The Horse, the Wheel, and Language

David W. Anthony

Published by Princeton University Press

→ For additional information about this book

https://muse.jhu.edu/book/36661

CHAPTER THREE
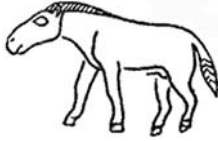
Language and Time 1
The Last Speakers of Proto-Indo-European

Time changes everything. Reading to my young children, I found that in mid-sentence I began to edit and replace words that suddenly looked archaic to me, in stories I had loved when I was young. The language of Robert Louis Stevenson and Jules Verne now seems surprisingly stiff and distant, and as for Shakespeare's English—we all need the glossary. What is true for modern languages was true for prehistoric languages. Over time, they changed. So what do we mean by Proto-Indo-European? If it changed over time, is it not a moving target? However we define it, for how long was Proto-Indo-European spoken? Most important, when was it spoken? How do we assign a date to a language that left no inscriptions, that died without ever being written down? It helps to divide any problem into parts, and this one can easily be divided into two: the birth date and the death date.

This chapter concentrates on the death date, the date *after which* Proto-Indo-European must have ceased to exist. But it helps to begin by considering how long a period probably preceded that. Given that the time between the birth and death dates of Proto-Indo-European could not have been infinite, precisely how long a time was it? Do languages, which are living, changing things, have life expectancies?

THE SIZE OF THE CHRONOLOGICAL WINDOW: HOW LONG DO LANGUAGES LAST?

If we were magically able to converse with an English speaker living a thousand years ago, as proposed in the last chapter, we would not understand each other. Very few natural languages, those that are learned and spoken at home, remain sufficiently unchanged after a thousand years to be considered the "same language." How can the rate of change be measured?

Languages normally have dialects—regional accents—and, within any region, they have innovating social sectors (entertainers, soldiers, traders) and conservative sectors (the very rich, the very poor). Depending on who you are, your language might be changing very rapidly or very slowly. Unstable conditions—invasions, famines, the fall of old prestige groups and the rise of new ones—increase the rate of change. Some parts of language change earlier and faster, whereas other parts are resistant. That last observation led the linguist Morris Swadesh to develop a standard word list chosen from the most resistant vocabulary, a group of words that tend to be retained, not replaced, in most languages around the world, even after invasions and conquests. Over the long term, he hoped, the average rate of replacement in this resistant vocabulary might yield a reliable standardized measurement of the speed of language change, what Swadesh called *glottochronology*.[1]

Between 1950 and 1952 Swadesh published a hundred-word and a two-hundred-word *basic core vocabulary*, a standardized list of resistant terms. All languages, he suggested, tend to retain their own words for certain kinds of meanings, including body parts (blood, foot); lower numerals (one, two, three); some kinship terms (mother, father); basic needs (eat, sleep); basic natural features (sun, moon, rain, river); some flora and fauna (tree, domesticated animals); some pronouns (this, that, he, she); and conjunctions (and, or, if). The content of the list can be and has been modified to suit vocabularies in different languages—in fact, the preferred two-hundred-meaning list in English contains 215 words. The English core vocabulary has proven extremely resistant to change. Although English has borrowed more than 50% of its *general* vocabulary from the Romance languages, mainly from French (reflecting the conquest of Anglo-Saxon England by the French-speaking Normans) and Latin (from centuries of technical and professional vocabulary training in courts, churches, and schools), only 4% of the English *core* vocabulary is borrowed from Romance. In its core vocabulary English remains a Germanic language, true to its origins among the Anglo-Saxons who migrated from northern Europe to Britain after the fall of the Roman Empire.

Comparing core vocabularies between old and new phases in languages with long historical records (Old English/Modern English, Middle Egyptian/Coptic, Ancient Chinese/Modern Mandarin, Late Latin/Modern French, and nine other pairs), Swadesh calculated an average replacement rate of 14% per thousand years for the hundred-word list, and 19% per thousand years for the two-hundred-word list. He suggested

that 19% was an acceptable average for all languages (usually rounded to 20%). To illustrate what that number means, Italian and French have distinct, unrelated words for 23% of the terms in the two-hundred-word list, and Spanish and Portuguese show a difference of 15%. As a general rule, if more than 10% of the core vocabulary is different between two dialects, they are either mutually unintelligible or approaching that state, that is, they are distinct languages or emerging languages. On average, then, with a replacement rate of 14–19% per thousand years in the core vocabulary, we should expect that most languages—including this one—would be incomprehensible to our own descendants a thousand years from now.

