persians and Europeans, as Hamites, in place of the more usual view that Europeans descended from Japhet. It did not succeed in overturning the latter view, and indeed “Japhetic” was a common label for the language family that came to be called Indo-European. The organizing idea of the Bryant-Jones scheme is that all the arts and sciences of primitive times are attributable to the Hamites, who were also the first to turn from the true religion, known to all mankind in the times of Noah, to idolatry; while it is among the Shemites that true religion was preserved, and the Japhetites fell out of the agrarian life established by God in Eden into a life of nomadism. Jones extends this scheme around the globe: for the New World, for example, he has a two-nation theory, such that the nomadic, hunting American Indians are Japhetites, but the Aztec and Inca civilizations are Hamitic. One becomes aware, by the end of the “Anniversary discourses,” that the whole scheme is a rational working-out of the story of the peopling of the world from a single stock by the descent of Noah, in a past contained within the chronology of Archbishop Ussher, according to which the world was created no further back than 4004 B.C. and human beings spread across the earth after the universal

the 1807 edition is the most widely available, or the newly published edition of Garland Cannon (1993).

flood, dated to 2349 B.C. The assumption, then, was that within the short time span of the biblical chronology, the human family split up into descent lines and fanned out rapidly across the face of the earth.

This brief look at the larger scheme of the “Anniversary discourses” is useful not so much in the way of drawing attention to Jones’s errors as in providing a critical distance on the normal view: that the idea of Indo-European arose immediately from the apprehension of similarity among Latin, Greek, Sanskrit, and Persian, and that historical linguistics came into being by a kind of spontaneous combustion in the mind of Jones. Jones himself promotes that view when he says of Sanskrit, Latin, and Greek, that “no philologist could examine them all three, without believing them to have sprung from some common source.” Taken literally, the statement implies a strict empiricism, according to which the historical relationship of the languages is directly seen in the languages themselves and requires only that they be placed side by side for this deeper seeing to take place. This is the “discovery” narrative at its simplest, and it shuts down further investigation into that account’s true nature before any inquiry can begin.

If the narrative of simple discovery were true, the Indo-European language family would have been brought to light long since, in antiquity. The Greeks, Persians, and Indians had dealings with one another, and there were undoubtedly many opportunities for the kind of direct comparison that, by Jones’s account, would have led spontaneously to the positing of a common ancestry. Yet not only did the ancients fail to discover the Indo-European language family, the Greeks took almost no interest in recording matters to do with languages in their accounts of Persia and India, such as the *Persika* of Ctesias or the *Indika* of Megasthenes. Jones remarks on this Greek indifference to other languages at the beginning of the “Third anniversary discourse”: “It is much to be lamented, that neither the *Greeks*, who attended ALEXANDER into *India*, nor those who were long connected with it under the *Bactrian* Princes, have left us any means of knowing with accuracy, what vernacular languages they found on their arrival in this Empire” (Jones 1788d:421–22). Nor did the sophisticated linguistic science of the Indians take any notice of Greek, beyond a few vague references to the language of the barbarians (*mleccha*). Those for whom the similarities of the ancient languages should have been most telling failed entirely to come to the conclusion that Jones represents as springing directly from the simple juxtaposition of languages.

The truth of the matter is that the brute facts of language similarity

do not interpret themselves, and indeed are subject to any number of interpretations. That this is so is proven by the many competing interpretations of these facts in the eighteenth and nineteenth centuries. Nathaniel Brassey Halhed, a friend of Jones, for example, had come to a similar conclusion about the historical relationship of Sanskrit and Greek somewhat earlier, in his *Grammar of the Bengal language* (Halhed 1778; Rosane Rocher 1983). On the other hand, Dugald Stewart, the leading philosopher of his generation, followed the lead of Christoph Meiners (*Historia doctrinae de vero Deo*, 1780) and Gottlieb Siegfried Bayer (*Historia regni Graecorum Bactriani in qua simul Graecarum in India coloniarum vetus memoria explicatur*, 1738) and made a fool of himself by publishing, as late as 1827, when the comparative philology of Indo-European was well on the way to its greatest successes, an elaborate demonstration that Sanskrit was similar to Greek because it *was* Greek, overheard by the wily brahmins from Alexander's soldiers during their incursion into India, and adapted by them as a kind of pig Latin with which to mystify the people and hold them in subjection to their priestcraft (Stewart 1827:110; Trautmann 1997:124–26). In the meantime Friedrich Schlegel (*Über die Sprache und Weisheit der Indier*, 1808) had accepted the Indo-European idea, but with a difference from the formulation of Jones that was deeply consequential, according to which Sanskrit is not the co-descendant of a lost ancestral language with Greek, Latin, Gothic, Celtic, and Persian, but is *itself* the ancestor and source of the other Indo-European languages. It was Schlegel's program to reengineer the conception of ancient India as the pure source of a lost, primitive innocence and ancient wisdom—the India of Romanticism, an India that embodied the childhood of the human race. While Franz Bopp (1816, 1833, 1845–53), the great pioneer of Indo-Europeanist comparative philology, resisted this view and held to the one Jones had expressed—that Sanskrit was the sibling and not the mother of the other Indo-European languages—other writers, perhaps the majority of them in the early nineteenth century, held some version of Schlegel's view. Thus the comparative examination of Sanskrit and Greek has led to three quite different interpretations: the identity of Sanskrit with Greek, the co-descent of the two from a common ancestor, and the derivation of Greek from Sanskrit.

Perhaps the most telling case that helps us free ourselves from the naturalizing narrative of the discovery of Indo-European by the mere inspection of Sanskrit, Latin, and Greek is that of the Jesuit missionary Gaston-Laurent Coeurdoux, who spent his whole adult life in South India. Only now, thanks to the work of Sylvia Murr (1977, 1987), do we

have a full appreciation of this fine scholar-missionary's Indological work, which, as Murr has shown conclusively, was recycled (to put it politely) by the Abbé Dubois in the well-known and much reprinted work *Hindu manners, customs and ceremonies* (3d ed., 1906). Father Coeurdoux also conceived a form of the Indo-European idea, and he did so before Jones, in a letter to the Académie des Inscriptions written in 1768. However, his ideas were only published in 1808—after his death and after Jones had published the “Third anniversary discourse”—thanks to the efforts of another French Indologist, Abraham-Hyacinthe Anquetil-Duperron.

After providing a number of examples of Latin, Greek, and Sanskrit words, Coeurdoux interprets the similarities among them in this way:

The Samskroutam language is that of the ancient Brahmes; they came to India from the north of that country, from Caucasia, from Tartary, which had been peopled by the descendants of Magog. Of the sons of Japhet, some spoke Greek, others Latin, still others Samskroutam. Before their total separation, their languages were somewhat mixed because of the communication they had among each other; and there remain vestiges of that ancient intercourse, in the common words which still exist, and of which I have reported a part. (My translation of a passage in Coeurdoux c. 1768:666; see also Murr 1987, pt.1, ch.7)

While Coeurdoux, like Jones, interprets the similarity among the three languages in biblical terms, that is, in the terms of the Mosaic ethnology, we see in this passage that this particular technology of location does not operate as an iron frame leading always to identical results. For Jones, the three nations of this passage are Hamites, but for Coeurdoux they are Japhetites; moreover, for Coeurdoux the brahmins of India are of the descendants of Japhet called Magog, who had migrated to Central Asia, whence the brahmins migrated to India. He accounts for the similarity of the three languages not by co-descent from a single ancestor language, as in Jones, but by mutual borrowing among languages long neighboring one another, though originally distinct. One supposes the author means that God made these languages completely different from one another following the building of the Tower of Babel, and thereafter they grew similar because of their communications with one another.

From this passage we can draw several conclusions. First, both Father Coeurdoux and Sir William Jones, independently of one another, observed similarities among Sanskrit, Latin, and Greek through comparison, and sought for interpretations from the Genesis narrative of the Confusion of Tongues and the Dispersal of Nations, in short, from the Mosaic ethnology. Second, they located the nations speaking these lan-

guages differently, Coeurdoux making them co-descendants of Japhet, Jones making them co-descendants of Ham. The way in which the Mosaic ethnology is applied is thus underdetermined, and the outcomes of its application are not predictable, though both Coeurdoux and Jones place the three nations in a common descent line. Third, we come to the crucial move: from the genealogy of *nations* to the genealogy of *languages*. Here Coeurdoux and Jones again differ, showing two very different totalizing conceptions. Coeurdoux gives an explanation of language similarity through *mixture*, positing a movement from original distinctness toward similarity. Jones gives an explanation of language similarity through *co-descent*, positing a movement from original unity to difference—a movement that mirrors the movement of the Tree of Nations from generation to generation of patrilineal descendants. Jones applies the figure of the Tree of Nations directly to language as a model of language history, and by his doing so language history becomes a remedy and substitute for the lost memory of the history of nations. Language, like the DNA in our cells, contains, unknown to its speakers, the hidden history of the human race. Thus for Jones, and not for Coeurdoux, the shape of language history tracks the shape of the history of the nations. It is the explanation of Jones that became the foundation of comparative philology.

The reason why the ancient Indians, Persians, and Greeks did not discover the Indo-European language family is now clear. The underlying technology of location, the Mosaic ethnology, did not come from the ancient Greeks, Persians, or Indians, but from the Bible.

The question remains, however, why the Indo-European conception did not arise earlier among those who did have the locational technology of the Mosaic ethnology, the Peoples of the Book: Jews, Muslims and Christians. One would especially expect it to have been made by Muslim scholars when the expansion of Islam brought them into contact with various branches of the Indo-European family. One thinks, for instance, of the great scholar al-Biruni, who had lived and studied in India and knew both Persian and Sanskrit. The writings of Muslim historians did make abundant use of the Mosaic ethnology to give meaningful location to the many foreign peoples they encountered—a fact that was very consequential for Orientalist scholarship in British India, for it was largely through Persian, the language of diplomacy and learning for the Mughal empire, that the British acquired knowledge of Indian civilization in the early days. Muslim histories of India—and so also Muslim interlocutors of British scholars in British India such as Jones—used the familiar Mo-

saic technology to determine the ethnological location of India. For example, a history of India written in Persian by Firishtah was translated and published at Calcutta in 1768 by Alexander Dow; in it we learn that Japhet had sons named Turc, Chin, and Rus, from whom the Turks, the Chinese, and the Russians are descended, while the Indians are descendants of a patriarch named Hind, not named in the Bible but reckoned as a son of Ham, and that he had sons named Purib, Bang (i.e., Bengal), Decan (Deccan), Narwaal, and so forth, who founded the nations of India (Firishtah 1768:7–9). Similarly, in Abu'l Fazl's *Akbar nama*, Ham has sons named Hind and Sindh (both names of India), and Japhet, called the most just of Noah's sons, is said to be the ancestor of the Mughal emperors (Abu'l Fazl 1908, vol. 1, ch. 1; first English translation by Gladwin 1783–80). These texts, among others, would have reinforced Jones in his adoption of Hamitic ancestry for the Indo-Europeans.

In short, Muslims, Christians, and Jews shared the Mosaic framework for ethnological classification, and in that sense belonged to a common intellectual world. Our problem is to account for the fact that the application of methods for determining the historical relations among nations was applied to the problem of finding the historical relations among languages, and to explain why using language history as an index of the historical relations among nations appears only late in the day, and only among the Europeans.

COMPARATIVE VOCABULARY, OR THE METHOD OF THE WORD LIST

To see how language history became an index of the history of nations we need to zoom out still further and see Jones and the project of the “Anniversary discourses” in a yet larger context.

It is a curious fact that the famous text of Jones in the “Third anniversary discourse” does not display the means by which the comparison of languages was undoubtedly accomplished: the comparative vocabulary. The “Anniversary discourses,” he makes us understand, being oral performances to mark an occasion, were no place for the extensive recitation of evidence, and like any good lecturer Jones knew to present listeners with plenty of conclusions without troubling them overmuch with tedious evidence.⁶ But his sketch of the Indo-European conception

6. At the end of the eighth anniversary discourse Jones says that in the ninth discourse “I shall resume the whole argument concisely and synthetically; and shall then have con-

rested upon inspection and comparison of the languages in question, and we may be sure from every parallel of his day—and they are abundant—that such a comparison was given focus and direction by the means, simple but effective, of the word list: arranging the words of several languages in columns against a standard list of categories.

The simplicity of the word list lends itself to the empiricist view, against which I have been arguing, that the mere comparative inspection of languages leads to the apprehension of their historical relationships. It seems the word list merely orders the material of languages in ways that make comparison possible and are more user-friendly than older methods, such as the polyglot Bibles or polyglot collections of Pater Nosters that began to be published almost as early as the introduction of print in Europe. It takes an effort to see that the simple word list is not, after all, simple. The comparative vocabulary is not a neutral enterprise but an abstraction from living languages that freezes and organizes certain aspects of them for a certain purpose.

It is unfortunate for our purposes that Jones did not publish the word lists upon which his comparison of the Indo-European languages surely rested, for we would like to examine them at this point in the argument. We know for a certainty that they existed, and we see something of their content in the later work of Alexander Hamilton—a contemporary of Jones who learned Sanskrit at Calcutta and went on to become the first professor of Sanskrit in Europe, at the East India College in England (at Hartford Castle, later Haileybury)—particularly in Hamilton’s review of Franz Bopp’s important first work on the comparison of Indo-European languages (Bopp 1816; Hamilton 1820; see also Rosane Rocher 1961). We are obliged to look elsewhere for an example to analyze, but we can make a virtue of the necessity by turning to two important short papers, written at more or less at the same time as Jones’s “Third anniversary discourse,” by William Marsden, another member of the Asiatic Society and also in the employ of the East India Company.

William Marsden is chiefly remembered for his *History of Sumatra* (1783, 1966), his dictionary (1801, 1812a) and his grammar of Malay (1812b), and the collection of 3,447 coins he gave to the British Mu-

densed in seven discourses a mass of evidence, which, if brevity had not been my object, might have been expanded into seven large volumes with no other trouble than that of holding the pen; but (to borrow a turn of expression from one of our poets) ‘for what I have produced, I claim only your indulgence; it is for what I have suppressed, that I am entitled to your thanks.’” (Jones 1807, 3:185).

seum (cf. Marsden 1823–25). He contributed extensively to linguistics, but (as he observed bitterly) his philological work was better appreciated on the Continent than in England, where the brilliant reputation of Jones had thrown Marsden's work into the shade.⁷ How true this is can be seen in the *Dictionary of national biography* entry for Marsden (s.v.), which describes him as “orientalist and numismatist” and does not mention one of his greatest achievements, the first demonstration of the Malayo-Polynesian family of languages, and barely mentions his paper on the Indian origin of the Gypsies. These two accomplishments, which remain valid today, were realized by the method of the word list.⁸

Marsden shipped out to Bencoolen, Sumatra in 1770. On 5 March 1780, having just returned to England at age twenty-five, he wrote a letter to Sir Joseph Banks, president of the Royal Society, subsequently published as “Remarks on the Sumatran languages” (Marsden 1782). The centerpiece of that article is a vocabulary of fifty English words, against which are listed in columns the equivalent words in thirteen languages (see figure 4). Six of these languages are Sumatran: Malay, Achenese, Batta, Lampoo, Nias, and Rejang. The remaining languages are Javanese, Malagash (of Madagascar), Mongeray, Macassar, Savu, Tahitian, and Chinese, although Chinese does not play much part in the analysis. Marsden's purpose in drawing up this chart, he says, was, first, to

7. “About the period at which I first submitted to the notice of the literary world such information as a residence in Sumatra had enabled me to acquire on the subject of the languages spoken by the inhabitants of that and other of the Eastern Islands, it happened that the richer and more important mine of SANSKRIT learning had been opened by the labours of WILKINS and of JONES. To this the attention of persons who took an interest in Oriental studies was almost exclusively directed, and little encouragement was given to philological (though much to physical) researches in the maritime and less civilized regions of the East. In latter years, however, a disposition has been shewn, especially amongst our continental neighbors, to bring these languages within the scope of critical investigation, to examine their structure, their analogies, ascertain the extent to which they prevail, and, if possible, to deduce their origin” (Marsden 1834, art. 1, “On the Polynesian, or east-insular languages,” pp. 1–116).