Swadesh hoped to use the replacement rate in the core vocabulary as a standardized clock to establish the date of splits and branches in unwritten languages. His own research involved the splits between American Indian language families in prehistoric North America, which were undatable by any other means. But the reliability of his standard replacement rate wilted under criticism. Extreme cases like Icelandic (very slow change, with a replacement rate of only 3–4% per thousand years) and English (very rapid, with a 26% replacement rate per thousand years) challenged the utility of the "average" rate.[2] The mathematics was affected if a language had multiple words for one meaning on the list. The dates given by glottochronology for many language splits contradicted known historical dates, generally by giving a date much later than it should have been. This direction in the errors suggested that real language change often was slower than Swadesh's model suggested—less than 19% per thousand years. A devastating critique of Swadesh's mathematics by Chretien, in 1962, seemed to drive a stake through the heart of glottochronology.

But in 1972 Chretien's critique was itself shown to be incorrect, and, since the 1980s, Sankoff and Embleton have introduced equations that include as critical values borrowing rates, the number of geographic borders with other languages, and a similarity index between the compared languages (because similar languages borrow in the core more easily then dissimilar languages). Multiple synonyms can each be given a fractional score. Studies incorporating these improved methods succeeded better in producing dates for splits between known languages that matched historical facts. More important, comparisons between most Indo-European languages still yielded replacement rates in the core vocabulary of about 10–20% per thousand years. Comparing the core vocabularies in ninety-five Indo-European languages, Kruskal and Black found that the most frequent date for the first splitting of Proto–Indo–European was about

3000 BCE. Although this estimate cannot be relied on absolutely, it is probably "in the ballpark" and should not be ignored.[3]

One simple point can be extracted from these debates: if the Proto-Indo-European core vocabulary changed at a rate ≥10% per millennium, or at the lower end of the expected range, Proto-Indo-European did not exist as a single language with a single grammar and vocabulary for as long as a thousand years. Proto-Indo-European grammar and vocabulary should have changed quite substantially over a thousand years. Yet the grammar of Proto-Indo-European, as reconstructed by linguists, is remarkably homogeneous both in morphology and phonology. Proto-Indo-European nouns and pronouns shared a set of cases, genders, and declensions that intersect with dozens of cognate phonological endings. Verbs had a shared system of tenses and aspects, again tagged by a shared set of phonological vowel changes (*run-ran*) and endings. This shared system of grammatical structures and phonological ways of labeling them looks like a single language. It suggests that reconstructed Proto-Indo-European probably refers to less than a thousand years of language change. It took less than a thousand years for late Vulgar Latin to evolve into seven Romance languages, and Proto-Indo-European does not contain nearly enough internal grammatical diversity to represent seven distinct grammars.

But considering that Proto-Indo-European is a fragmentary reconstruction, not an actual language, we should allow it more time to account for the gaps in our knowledge (more on this in chapter 5). Let us assign a nominal lifetime of two thousand years to the phase of language history represented by reconstructed Proto-Indo-European. In the history of English two thousand years would take us all the way back to the origins of the sound shifts that defined Proto-Germanic, and would include all the variation in all the Germanic languages ever spoken, from Hlewagasti of Holt to Puff Daddy of hip-hop fame. Proto-Indo-European does not seem to contain that much variation, so two thousand years probably is too long. But for archaeological purposes it is quite helpful to be able to say that the time period we are trying to identify is no longer than two thousand years.

What is the end date for that two-thousand-year window of time?

The Terminal Date for Proto-Indo-European: The Mother Becomes Her Daughters

The terminal date for reconstructed Proto-Indo-European—the date after which it becomes an anachronism—should be close to the date when its oldest daughters were born. Proto-Indo-European was reconstructed on

the basis of systematic comparisons between all the Indo-European daughter languages. The mother tongue cannot be placed later than the daughters. Of course, it would have survived after the detachment and isolation of the oldest daughter, but as time passed, if that daughter dialect remained isolated from the Proto-Indo-European speech community, each would have developed its own peculiar innovations. The image of the mother that is retained through each of the daughters is the form the mother had *before* the detachment of that daughter branch. Each daughter, therefore, preserves a somewhat different image of the mother.

Linguists have exploited this fact and other aspects of internal variation to identify chronological phases within Proto-Indo-European. The number of phases defined by different linguists varies from three (early, middle, late) to six.[4] But if we define Proto-Indo-European as the language that was ancestral to *all* the Indo-European daughters, then it is the *oldest* reconstructable form, the *earliest* phase of Proto-Indo-European, that we are talking about. The later daughters did not evolve directly from this early kind of Proto-Indo-European but from some intermediate, evolved set of late Indo-European languages that preserved aspects of the mother tongue and passed them along.

So when did the oldest daughter separate? The answer to that question depends very much on the accidental survival of written inscriptions. And the oldest daughter preserved in written inscriptions is so peculiar that it is probably safer to rely on the image of the mother preserved within the second set of daughters. What's wrong with the oldest daughter?

The Oldest and Strangest Daughter (or Cousin?): Anatolian

The oldest written Indo-European languages belonged to the Anatolian branch. The Anatolian branch had three early stems: Hittite, Luwian, and Palaic.[5] All three languages are extinct but once were spoken over large parts of ancient Anatolia, modern Turkey (figure 3.1). Hittite is by far the best known of the three, as it was the palace and administrative language of the Hittite Empire.

Inscriptions place Hittite speakers in Anatolia as early as 1900 BCE, but the empire was created only about 1650–1600 BCE, when Hittite warlords conquered and united several independent native Hattic kingdoms in central Anatolia around modern Kayseri. The name *Hittite* was given to them by Egyptian and Syrian scribes who failed to distinguish the Hittite kings from the Hattic kings they had conquered. The Hittites called themselves *Neshites* after the Anatolian city, Kanesh, where they

Figure 3.1 The ancient languages of Anatolia at about 1500 BCE.

rose to power. But Kanesh had earlier been a Hattic city; its name was Hattic. Hattic-speakers also named the city that became the capital of the Hittite Empire, Hattušas. Hattic was a non–Indo-European language, probably linked distantly to the Caucasian languages. The Hittites borrowed Hattic words for throne, lord, king, queen, queen mother, heir apparent, priest, and a long list of palace officials and cult leaders—probably in a historical setting where the Hattic languages were the languages of royalty. Palaic, the second Anatolian language, also borrowed vocabulary from Hattic. Palaic was spoken in a city called Pala probably located in north-central Anatolia north of Ankara. Given the geography of Hattic place-names and Hattic→ Palaic/Hittite loans, Hattic seems to have been spoken across all of central Anatolia before Hittite or Palaic was spoken there. The early speakers of Hittite and Palaic were intruders in a non–Indo-European central Anatolian landscape dominated by Hattic speakers who had already founded cities, acquired literate bureaucracies, and established kingdoms and palace cults.[6]

After Hittite speakers usurped the Hattic kingdom they enjoyed a period of prosperity enriched by Assyrian trade, and then endured defeats that later were dimly but bitterly recalled. They remained confined to the center of the Anatolian plateau until about 1650 BCE, when Hittite armies became mighty enough to challenge the great powers of the Near East and the imperial era began. The Hittites looted Babylon, took other cities from the Assyrians, and fought the Egyptian pharaoh Ramses II to a standstill at the greatest chariot battle of ancient times, at Kadesh, on

the banks of the Orontes River in Syria, in 1286 BCE. A Hittite monarch married an Egyptian princess. The Hittite kings also knew and negotiated with the princes who ruled Troy, probably the place referred to in the Hittite archives as *steep Wilusa* (*Ilios*).[7] The Hittite capital city, Hattušas, was burned in a general calamity that brought down the Hittite kings, their army, and their cities about 1180 BCE. The Hittite language then quickly disappeared; apparently only the ruling élite ever spoke it.

The third early Anatolian language, Luwian, was spoken by more people over a larger area, and it continued to be spoken after the end of the empire. During the later Hittite empire Luwian was the dominant spoken language even in the Hittite royal court. Luwian did not borrow from Hattic and so might have been spoken originally in western Anatolia, outside the Hattic core region—perhaps even in Troy, where a Luwian inscription was found on a seal in Troy level VI—the Troy of the Trojan War. On the other hand, Luwian did borrow from other, unknown non–Indo-European language(s). Hittite and Luwian texts are abundant from the empire period, 1650–1180 BCE. These are the earliest complete texts in any Indo-European language. But individual Hittite and Luwian words survive from an earlier era, before the empire began.[8]

The oldest Hittite and Luwian names and words appeared in the business records of Assyrian merchants who lived in a commercial district, or *karum*, outside the walls of Kanesh, the city celebrated by the later Hittites as the place where they first became kings. Archaeological excavations here, on the banks of the Halys River in central Anatolia, have shown that the Assyrian *karum*, a foreigners' enclave that covered more than eighty acres outside the Kanesh city walls, operated from about 1920 to 1850 BCE (level II), was burned, rebuilt, and operated again (level Ib) until about 1750 BCE, when it was burned again. After that the Assyrians abandoned the karum system in Anatolia, so the Kanesh karum is a closed archaeological deposit dated between 1920 and 1750 BCE. The Kanesh karum was the central office for a network of literate Assyrian merchants who oversaw trade between the Assyrian state and the warring kingdoms of Late Bronze Age Anatolia. The Assyrian decision to make Kanesh their distribution center greatly increased the power of its Hittite and Luwian occupants.