8. Marsden's publications and manuscripts show abundant interest in the comparison of languages. The published works I may mention include: *A catalogue of the dictionaries, vocabularies, grammars and alphabets of all languages and dialects* (1796), *Bibliotheca Marsdeniana philologica et orientalis, a catalogue of works and manuscripts collected with a view to the general comparison of languages and to the study of Oriental literature* (1827), and articles on the languages of Polynesia (“The Polynesian, or east-insular languages”) and on the romanization of Oriental languages (“On a conventional Roman alphabet, applicable to Oriental languages”) in *Miscellaneous works* (1834), art. 1, pp. 1–116, and art. 2, pp. 1–27. Among the manuscripts of Marsden at the School of Oriental and African Studies of the University of London there is, for example, a collection of lists of fifteen numerals, common nouns, and adjectives in languages of all parts of the world (MS 12283–12304).

English	Malay	Acheen	Batta	Lampoon	Neeta	Reyng
One	Satoo	Sah	Salah	Sye	Semboua	Do
Two	Dua	Dua	Duo	Rowah	Dembooa	Dooy
Three	Tergo	Tioo	Tolo	Tuloo	Tuloo	Telloo
Four	Ampat	Paat	Opat	Ampah	Oopha	'Mpat
Five	Luma	Lumung	Loemah	Leemah	Leema	Lemo
Six	Anam	'Nam	Onam	Anemah	Oonoo	Noom
Seven	Toojoo	Toojoo	Paitoo	Pectoo	P'heetoo	Toojooa
Eight	Slappan	D'lappan	Ooalloo	Ooalloo	Ooalloo	Delaplan
Nine	Sambilan	Sakoorang	Seeah	Seewah	Seewa	Sembilan
Ten	Sapooloo	Saploo	Sapooloo	Pooloo	Phooloo	Depooloo
An hundred	Sa-ratoos	Sa-ratoos	Saratoos	Saratoos	Ogboo	Sotofe
Husband	Lackee	Lackaye	Merah	Cadjoon	Dongagoo	Sacky
Wife	Beenee	Beenaye	Aboo	Cadjoon	Seeslavee	Sooma
Father	Ilapa	Ilah	Ammah	Ilapa	Amah	Bapa
Mother	Mau	Mau	Ennah	Ennah	Eenah	Intoo
Brother	Sadarroo	Addooth	Ahhah	Adding	Tallephoofooa	Cacoon
Head	Capallo	Ooleu	Ooloo	Ooloo	Hugu	Ooloo
Eyes	Matro	Matra	Mahtah	Matlah	Huru	Matty
Nose	Edong	Eedoon	Aygon	Kerong	Eegho	Eeong
Hair	Ramboot	Oh	Oboo	Boobo	Boo	Boo
Checks	Pepee	Meung	Oroom	Bechun	Bo-ooah	Cubbole
Belly	Proot	Proot	Buttoobah	Tunnay	Talloo	Tennay
Hand	Tangan	Iarrooy	Tangan	Chooloo	Tanga	Tangoon
Legs	Cakce	Butteea	Paat	Binto	Apeh	Bettes
Garment	Badjoo	Badjoo	Ahben	Caway	Baroo	Badjoo
Day	Aree	Oocaye	Torang harree	Rannee	Loo-oh	Beely loong
Night	Mallam	Mallam	Borgning	Beengbee	Boong-ee	B. calemoon
White	Postee	Postee	Nabottar	Mandack	Aphooce	Pootoa
Sick	Itain	Itain	Nabecrong	Malloom	Aytoo	Meloo
Good	Ilave	Gaet	Dengan	Buttie	Songhee	Eaye
Use	Mittee	Mattay	Mahtay	Jahal	Mate	Mattooe
Fire	Appee	Appooy	Ahlee	Aphooy	Aleeroo	Opooy
Water	Ayer	Ayck	Ayck	Ayck	Eedano	Beole
Earth	Faza	Tano	Halla	Tanno	Tano	Peera
People	Orang	Oreecing	Halla	Ooloon	Neeha	Toon
Cocnut	Clappo	Oo	Cranchce	Clappah	Bunneco	Neole
'Terah	Gejoo	Gegoy	Ningee	Eclan	Eephoo	Aypen
Hog	Babe	Babe	Ilabee	Babooye	Ilavee	Sooecetmba
Bird	Boorring	Cheechin	Peerah	Boorring	Foopho	Benono
Egg	Telloot	Boo	Peedong, Mannok	Tullooy	Adoolo	Tennole
Fish	Eecun	Incoor	Deklay	Eerah	Eerah	'Conn
Rice	Ilray	Breeagh	Duhans	Beez	Booragh	Blas
Potatoes	Oslee	Gadong	Gadong	Cutillah	Gorvee	Ooby
Son	Matto Aree	Mattowaye	Marah harree	Matra ronce	Secno	Mattu beely
Moon	Boolan	Booloon	Boolan	Boolan	Boon	Booloon
Stars	Beentang	Beentang	Bintang	Bintang	Doothee	Beetang
I	Anbo, Sayo	Oooloon	Apoo	'Gnah	Beoo	Ookoo
Yes	Foo	Nyoh	Olo	Eer	Eh	Aou
Come hither	Marce feenee	Jah knyoi	Maré tofone	Eja dejah	Einee undeh	Conceendi
God	Allah taliah	Allah	Dabattah	Alla Yella	Loma langee	Oola tallo

Figure 4. William Marsden's comparative vocabulary of twelve Malayo-Polynesian languages and Chinese. (From Marsden 1782: facing p. 154, table 1.)

trace, if possible, a common origin among these far-flung languages, and, second, to determine whether the languages spoken by the “various independent and unconnected nations” of the interior of Sumatra are radically different, “as is generally supposed by the Europeans resident there,” or are different dialects of a single language.