Most of the local names recorded by the merchants in the Kanesh karum accounts were Hittite or Luwian, beginning with the earliest records of about 1900 BCE. Many still were Hattic. But Hittite speakers seem to have controlled business with the Assyrian karum. The Assyrian merchants were so accustomed to doing business with Hittite speakers that they adopted Hittite words for *contract* and *lodging* even in their

private correspondence. Palaic, the third language of the Anatolian branch, is not known from the Kanesh records. Palaic died out as a spoken language probably before 1500 BCE. It presumably was spoken in Anatolia during the karum period but not at Kanesh.

Hittite, Luwian, and Palaic had evolved already by 1900 BCE. This is a critical piece of information in any attempt to date Proto-Indo-European. All three were descended from the same root language, Proto-Anatolian. The linguist Craig Melchert described Luwian and Hittite of the empire period, ca. 1400 BCE, as sisters about as different as twentieth-century Welsh and Irish.[9] Welsh and Irish probably share a common origin of about two thousand years ago. If Luwian and Hittite separated from Proto-Anatolian two thousand years before 1400 BCE, then Proto-Anatolian should be placed at about 3400 BCE. What about *its* ancestor? When did the root of the Anatolian branch separate from the rest of Proto-Indo-European?

Dating Proto-Anatolian: The Definition of Proto- and Pre-Languages

Linguists do not use the term *proto-* in a consistent way, so I should be clear about what I mean by Proto-Anatolian. Proto-Anatolian is the language that was *immediately ancestral* to the three known daughter languages in the Anatolian branch. Proto-Anatolian can be described fairly accurately on the basis of the shared traits of Hittite, Luwian, and Palaic. But Proto-Anatolian occupies just the *later portion* of an undocumented period of linguistic change that must have occurred between it and Proto-Indo-European. The hypothetical language stage in between can be called *Pre-Anatolian*. Proto-Anatolian is a fairly concrete linguistic entity closely related to its known daughters. But Pre-Anatolian represents an *evolutionary period*. Pre-Anatolian is a phase defined by Proto-Anatolian at one end and Proto-Indo-European at the other. How can we determine when Pre-Anatolian separated from Proto-Indo-European?

The ultimate age of the Anatolian branch is based partly on objective external evidence (dated documents at Kanesh), partly on presumed rates of language change over time, and partly on internal evidence within the Anatolian languages. The Anatolian languages are quite different phonologically and grammatically from all the other known Indo-European daughter languages. They are so peculiar that many specialists think they do not really belong with the other daughters.

Many of the peculiar features of Anatolian look like archaisms, characteristics thought to have existed in an extremely early stage of Proto-

Indo-European. For example, Hittite had a kind of consonant that has become famous in Indo-European linguistics (yes, consonants can be famous): h_2, a guttural sound or *laryngeal*. In 1879 a Swiss linguist, Ferdinand de Saussure, realized that several seemingly random differences in vowel pronunciation between the Indo-European languages could be brought under one explanatory rule if he assumed that the pronunciation of these vowels had been affected by a "lost" consonant that no longer existed in any Indo-European language. He proposed that such a lost sound had existed in Proto-Indo-European. It was the first time a linguist had been so bold as to reconstruct a feature for Proto-Indo-European that no longer existed in any Indo-European language. The discovery and decipherment of Hittite forty years later proved Saussure right. In a stunning confirmation of the predictive power of comparative linguistics, the Hittite laryngeal h_2 (and traces of a slightly different laryngeal, h_3) appeared in Hittite inscriptions in just those positions Saussure had predicted for his "lost" consonant. Most Indo-Europeanists now accept that archaic Proto-Indo-European contained laryngeal sounds (probably three different ones, usually transcribed as $*h_1$, $*h_2$, $*h_3$) that were preserved clearly only in the Anatolian branch.[10] The best explanation for why Anatolian has laryngeals is that Pre-Anatolian speakers became separated from the Proto-Indo-European language community at a very early date, when a laryngeal-rich phonology was still characteristic of archaic Proto-Indo-European. But then what does *archaic* mean? What, exactly, did Pre-Anatolian separate from?

The Indo–Hittite Hypothesis

The Anatolian branch either lost or never possessed other features that were present in all other Indo-European branches. In verbs, for example, the Anatolian languages had only two tenses, a present and a past, whereas the other ancient Indo-European languages had as many as six tenses. In nouns, Anatolian had just animate and neuter; it had no feminine case. The other ancient Indo-European languages had feminine, masculine, and neuter cases. The Anatolian languages also lacked the dual, a form that was used in other early Indo-European languages for objects that were doubled like eyes or ears. (Example: Sanskrit dēvas 'one god', but dēvau 'double gods'.) Alexander Lehrman identified ten such traits that probably were innovations in Proto-Indo-European after Pre-Anatolian split away.[11]