The vocabulary appears to be a mere list, and its evidence is offered modestly by the young author, with but a minimum of analysis, in deference to what the wiser heads of the Royal Society will make of it. But the list is, in fact, carefully constructed and has a theory behind it. It

Javan	Malagash	Moogeraye	Muraffir	Sava	Orabete	Chineſe
Seejee	Eranke	Eftakoo	Sayden	Uffe	Atahay	Chieed
Roro	Dooe	Lolaye	Dooa	Rooe	Erooa	No
Tullooo	Teloo	Loolectoo	Tullooo	Tullooo	Tofficoo	Sinh
Papat	Ephat	Lopah	Pa-me	Uppa	Atlas	See
Leemo	Leemoo	Leemo	Leema	Luuſoo	Ereema	Go
Nanam	Enema	Daho	Atan	Unna	Aono	Lacq
Pectoo	Pheetoo	Pheetoo	Pectoo	Pectoo	Abectoo	Shit
Oloo	Valoo	Apho	Arrooa	Aroo	Awarroo	Peel
Sanga	Seevee	Seeva	Affarra	Saio	Aeeva	Caow
Sapooloo	Phooloo	Tooroo	Sapooloo	Singooroo	Ahooroo	Chap
Satoo	Tatoo		Sangatoos			Chepe
Lanang	Lake lake	Namee	Earonee		Tane	Ang
Oodone	Ampela	Jah	Makoonraye		Huabeine	Poh
Paman	Appa	Iapa	Ambo		Medooa tane	Enteeah
Berang	Nenay	Mau	Endo		Med, wahine	Neha
Sadooloor	Ranowla	Noko	Sadjeo		Teine	Suotee
Eodaf	Loobah	Jahé	Ooloo		Oupo	Toucah
Matta	Meffoo	Nana	Mattaye	Madda	Mazau	Bixu
Eerong	Oorong	Mernee	Eengana	Sivanga	Eahoo	Peel
Ramboot	Vooloo	Jahé	Gummomna	Rov	Eraowroo	Toumo
Peepce	Takoolaka		Dowcheelee	Cavaranga	Paparea	Shreepy
Ouatrang	Keeboo	Ataliba	Allay	Dullooo	Biooo	Pooto
Lungan	Feletanan	Tanaraga	Tapah-lamay	Wahleat	Ereema	Tehoo
Seckel	Toongoota	Eetee	Aajungna	Bubo	Awy	Calcoot
Calumbee	Ahanzoa	Moortana	Cabudja	Cova	Ahoo	Sant
Deena	Hareeanroo	Oofa	Afo		Milana	Iet
Oongee	Haree Vah	Gumoo	Bumnee-ee		Eaoo	May
Pootee	Phootee	Bootee	Pootee		Tea	Pay
Eerung	Minetee	Metna	Lotong	Bulle	Ere ere	Oh
Saye	Soooh	Rov	Maclechang		Myty	Hoh
Mattee	Mattce	Hommoo	Mattce		Mattce	See
Geennee	Aphoo	Arta	Appee		Wahaa	Whory
Bannee	Rano	Eera	Oonye		Avy	Choee
Lumma	Tana	Tazo	Humoo	Monome	Fenooa	Toh
Wong	Ooloo	Anoooo	Taow		Tara	Lang
Clappo	No Word	Conta	Calookoo		Taro - Arce	Een
Oootoo	Neepee	Ootfee	Affimna		Eerebee	Cheekce
Cheling	Lunboo	Iaye	Pahee	Varre	Boesh	Tee
Mano	Voorong	Oio	Mahoomanoo	Doolia	Mahoo	Chooa
Endo	Atoodee	Affowa	Tellu	Dullooo	Auuro	Noocy
Eewa	Phceah	Appee	looco-edja	Nudoo	Eya	Hee
Bras	Varay	Reia	Era	Arree		Be
Cuffela	Dovee	Janmais	Oober		Oonarruh	Whunchee
Surungee	Malloo anroo	Oootoo	Mitangallo	Warroo	Malana	Jerrou
Oolan	Voo'an	Ooroo	Ooiaag		Marana	Gooly
Ooentang	Vintan	Eepce berry			Ebitoo	Schyo
Coula	Zaho	Anee	Eetee		Waww, Mec	Gooa
Inghee	No word	Eoo	Fey-ma	O	Al	Hoh
Marenee	Avech	Maoo	Sakoonaye		Harre mai	Lajeno
Dewah	Dernakaree		Allah tallah		Patna	Teehn tay

consists of words “of universal use from the nature of the ideas they express,” namely, the numbers from *one* to *ten* and *one hundred*; kinship terms (*husband, wife, father, mother, brother*); parts of the body (*head, eyes, nose, hair, cheeks, belly, hands, legs*); garments; *day* and *night*; *white* and *black*; *good*; *die*; the elements (*fire, water, earth*); *people*; foodstuffs (*coconut, hog, bird, egg, fish, rice, potatoes*—with the word *teeth* randomly stuck in); heavenly bodies (*sun, moon, stars*), and a miscellany (*I, yes, come hither, God*). The vocabulary, then, is not a random sample of the abundance to be found in a living language or a dictionary but is comprised of what are conceived to be the simplest, most primitive and necessary conceptions that languages must name at their very creation—words which constitute, for that reason, the native core of those languages.

From this simple-seeming device Marsden then abstracted the most obvious cognates, constructing a second list that reversed the order of the original word list, putting the cognates first and the English glosses second. These “examples of words in the Sumatran and neighbouring languages[,] corresponding in sound and significance with others in places remote from thence” (see figure 5) and consisting of about half the words of the original vocabulary list, display correspondences across languages from Madagascar to New Zealand and Easter Island. Marsden, with a show of diffidence, says that though he has been only mildly successful in tracing a common origin among the languages in question, the only general inference that can be drawn is that “from Madagascar eastward to the Marquesas, or nearly from the east coast of Africa to the west coast of America, there is a manifest connexion in many of the words by which the inhabitants of the islands express their simple ideas, and between some of the most distant, a striking affinity.” This is the very first statement of the bounds of the language family we now call Malayo-Polynesian.

Marsden goes on to suggest that Central Asia (“Tartary”), that *officina gentium*, or womb of nations, may have been the source of the peoples of the archipelago, and a knowledge of Siamese, Lao, Cambodian, and Peguan—languages lying between Central Asia and the islands whose languages he has been studying—would be the “readiest clue to a discovery of that kind.” A few decades later John Leyden, also of the East India Company, published two short works on the languages of the Indochina Peninsula, as if in response to this expressed need (Leyden 1810, 1812). The theory of a Central Asian origin of these languages—a Japhetic origin, in Mosaic terms—did not prove useful. Nevertheless, Marsden had accomplished a basic demonstration of the relatedness of many of the languages making up the Malayo-Polynesian family, mostly from materials collected by himself, as he says, from the mouths of natives (excepting those for Savu and Tahitian) and not from books. It was a stunning accomplishment, equal in every way to Jones’s proposal of the Indo-European concept.

Marsden’s discussion, brief as it is, throws much light on the underlying assumptions of the whole enterprise of collecting words against a standardized word list. Regarding the second question he had set out to examine, the relationships among the languages of the Sumatran interior, he regarded his table as showing them to be dialects of a single language, against the view prevailing among Europeans there, and among the Sumatrans themselves, that the languages were unrelated. To be sure—

and here Marsden gives the arguments of the opposing view—the fact that several of these languages had writing systems all their own seemed to prove that the languages were of different origins and that the similarities among them had been produced by the borrowing of words from one another. The different Sumatran peoples do not comprehend one another's language, are different in their persons, and their manners and customs are “as unlike as those of the most distant nations.” But, Marsden replies to his own objection, the words they have in common “are radical and such whose correspondent ideas must have existed and been described prior to all intercourse with either remote or neighbouring people; as will appear from an inspection of the comparative specimens, and consequently that the dissimilarity, not the similarity, must have been induced by degrees” (Marsden 1782:156). The critical aspect of the proof is that the vocabulary is made up of words that are radical, that is, root words native to each language.