For some Indo-Europeanists these traits suggest that the Anatolian branch did not develop from Proto-Indo-European at all but rather evolved

from an older Pre-Proto-Indo-European ancestor. This ancestral language was called Indo-Hittite by William Sturtevant. According to the Indo-Hittite hypothesis, Anatolian is an Indo-European language only in the broadest sense, as it did not develop from Proto-Indo-European. But it did preserve, uniquely, features of an earlier language community from which they both evolved. I cannot solve the debate over the categorization of Anatolian here, although it is obviously true that Proto-Indo-European must have evolved from an earlier language community, and we can use *Indo-Hittite* to refer to that hypothetical earlier stage. The Proto-Indo-European language community was a chain of dialects with both geographic and chronological differences. The Anatolian branch seems to have separated from an archaic chronological stage in the evolution of Proto-Indo-European, and it probably separated from a different geographic dialect as well, but I will call it archaic Proto-Indo-European rather than Indo-Hittite.[12]

A substantial period of time is needed for the Pre-Anatolian phase. Craig Melchert and Alexander Lehrman agreed that a separation date of about 4000 BCE between Pre-Anatolian and the archaic Proto-Indo-European language community seems reasonable. The millennium or so around 4000 BCE, say 4500 to 3500 BCE, constitutes the *latest* window within which Pre-Anatolian is likely to have separated.

Unfortunately the oldest daughter of Proto-Indo-European looks so peculiar that we cannot be certain she is a daughter rather than a cousin. Pre-Anatolian could have emerged from Indo-Hittite, not from Proto-Indo-European. So we cannot confidently assign a terminal date to Proto-Indo-European based on the birth of Anatolian.

The Next Oldest Inscriptions: Greek and Old Indic

Luckily we have well-dated inscriptions in two other Indo-European languages from the same era as the Hittite empire. The first was Greek, the language of the palace-centered Bronze Age warrior kings who ruled at Mycenae, Pylos, and other strongholds in Greece beginning about 1650 BCE. The Mycenaean civilization appeared rather suddenly with the construction of the spectacular royal Shaft Graves at Mycenae, dated about 1650 BCE, about the same time as the rise of the Hittite empire in Anatolia. The Shaft Graves, with their golden death masks, swords, spears, and images of men in chariots, signified the elevation of a new Greek-speaking dynasty of unprecedented wealth whose economic power depended on long-distance sea trade. The Mycenaean kingdoms were

destroyed during the same period of unrest and pillage that brought down the Hittite Empire about 1150 BCE. Mycenaean Greek, the language of palace administration as recorded in the Linear B tablets, was clearly Greek, not Proto-Greek, by 1450 BCE, the date of the oldest preserved inscriptions. The people who spoke it were the models for Nestor and Agamemnon, whose deeds, dimly remembered and elevated to epic, were celebrated centuries later by Homer in the *Iliad* and the *Odyssey*. We do not know when Greek speakers appeared in Greece, but it happened no later than 1650 BCE. As with Anatolian, there are numerous indications that Mycenaean Greek was an intrusive language in a land where non-Greek languages had been spoken before the Mycenaean age.[13] The Mycenaeans almost certainly were unaware that another Indo-European language was being used in palaces not far away.

Old Indic, the language of the *Rig Veda*, was recorded in inscriptions not long after 1500 BCE but in a puzzling place. Most Vedic specialists agree that the 1,028 hymns of the *Rig Veda* were compiled into what became the sacred form in the Punjab, in northwestern India and Pakistan, probably between about 1500 and 1300 BCE. But the deities, moral concepts, and Old Indic language of the *Rig Veda* first appeared in written documents not in India but in *northern Syria*.[14]

The Mitanni dynasty ruled over what is today northern Syria between 1500 and 1350 BCE. The Mitanni kings regularly spoke a *non*–Indo-European language, Hurrian, then the dominant local language in much of northern Syria and eastern Turkey. Like Hattic, Hurrian was a native language of the Anatolian uplands, related to the Caucasian languages. But all the Mitanni kings, first to last, took Old Indic throne names, even if they had Hurrian names before being crowned. Tus'ratta I was Old Indic *Tvesa-ratha* 'having an attacking chariot', Artatama I was *Rta-dhaaman* 'having the abode of r'ta', Artas's'umara was *Rta-smara* 'remembering r'ta', and S'attuara I was *Satvar* 'warrior'.[15] The name of the Mitanni capital city, Waššukanni, was Old Indic *vasu-khani*, literally "wealth-mine." The Mitanni were famous as charioteers, and, in the oldest surviving horse-training manual in the world, a Mitanni horse trainer named Kikkuli (a Hurrian name) used many Old Indic terms for technical details, including horse colors and numbers of laps. The Mitanni military aristocracy was composed of chariot warriors called *maryanna*, probably from an Indic term *márya* meaning "young man," employed in the *Rig Veda* to refer to the heavenly war-band assembled around Indra. Several royal Mitanni names contained the Old Indic term *r'ta*, which meant "cosmic order and truth," the central moral concept of the *Rig Veda*. The Mitanni king