In this passage we see Marsden considering the two great alternatives of interpretation for similarity among languages: *mixture* of languages through borrowing versus gradual *differentiation* of languages sharing descent from a common original language. He decides that language similarity is, on the whole, the vestigial evidence of co-descent and not the product of linguistic mixture and convergence. This is the very issue that separates Jones and Coeurdoux in the interpretation of the resemblance among Indo-European languages, Jones arguing (correctly, in this case) for co-descent and Coeurdoux arguing (wrongly) for convergence due to borrowing. The issue is inescapable in any comparison of two or more languages for purposes of determining historical relations, and the method of the word list is a kind of technology for getting at relations of co-descent. It does this by what it aims to include—radical (root) words of the language, as Marsden puts it. But it also operates by what it excludes from the list, namely, those words corresponding to more complex ideas, words of art and science that develop late through commerce with other nations and the combination of simple ideas into complex ones. The philosophical distinction between simple and complex ideas, as well as Locke's conception of words as names of ideas and his developmentalist, progressive notion of knowledge (Aasleff 1982), are the theoretical underpinnings of the simple vocabulary list as a strategic method for uncovering relations of co-descent.

The second paper by Marsden, “Observations on the language of the people commonly called Gypsies” (1785), offers further testimony to the efficacy of the comparative vocabulary as a tool by which reliable new

Malay Achenefe Batta Neas Rejang Javancie	Mattee Matray Mahtay Mate Mattooeo Mattee	Die, Dead	Otaheite Garageco Madagascar Bugguefs or Macassar	Matte Mattee Mattee Mattee
Malay Achen. Batta Rejang Lampoon Javan.	Matto Matta Mahtah Matty Matrah Matta	Eyes	Otaheite Savu Garageco Bugguefs Easter island Marquefas Amsterdam N. Zealand Malicolo	Matta Madda Matta Mattaye Matta Matta Matta Matta Maitang
Malay Achen. Batta Neas	Babbee Booy Babee Bavee	Hog	Otaheite Marquefas Amsterdam Savu Mongeraije Malicolo Tanna	Booa Booa Booacha Vavee Baye Brooas Boogas
Batta Lampoon Javan.	Manook Manoo Mano	Bird, Fowl	Otaheite Easter Island Amsterdam Tanna N. Caladonia Bugguefs	Manoo Manoo Manoo Manoo Maneck Manoo manoo
Rejang Neas	Neole Bunneco	Coconut	Amsterdam Chinefe N. Caladonia	Eeoo Eea Neco
Malay Neas Rejang Javancie	Eecun Eeagh Eewah Eewah	Fish	Otaheite Easter Island N. Zealand Amsterdam Madagascar	Eya Eeka Eeka Eeka P'heeah
Malay Achen. Batta Lampoon	Ayer Eer Ayck Wye	Water	Bugguefs Mongeraye Otaheite Easter Island N. Caladonia	Ooaye Eera Avy Evy Ooee
Malay Neas Rejang	Oobee Gooee Ooby	Potatoes and Yams	Otaheite Easter Island Amsterdam N. Caladonia Madagascar	Eoobe Oche Oobe Oovee
Malay Lampoon Javan.	Orang Oooloon Woug	Man Perfon People	Chinefe Madagascar	Lang Ooloo
Achen. Lampoon	Appooy Apooy	Fire	Madagascar Chinefe	Aphoo Whooce

Figure 5. William Marsden's list of cognate words in Sumatran and neighboring languages. (From Marsden 1782: facing p. 154, table 2.)

Malay Achenefe Batta Neas Rejang Lampoon Javanefe	Duo Dua Duo Dembooa Dooy Rowah Rota	Two	Otaheite Savu Madagafcar Bugguefs Eafter Ifland Marquefas Amfterdam Tana	Erooa Roee Doocce Dooa Rooa Aooa Eooa
Malay Achen. Batta Neas Rejang Lampoon Javan.	Teego Tloo Toloo Tuloo Telou Tulloo Tulloo	Three	Otaheite Savu Madagafcar Bugguefs Eafter Ifland Marquefas Amfterdam	Torhoo Tulloo Teloo Tuloo Toroo Atoroo Toroo
Malay Achen. Batta Necas Rejang Lampoon	Ampat Paat Opat Oopha Mpat Ampah	Four	Otaheite Savu Madagafcar Eafter Ifland Marquefas Malicolo	Atta Uppa Ephat Faa Afaa Ebat
Malays Achen. Batta Neas	Toojoo Toojoo Pailoo Pheetoo	Seven	Otaheite Savu Madagafcar Mongeray Marquefah	Aheetoo Pectoo Pheetoo Pheetoo Acheettoo
Malay Batta Neas	Sambilan Secah Sewah	Nine	Otaheite Savu Madagafcar Eafter Ifland	Aeeva Saio Seevec Heeva
Neas Lampoon	Adooloo Tulloy	Egg	Savu Bugguefs	Dulloo Tello
Malay	Papatcel	A Tool	N. Zealand	Patoo patoo
Malay	Telingo	Ear	Malicolo	Talingan
Malay	Tapa	Sole of Foot	Otaheite	Tapooy
Javan.	Eerung	Black	Otaheite	Ere ere
Malay	Momotong	Cut	Otaheite	Motoo
Malay	Paya	Fatigued	Otaheite	Paya Faeca
Neas	Taloo	Belly	Savu	Dulloo
Malay Achen. Neas Rejang Lampoon Javan.	Bray Breeagh Booragh Blas Beas Bras	Rice	Savu Madagafcar Bugguefs Chinefe	Arre Vary Bra Bea
Malay	Bapa	Father	Otaheite	Papa

Comparisons of the Gypsey and Hindostanic languages.				
	English Gypses.	Turkish Gypses or Chingharis.	Cingari vel Eronnes Nobiani.	Hindostanic.
One	Aick, yek	Yock		Aick, ev, yek
Two	Dooce	Doy		Do, daw
Three	Trin	Trin		Teen
Four	Sow, flaur, fluzr	Shtier		Chaur
Five	Pange	Panch		Paunch
Six	Shove	Shove		Chaye, chey
Seven	Heftau	Efta		Saib, fait
Eight		Oktô		Aath, aut
Nine	Henrya	Enia		Noh, no
Ten	Defli	Defli		Dus, doll (Bengalî)
Man	Râpe, gajo	Rom, manush	Manush	Maandho (Bengalî); Mannos (Mahratta)
Woman or Lady	Raunte, gaujet	Roncee		Renlee, raunte
Head	Bol-shuroo, sharo	Shero	Scheiro	Seer, firr
Eyes	Yack-aa, yock	Yack	Jaka	Aunk, choke, shâyo, (Bengalî)
Nose	Bol-nok	Nack	Nak	Natck
Hair	Balla, lolau	Bal	Bal	Baul, bol
Teeth	Dan-ru	Dan		Dhant, doort, (Bengalî)
Ear	Kamana		Can	Kaun
Good	Quitto	Lutchi		Acha
Day	Deras, detas, devas,	Deevas		Deru, derwat, (Mahratt)
Night	Rautee	Ratee		Raut, raat
White	Pakno	Pancee		Paandra, (Mahratta)
Black	Kaulo	Caglee		Kaulla
Fire	Yaug	Yagg	Vag	Ag
Water	Paunce	Pagnee	Pani	Paunce
Dead	Mullo, mouly	Moolo		Moolah, maylay, (Mahratta)
Cow	Gouvinee		Carceeni	Gauee
Sheep	Baukro		Bakro	Biare
Hog	Bauo	Bâo	Falo	
Fish	Matcho	Matcho		Matcher, matchot
Bird	Chersak, chillakoo	Chirakio		Chersa
Knife	Kair		Ker	Ghair
Moan	Chooere			Chooere
Salt	Leos	Choun	Chou	Chound
Gold	Soma-kai			Nooor, (Bengalî)
Silver	Reop			Souna
God	Me-devel	Deriaa		Roopou Dawa, (Bengalî)

Figure 6. William Marsden's comparative vocabulary of Romani and Hindustani. (From Marsden 1785: 386.)

scientific knowledge could be achieved. From a specimen of words of the Cingari, or Gypsies, in a history of Ethiopia by Ludolphus, “which he had collected from these people in his travels, with a view of determining their origin,” Marsden was surprised to find many words familiar to him from his knowledge of Hindustani. The similarity seemed so extraordinary, he says, that he suspected an error in the publication, “which might have arisen from a confusion of obscure vocabularies in the author’s possession,” but he verified its accuracy by gathering further vocabularies from Gypsies in England and, through a correspondent, in Turkey. He displayed the results in a table using a vocabulary very similar to that of his article on the Sumatran languages, but slightly shorter (39 words) (see figure 6). He observes also that some of the Gypsy words correspond to words in Marathi and Bengali, and that “it is not a little singular that the terms for the numerals *seven*, *eight* and *nine*, are purely *Greek*, though the first five, and that for *ten*, are indisputably *Indian*.”

Since the name “Gypsy” was a corruption of “Egyptian,” some scholars thought the Gypsy language might be traced to Coptic, a view which Marsden was able to disprove: only a single word bore a resemblance to Coptic. Thus in a second brief article, Marsden had established the Indian origin of the Romani, or Gypsy, language, by means of the vocabulary list. More or less at the same time Jones stated, on the basis of Grellman’s recently published dictionary, that the Gypsy language “contains so many Sanscrit words, that their Indian origin can hardly be doubted,” giving a list of such words “pure Sanscrit scarce changed in a single letter” (“The eighth anniversary discourse,” delivered 24 Feb 1791, Jones 1807, 3:170–71).⁹

These three accomplishments of the 1780s, the first formulation of the Indo-European concept by Jones, the first formulation of the Malayo-Polynesian concept by Marsden, and Marsden’s identification of the Indian (or as we now say, Indo-Aryan) origin of Romani, share a number of characteristics with one another and with a fourth publication, also associated with the East India Company, that is the central object of this book: F. W. Ellis’s 1816 proof of the unity and non-Sanskritic origin of the South Indian languages, now called Dravidian. Each of these discoveries, as also others like them in the eighteenth and early nineteenth centuries, was truly new and unexpected, revealing historical connections among languages for which there was no historical memory that have proven sound and remain valid today.

The word list, then, has been a deceptively simple, highly effective tool for eliciting historical relations among languages as indices of historical relations among nations. How did it come about that European travelers to Asia and elsewhere took with them a more or less standardized word list, with a view to participating in the great revolution in ethnological knowledge that resulted from its use? No doubt the sources of this program were many, but a short text by G. W. Leibniz, published in 1718, shortly after his death, is of critical importance and may well be the model for subsequent lists, including Marsden’s.

The text in question is Leibniz’s “Appeal concerning languages of

9. According to the 11th edition of the *Encyclopaedia Britannica*, s.v. “Gipsies”: “Ruedinger (1782), Grellman (1783) and Marsden (1783) almost simultaneously and independently of one another came to the same conclusion, that the languages of the Gipsies, until then considered a thieves’ jargon, was in reality a language closely allied with some Indian speech.” Jones identified in Grellman’s list Romani words corresponding to Sanskrit *aṅgāra*, “charcoal,” *kāṣṭha*, “wood,” *pāra*, “a bank,” and *bhū*, “earth.” His contribution to the question of Roma origins has not been previously noted.

peoples,” a letter of inquiry he sent to the imperial interpreter.¹⁰ It begins: “As nothing throws greater light indicating the ancient origins of peoples than the collation of languages, I often wonder that geographers and travelers neglect to write of languages, and but rarely exhibit specimens of them.” His request for information amounts to instruction to travelers about how such linguistic specimens should be collected (Leibniz 1718:49). He asks for collections of Pater Nosters in foreign languages, a well-established practice—he tells us that we have examples of the Pater Noster for the languages of the Poles, Serbs, Dalmatians, Croats, and Russians, all of the Slavonic (i.e., Slavic) family, and of the Wallachians, Cettos and Livonians, Turks, Persians, and Chinese. He asks for texts with interlinear translations into languages known to Europeans. But he also asks—and this seems to be new—for “a few examples of their words, expressing common things,” which he spells out in detail:

Names of numbers: *one, two, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 100, 1000.*

Relatives and ages: *father, mother, grandfather, son, daughter, brother, sister, father’s brother, husband, wife, father-in-law, son-in-law, man, woman, child, youth, old man.*

Parts of the body: *body, flesh, skin, blood, bone, head, brow, nose, eye, pupil, ear, beard, mouth, tongue, tooth, chest, heart, throat, jaw, foot, finger, hair, belly, breasts.*

Necessities: *food, drink, bread, water, milk, wine, herbs, fruit, salt, fish, ox, sheep, horse, clothing, hide, house, wagon, sword, bow, arrow, lance, slingshot.*

Naturalia: *god, man, heaven, sun, moon, star, air, rain, thunder, lightning, cloud, frost, hail, snow, ice, fire, heat, light, smoke, earth, field, mountain, valley, sea, river, stone, sand, dog, wolf, deer, bear, fox, bird, snake, mouse.*

Actions: *to eat, to drink, to speak, to see, to be, to stand, to go, to strike, to laugh, to sleep, to know, to pluck, and so forth.*

It is striking that Marsden (who does not mention Leibniz or indeed any prior authority for his list) uses much the same, though not identical,

10. I am grateful to Hans Aarsleff for drawing my attention to this work, which is discussed in his book, *From Locke to Saussure* (1982), a work of profound scholarship. On this text of Leibniz, see also Gulya 1974.

categories in much the same order: numbers, kinship terms, parts of the body, necessities, naturalia. Other of the innumerable examples of the comparative vocabulary bear a family resemblance to this one, and it is possible that it is the ultimate source for them all.

Leibniz's formulation of the vocabulary list was contained within a definite program, as we can make out from the terms of his request. As a northern European he seeks information on the northern peoples who lie to the east and who, since much of Asia and part of Europe received colonies from such northern peoples, he considers most worth knowing yet are known least of all. He especially wants to explore the relations between the Germanic and Slavonic, or Slavic, language families to determine whether there are other language families that belong to neither. He seeks, for example, linguistic specimens from Transylvanian Saxons "not as cultivated men speak, but as the common people speak, that they may be compared with the language of our Saxons. It is said that there are among the common people many words neither Hungarian, nor Slavonic, and moreover which generally cannot be understood by other Germans." He also asks for information about an enclave of Germans in Crimean Tartary, "or rather of the ancient Goths," who, according to a report by an imperial ambassador to the Ottoman court, use a dialect of German. He requests information about non-Slavic languages of the Muscovite Empire: the languages of Siberia, the Black Sea, and other places. He wonders whether records can be had about a certain Hungarian Jesuit, made captive in Tartary and sold to barbarians beyond the Caspian Sea, who discovered that the language there was related to Hungarian. He asks whether in Albania and Bulgaria there is to be found any language completely different from Slavonic, Hungarian, Greek, and Turkish. The point of all these questions is evident: his program is to map the entire ethnological history of Eurasia through its languages and to determine the place in that history of the German language and people.

In the course of this brief letter—less than three pages in print—we see that Leibniz has engineered into, and out of, his word list a number of characteristics: the language of common people, not the learned; words that are necessary for the immediate needs of life, not the more complex notions of art and science; primitive words rather than recent ones; simple rather than complex words; native words rather than foreign borrowings. Those are the attributes of the native core of a language that the vocabulary seeks to capture.

It follows that the conception of the language is an abstraction from the living language; a thick abstraction, perhaps, but an abstraction nev-

ertheless, that is formed by identifying its core and removing from consideration the late, learned, foreign, borrowed, complex accretions of later ages. The method of the word list constitutes in its seeming simplicity the first, surgical move of historical linguistics: the cutting away of the later, borrowed, and complex accretions to reveal the native core of language, so that the operation of comparison can be performed on the authentic body of the language. This allows the historical relations among languages to be figured as the radiating branches of a tree, since the borrowings or mixtures that would make the branches grow into one another have been discarded by analysis. It is well to keep in mind the conception of language that undergirds the genealogies of languages in historical linguistics.