Kurtiwaza explicitly named four Old Indic gods (Indra, Varuna, Mithra, and the Nāsatyas), among many native Hurrian deities, to witness his treaty with the Hittite monarch around 1380 BCE. And these were not just any Old Indic gods. Three of them—Indra, Varuna, and the Nāsatyas or Divine Twins—were the three most important deities in the *Rig Veda*. So the Mitanni texts prove not only that the Old Indic language existed by 1500 BCE but also that the central religious pantheon and moral beliefs enshrined in the *Rig Veda* existed equally early.

Why did Hurrian-speaking kings in Syria use Old Indic names, words, and religious terms in these ways? A good guess is that the Mitanni kingdom was founded by Old Indic-speaking mercenaries, perhaps charioteers, who regularly recited the kinds of hymns and prayers that were collected at about the same time far to the east by the compilers of the *Rig Veda*. Hired by a Hurrian king about 1500 BCE, they usurped his throne and founded a dynasty, a very common pattern in Near Eastern and Iranian dynastic histories. The dynasty quickly became Hurrian in almost every sense but clung to a tradition of using Old Indic royal names, some Vedic deity names, and Old Indic technical terms related to chariotry long after its founders faded into history. This is, of course, a guess, but something like it seems almost necessary to explain the distribution and usage of Old Indic by the Mitanni.

The Mitanni inscriptions establish that Old Indic was being spoken before 1500 BCE in the Near East. By 1500 BCE Proto-Indo-European had differentiated into at least Old Indic, Mycenaean Greek, and the three known daughters of Proto-Anatolian. What does this suggest about the terminal date for Proto-Indo-European?

Counting the Relatives: How Many in 1500 BCE?

To answer this question we first have to understand where Greek and Old Indic are placed among the known branches of the Indo-European family. Mycenaean Greek is the oldest recorded language in the Greek branch. It is an isolated language; it has no recorded close relatives or sister languages. It probably had unrecorded sisters, but none survived in written records. The appearance of the Shaft-Grave princes about 1650 BCE represents the latest possible arrival of Greek speakers in Greece. The Shaft-Grave princes probably already spoke an early form of Greek, not Proto-Greek, since their descendants' oldest preserved inscriptions at about 1450 BCE were in Greek. Proto-Greek might be dated *at the latest* between about 2000 and 1650 BCE. Pre-Greek, the phase that preceded

Proto-Greek, probably originated as a dialect of late Proto-Indo-European *at least* five hundred to seven hundred years before the appearance of Mycenaean Greek, and very probably earlier—*minimally* about 2400–2200 BCE. The terminal date for Proto-Indo-European can be set at about 2400–2200 BCE—it could not have been later than this—from the perspective of the Greek branch. What about Old Indic?

Unlike Mycenaean Greek, Old Indic *does* have a known sister language, Avestan Iranian, which we must take into account. Avestan is the oldest of the Iranian languages that would later be spoken by Persian emperors and Scythian nomads alike, and today are spoken in Iran and Tajikistan. Avestan Iranian was the language of the *Avesta*, the holiest text of Zorastrianism. The oldest parts of the *Avesta*, the *Gathas*, probably were composed by Zoroaster (the Greek form of the name) or by Zarathustra (the original Iranian form) himself. Zarathustra was a religious reformer who lived in eastern Iran, judging from the places he named, probably between 1200 and 1000 BCE.[16] His theology was partly a reaction against the glorification of war and blood sacrifice by the poets of the *Rig Veda*. One of the oldest Gathas was "the lament of the cow," a protest against cattle stealing from the cow's point of view. But the *Avesta* and the *Rig Veda* were closely related in both language and thought. They used the same deity names (although Old Indic gods were demonized in the *Avesta*), employed the same poetic conventions, and shared specific rituals. For example, they used a cognate term for the ritual of spreading straw for the seat of the attending god before a sacrifice (Vedic *barhis,* Avestan *baresman*); and both traditions termed a pious man "one who spread the straw." In many small details they revealed their kinship in a shared Indo-Iranian past. The two languages, Avestan Iranian and Old Indic, developed from a shared parent language, Indo-Iranian, which is not documented.