We are now in a position to answer the question why it was in eighteenth-century European thought, and not earlier in the Christian West nor among Muslim scholars of the great age of the expansion of Islam, that the twinning of languages and nations took such a decisive turning, thereby completely rewriting the deep history of the world. It was in this time and context that a method was devised—the word list—that could put the posited close relation of languages and nations to work, such that language relations could be elicited even in the absence of historical memory and so serve as a key to ethnological relations, that is, the history of nations.

Initially, comparative vocabulary showed the way to the ancient, native core of languages. Comparative grammar followed, offering, if anything, a surer way into the native heart of languages. Already in Jones we see vocabulary *and* grammar providing evidence of historical relatedness among languages, and the comparative grammar of Indo-European was systematized by Franz Bopp, who made it his life's work.

The pairing of languages and nations or races, a strong tendency in biblical thought and in European thought from late antiquity, became intensified only at the moment when it was given a rational, scientific method, the method of the vocabulary. The application of this seemingly simple means in an age of worldwide expansion of European power yielded astonishing results, whose very durability tells us that they hold a great deal of truth. The presumption upon which this pairing rests—that the native core of a language is intimately bound to the nation that speaks it—was unquestioned in the age of its invention. It was only later, in the mid-nineteenth century, that one begins to hear a growing chorus of voices asserting, as a new and surprising truth, that race and language do not necessarily coincide with one another (see chapter 7). Hitherto, for millennia, the assumption had been that they did.

GRAMMAR MANIA

Examining the text of Leibniz, we come to see that behind what appears to be a *mere* word list is a formative theory that determines what gets included in the list and what does not. The word list, with its names for numbers, kinship terms, parts of the body, forces of nature, and objects of common use, incorporates within it a conception of the primitive core of a language, as distinguished from its later accretions, and sets up a series of binaries (primitive/recent, native/foreign, common/learned) that are used to accomplish the first move of the emerging analytic: the identification of the primitive core and the removal, for purposes of analysis, of everything that is not part of the core.

I say again that this method has been highly effective, astonishingly so and out of all proportion to its low level of complexity. Even before a satisfactory and unitary scientific standard of etymology had been found through the recognition of lawlike regularity in sound shifts within languages, beginning with Grimm's law in the early nineteenth century, the word list was the means to new discoveries that have endured. Simple as its form is, outmoded as is the biblical frame of historical time within which it was conceived, lacking as it does a method of determining which similarities count as signs of relationships and which do not, or a method of discerning relationships among words that do *not* look similar yet are cognate (the greater analytic power achieved when comparative philology arrived at the laws of sound shifts)—despite all these shortcomings, the method of the word list, even in its simplest, eighteenth-century form, must be close enough to the truth of things to have been able to deliver so much knowledge that was new and lasting.

The word list is the middle term that allows for the complete assimilation of language to the terms of the Mosaic ethnology, permitting languages to be treated as kinsmen in the branches of a segmentary lineage and as growing ever more different and distant, like lineage segments of the Nuer or Arabic-speaking Muslim lineages in Yemen (Dresch 1988, 1989) or Libya (Davis 1987).

Jones presented his pioneering statement about the Indo-European languages not as an end in itself, as a finding about languages, but as an objective means to make up for defective historical memory. For Jones, the three faculties of the mind were *memory*, *reason*, and *imagination*—the respective sources of history, science, and the arts (Jones 1788b:xiii). History, then, is founded on memory, which is to say that to know the history of India one must consult the historical memory of the Indian na-

tion as captured in histories written by Indians. Jones and his British colleagues had sought Indian history in the texts called *purāṇas* (antiquities), and one of the earliest texts composed in Sanskrit at British request, written before Jones arrived in India, was a synopsis of the *purāṇas*, titled *Purāṇārthaprākāśa*, by Radhakanta Sharma (see Ludo Rocher 1986, text published in Rocher and Rocher 1994–95). Yet the high hopes with which the British pursued history through the *purāṇas* were only partially fulfilled. Jones himself believed that the *purāṇas* contained a genuine recollection of the flood of Noah and thus corroborated the truth of the Bible, for example. But the *purāṇas*' genuine memory of civil history in India prior to about 200 B.C., in his view, "is involved in a cloud of fables" (Jones 1788d:421). To make good this defect, then, four means could be introduced from outside the national memory, namely, the study of languages and letters, the study of philosophy and religion, study of the remains of old sculpture and architecture, and study of the written texts of the sciences and arts. It is in the context of the first of these that Jones made his famous statement about Indo-European; that is, language comparison functions as an exterior means by which to recover a history obscured by "fable." If the method was useful in India, with its long and ancient literary tradition, how much more useful it must be as a way of finding history in places where there was no writing in which memory could be made fast.

Such a program for recovering the history of the nations of North American Indians had been proposed before, in 1744, by the Jesuit missionary Francois-Xavier de Charlevoix. He argued that previous writers had compared the morals, customs, religions, and traditions of the American Indians with those of Old World cultures to determine their origins but had neglected the one means that would solve the problem: comparison of their languages. For languages change but slowly and remain distinct from one another, so that if the American languages could be shown to have the features of the mother languages, they could be traced back to the creation of different languages at the Tower of Babel and so prove that America had been peopled by the great-grandchildren of Noah (Charlevoix 1994, 1:153–54). This program was restated by Thomas Jefferson in *Notes on the state of Virginia* (c. 1782) and put into effect by means of a printed vocabulary, similar to the one of Leibniz, which may still be seen in the library of the American Philosophical Society, against which were recorded the words of different Indian languages in columns. This project was carried forward and completed by Stephan Du Ponceau,

protégé and successor of Jefferson in the presidency of the American Philosophical Society (Du Ponceau 1838).

At the same time as Jefferson and Du Ponceau were investigating the history of the New World nations through the history of their languages, Catherine of Russia, with the help of a vast imperial bureaucracy, was collecting a comparative vocabulary of her empire, a vocabulary she hoped would encompass the whole world. Here again the word list had a Leibnizian form, and the result was published by Pallas in 1786–89, in the two-volume work grandly called *Linguarum totius orbis vocabularia comparativa*—at the very moment, one might say, that Jones was proposing the Indo-European concept before the Asiatic Society in Calcutta. Thus the project of Jones was not at all singular but was part of a broad movement that, because of the worldwide expansion of European power in the eighteenth century, reached the far corners of the earth.

By means of the theory encoded in the structure of the comparative vocabulary, languages and nations or races were understood as being so closely identified with one another as to have no gap between them—like adjacent stones in one of those cyclopean walls of the Incas or the Pyramid of Cheops. Every statement about the relations among languages was a statement about the relations among nations. This was the first effect of the full assimilation of language to the locational technology of the Mosaic ethnology.

The second effect was the radical redefinition of what a language is. Eighteenth-century European discussions still had room for distinctions between stock, or primitive, languages and mixed languages formed from the blending of stock languages. But the logic of the word list tended away from that kind of thinking and toward identification of the primitive core as constituting the authentic language itself. Though mixture is an aspect of every natural language, for a natural language to enter into the historical-linguistic tree, it must first give up its borrowed clothes and get naked. Much as ancient genealogists learned to sculpt the relations of kinship to form beautiful, treelike branching structures by throwing away all relations of descent and marriage except those that pass exclusively through males, modern linguists could map languages onto such structures only after the languages had been thinned down, carved out, and abstracted from all that time, proximity, and communication do to make neighboring natural languages, such as the languages of India, resemble one another. It will be well to keep in mind the highly artificial nature of both these effects, and to remember that they flow from the word list and

its theory. While in the eighteenth century there was considerable vacillation between mixture and genealogy as the explanation of similarity among languages, the inexorable tendency, or at any rate the tendency that won out, was the latter, epitomized in the slogan of the Young Grammmarians of the mid-nineteenth century: there is no such thing as a mixed language (*Es gibt kein Mischgesprach*). The audacity of this statement is breathtaking, for in truth there is no such thing as a pure language, a language that is not mixed, if we are speaking of natural, actually existing languages spoken by human beings rather than artificial languages abstracted from them for purposes of comparative study.

It has become customary to the point of cliché, when examining the pairing of languages and nations in European thought, to invoke the words and thoughts of Herder. The idea that Herder is at the bottom of this has been overdone, and we need to focus, rather, on Locke and Leibniz.

Locke's *Essay concerning human understanding* (1689) discusses language at length but not languages, *langue* but not *parole*, and it did not generate a program for research into the world's languages. But Leibniz's *Nouveaux essais sur l'entendement humaine* (1765, published posthumously; see Leibniz 1981), which was inspired and provoked by Locke and is a sort of commentary on Locke's book in the form of a dialogue between two speakers, one a surrogate for Locke and the other for Leibniz, does develop such a view and, indeed, imagines the explosion in the grammar factory before it occurred. It is useful, therefore, to take the two together. Briefly, what Locke's view of language supplies to Leibniz is the idea that a language is not created once for all time by God but is an historical entity growing and becoming progressively more copious in vocabulary as it develops complex ideas through the cumulation of experiences of nature and commerce with other nations. The growth of ideas from simple to complex is tracked by the growth of language from rude to polished, savage to refined.

The following is what Leibniz builds upon that Lockean base: "Languages in general, being the oldest monuments of peoples, earlier than writing and the [practical] arts, best indicate their origins, kinships and migrations." It is for this essentially ethnological reason that it is useful and important to study etymology. But in doing so one must show the relationships among various peoples and should not make leaps from one nation to another remote one without solid confirming evidence, especially that provided by intervening peoples. Etymologies should not be trusted without a great deal of corroborating data; to do otherwise is to "goropize," that is, emulate the absurd etymologies of the Dutch scholar

Goropius Becanus, who proclaimed Dutch as the original language (Leibniz 1981:285). When European scholars have used up the ancient writings of the Romans, Greeks, Hebrews, and Arabs, there will be the Chinese, Persians, Armenians, Copts, and Brahmins to supply materials for further progress. “And when there are no more ancient books to examine, their place will be taken by mankind’s most ancient monument—languages. *Eventually every language in the universe will be recorded, and contained in dictionaries and grammars; and comparisons will be made amongst them*” (emphasis added). Thus does Leibniz predict the explosion in the grammar factory. The comparison of languages will extend the genealogy of nations beyond the reach of written histories. Study of foreign languages will be useful for the study of *things*, “since their properties are often reflected in their names (as can be seen from the names of plants among different nations)” and for the knowledge of the *mind* and the variety of its operations, but above all it will also be useful for searching out the origins of *peoples*, through etymologies, which can best be ascertained through language comparison (Leibniz 1981:336–37).

By the end of the eighteenth century, as Leibniz predicted, a veritable explosion in the grammar factory had occurred in Europe, by which I mean an impulse to blanket the world in grammars and dictionaries. This impulse was indeed borne around the globe by mercantile and imperial expansion. The explosion in the grammar factory is still underway today. We continue to think it obvious that languages, whether written or not, have a grammar, and that linguists would want to write grammars and dictionaries for every language in the world and would give their books titles like *Les langues du monde* (Meillet 1952)—a late echo of the *Linguarum totius orbis vocabularia comparative* (Pallas 1786–89).

It was not always so. A remarkable passage from an unpublished manuscript in the British Library brings out clearly how very characteristic the explosion of grammar- and dictionary-making was of modern-era Europe, and how different it was from Europe’s supposed ancestors, Rome and Greece. The passage is from a text of another East India Company servant and Orientalist, John Leyden, of whom I will have occasion to speak later. In a proposal to the governor-general seeking sanction for an ambitious plan of making a whole set of grammars and dictionaries for the languages of India and Southeast Asia, Leyden lays out the value of languages of the region for administration, diplomacy and trade, and goes on to consider the importance of studying all languages, not just the useful ones, “in a literary point of view,” which is to say, a purely scholarly one:

Almost all the languages and dialects of India have been more or less cultivated by writing. Almost all of them contain a variety of compositions and untill these compositions be examined by Europeans the history of the tribes and nations to which they belong can never be elucidated in a satisfactory manner. India is as it were the literary property of the English nation. The facility of research is in their hands and in their[s] alone, as it is only by the individuals of this nation that the literature and languages of India whether of primary or secondary importance, can be properly investigated. The peculiar advantages therefore which the English enjoy, render them responsible to the general commonwealth of letters and pledged to all posterity as long as civilized men shall exist in the earth. The fame of some great conquering nations and in particular of the Romans is deeply sullied by their inattention to the literature of the nations they conquered. The sanguine flood of Roman glory not only swept away ancient nations but obliterated the accumulated records of the primeval times. The Roman warriors trod under foot the arts and sciences of Etruria and left nothing but the fragments of their vases to demonstrate the existence of former civilization. The ancient literature of Spain the monuments of the Carthaginians and Phoenicians & the venerable institutions of the Druids of Gaul & Britain they extinguished, and [of] the literature of ancient Egypt, they left us nothing but the hieroglyphics as a riddle to perplex future ages, a cypher of which they destroyed the key. Roman literature as it is preserved, is a ruin of ruins, a military trophy, raised from the fragments of ancient monuments. (Leyden, "Plan for investigating the languages, literatures, antiquities and history of the Dekhan"; BL Add. Mss. 26,567, ff. 112v-113)

This passage is remarkable for its frank imperialism, but, given the tendency of the British to imagine themselves in togas, it is remarkable as well in condemning the Greeks and Romans for failing to preserve knowledge of the Etruscan and ancient Egyptian languages. Leyden is quite right to point out the Greeks' and Romans' utter indifference to the languages of others, in the sense of failing to make grammars and dictionaries of them and conserve knowledge of them for the future.

The explosion in the grammar factory, then, is a specifically European phenomenon, and one in which its biblical heritage is a crucial ingredient that is absent from ancient Rome and Greece. The twinning of languages and nations begins in Genesis, with its seventy-two patriarchs and seventy-two languages, as St. Augustine has it. The explosive growth of grammars and dictionaries in the eighteenth century is perhaps a late outcome of the missionary tendency of Christianity; for both the translation of the Bible into the vernaculars stimulated by the Protestant Reformation, and the making of grammars and dictionaries for missionary purposes stimulated by the Counter-Reformation, with the great Jesuit projects in the vanguard, were important precursors of the eighteenth-

century explosion. No doubt there were other causes. Benedict Anderson, noting the dramatic increase in the writing of grammars and dictionaries for European languages beginning in the eighteenth century, makes a powerful case for the connection of this growth industry with nationalism and the means of its propagation: print capitalism (Anderson 1991:70–82). Bernard Cohn, describes the grammar- and dictionary-writing fever in early British-Indian Calcutta and makes a powerful case for the connection of this growth industry with British colonial rule and the urgent need to master the languages of India in order to master India itself (Cohn 1985). But nationalism and colonialism do not encompass the field constituted by *all* languages, however distant and obscure. The explosion in the grammar factory is fired by the forces of national self-love and colonial utility, no doubt, but it exceeds them as well. Nor is it a natural product of these forces. It has a history, and it began rather recently. It is not a constant of history.