The Mitanni inscriptions establish that Old Indic had appeared as a distinct language by 1500 BCE. Common Indo-Iranian must be earlier. It probably dates back *at least* to 1700 BCE. Proto-Indo-Iranian—a dialect that had some of the innovations of Indo-Iranian but not yet all of them—has to be placed earlier still, at or before 2000 BCE. Pre-Indo-Iranian was an eastern dialect of Proto-Indo-European, and must then have existed at the *latest* around 2500–2300 BCE. As with Greek, the period from 2500 to 2300 BCE, give or take a few centuries, is the *minimal* age for the separation of Pre-Indo-Iranian from Proto-Indo-European.

So the terminal date for Proto-Indo-European—the date after which our reconstructed form of the language becomes an anachronism—can be set around 2500 BCE, more or less, from the perspective of Greek and

Old Indic. It might be extended a century or two later, but, as far as *these two languages* are concerned, a terminal date *much* later than 2500 BCE—say, as late as 2000 BCE—is impossible. And, of course, Anatolian must have separated long before 2500 BCE. By about 2500 BCE Proto-Indo-European had changed and fragmented into a variety of late dialects and daughter languages—including at least the Anatolian group, Pre-Greek and Pre-Indo-Iranian. Can other daughters be dated to the same period? How many other daughters existed by 2500 BCE?

More Help from the Other Daughters: Who's the Oldest of Them All?

In fact, some other daughters not only *can* be placed this early—they *must* be. Again, to understand why, we have to understand where Greek and Old Indic stand within the known branches of the Indo-European language family. Neither Greek nor Indo-Iranian can be placed among the very oldest Indo-European daughter branches. They are the oldest daughters to survive in inscriptions (along with Anatolian), but that is an accident of history (table 3.1). From the perspective of historical linguistics, Old Indic and Greek must be classified as *late* Indo-European daughters. Why?

Linguists distinguish older daughter branches from younger ones on the basis of shared innovations and archaisms. Older branches seem to have separated earlier because they lack innovations characteristic of the later branches, and they retain archaic features. Anatolian is a good example; it retains some phonetic traits that definitely are archaic (laryngeals) and lacks other features that probably represent innovations. Indo-Iranian, on the other hand, exhibits three innovations that identify it as a later branch.

Indo-Iranian shared one innovation with a group of languages that linguists labeled the *satəm* group: Indo-Iranian, Slavic, Baltic, Albanian, Armenian, and perhaps Phrygian. Among the *satəm* languages, Proto-Indo-European *k*- before a front vowel (like *k'mtom* 'hundred') was regularly shifted to *š*- or *s*- (like Avestan Iranian *satəm*). This same group of languages exhibited a second shared innovation: Proto-Indo-European *k^w*- (called a labiovelar, pronounced like the first sound in *queen*) changed to *k*-. The third innovation was shared between just a subgroup within the *satəm* languages: Indo-Iranian, Baltic, and Slavic. It is called the *ruki*-rule: the original sound [*-s] in Proto-Indo-European was shifted to [*-sh] after the consonants *r*, *u*, *k*, and *i*. Language branches that do not share these innovations are assumed to have split away and lost regular contact with the *satəm* and *ruki* groups before they occurred.

Table 3.1

The First Appearance in Written Records of the Twelve Branches of Indo–European

Language Branch	Oldest Documents or Inscriptions	Diversity at That Date	Latest Date for Proto–Language for the Branch	Grouped with
Anatolian	1920 BCE	Three closely related languages	2800–2300 BCE	No close sisters
Indo–Iranian	1450 BCE	Two very closely related languages	2000–1500 BCE	Greek, Balto–Slavic
Greek	1450 BCE	One dialect recorded, but others probably existed	2000–1500 BCE	Indo–Iranian, Armenian
Phrygian	750 BCE	Poorly documented	1200–800 BCE	Greek? Italo–Celtic?
Italic	600–400 BCE	Four languages, grouped into two quite distinct sub–branches	1600–1100 BCE	Celtic
Celtic	600–300 BCE	Three broad groups with different SVO syntax	1350–850 BCE	Italic
Germanic	0–200 CE	Low diversity; probably the innovations that defined Germanic were recent and still spreading through the Pre–Germanic speech community	500–0 BCE	Baltic/Slavic

Table 3.1 (*continued*)

Language Branch	Oldest Documents or Inscriptions	Diversity at That Date	Latest Date for Proto–Language for the Branch	Grouped with
Armenian	400 CE	Only one dialect documented, but Armina was a Persian province ca. 500 BCE so other dialects probably existed 400 CE	500 BCE–0 CE?	Greek, Phrygian?
Tocharian	500 CE	Two (perhaps three) quite distinct languages	500 BCE–0 CE	No close sisters
Slavic	865 CE	Only one dialect documented (OCS), but the West, South, and East Slavic branches must have existed already	0–500 CE	Baltic
Baltic	1400 CE	Three languages	0–500 CE	Slavic
Albanian	1480 CE	Two dialects	0–500 CE	Dacian–Thracian? No close sisters

The Celtic and Italic branches do not display the *satəm* innovations or the *ruki* rule; both exhibit a number of archaic features and also share a few innovations. Celtic languages, today limited to the British Isles and nearby coastal France, were spoken over much of central and western Europe, from Austria to Spain, around 600–300 BCE, when the earliest records of Celtic appeared. Italic languages were spoken in the Italian peninsula at about 600–500 BCE, but today, of course, Latin has many daughters—the Romance languages. In most comparative studies of the Indo-European languages, Italic and Celtic would be placed among the earliest branches to separate from the main trunk. The people who spoke Pre-Celtic and Pre-Italic lost contact with the eastern and northern groups of Indo-European speakers before the *satəm* and *ruki* innovations occurred. We cannot yet discuss where the boundaries of these linguistic regions were, but we can say that Pre-Italic and Pre-Celtic departed to form a western regional–chronological block, whereas the ancestors of Indo-Iranian, Baltic, Slavic, and Armenian stayed behind and shared a set of later innovations. Tocharian, the easternmost Indo-European language, spoken in the Silk Road caravan cities of the Tarim Basin in northwestern China, also lacked the *satəm* and *ruki* innovations, so it seems to have departed equally early to form an eastern branch.

Greek shared a series of linguistic features uniquely with the Indo-Iranian languages, but it did not adopt the *satəm* innovation or the *ruki* rule.[17] Pre-Greek and Pre-Indo-Iranian must have developed in neighboring regions, but the speakers of Pre-Greek departed before the *satəm* or the *ruki* innovations appeared. The shared features included morphological innovations, conventions in heroic poetry, and vocabulary. In morphology, Greek and Indo-Iranian shared two important innovations: the augment, a prefix *e-* before past tenses (although, because it is not well attested in the earliest forms of Greek and Indo-Iranian, the augment *might* have developed independently in each branch much later); and a mediopassive verb form with a suffixed *–i*. In weapon vocabulary they shared common terms for *bow* (**taksos*), *arrow* (**eis-*), *bowstring* (**jya-*), and *club* (**uágros*), or *cudgel*, the weapon specifically associated with Indra and his Greek counterpart Herakles. In ritual they shared a unique term for a specific ritual, the *hecatomb*, or sacrifice of a hundred cows; and they referred to the gods with the same shared epithet, *those who give riches*. They retained shared cognate names for at least three deities: (1) *Erinys/Saraṇ, yū*, a horse-goddess in both traditions, born of a primeval creator-god and the mother of a winged horse in Greek or of the Divine Twins in

Indo-Iranian, who are often represented as horses; (2) *Kérberos/Śárvara*, the multiheaded dog that guarded the entrance to the Otherworld; and (3) *Pan/Pūṣán*, a pastoral god that guarded the flocks, symbolically associated in both traditions with the goat. In both traditions, goat entrails were the specific funeral offering made to the hell-hound *Kérberos/Śárvara* during a funeral ceremony. In poetry, ancient Greek, like Indo-Iranian, had two kinds of verse: one with a twelve-syllable line (the Sapphic/Alcaic line) and another with an eight-syllable line. No other Indo-European poetic tradition shared both these forms. They also shared a specific poetic formula, meaning "fame everlasting," applied to heroes, found in this exact form only in the *Rig Veda* and Homer. Both Greek and Indo-Iranian used a specific verb tense, the imperfect, in poetic narratives about past events.[18]

It is unlikely that such a large bundle of common innovations, vocabulary, and poetic forms arose independently in two branches. Therefore, Pre-Greek and Pre-Indo-Iranian almost certainly were neighboring late Indo-European dialects, spoken near enough to each other so that words related to warfare and ritual, names of gods and goddesses, and poetic forms were shared. Greek did not adopt the *ruki* rule or the *satəm* shift, so we can define two strata here: the older links Pre-Greek and Pre-Indo-Iranian, and the later separates Proto-Greek from Proto-Indo-Iranian.

The Birth Order of the Daughters and the Death of the Mother

The *ruki* rule, the *centum/satəm* split, and sixty-three possible variations on seventeen other morphological and phonological traits were analyzed mathematically to generate thousands of possible branching diagrams by Don Ringe, Wendy Tarnow, and colleagues at the University of Pennsylvania.[19] The cladistic method they used was borrowed from evolutionary biology but was adapted to compare linguistic innovations rather than genetic ones. A program selected the trees that emerged most often from among *all possible* evolutionary trees. The evolutionary trees identified by this method agreed well with branching diagrams proposed on more traditional grounds. The oldest branch to split away was, without any doubt, Pre-Anatolian (figure 3.2). Pre-Tocharian probably separated next, although it also showed some later traits. The next branching event separated Pre-Celtic and Pre-Italic from the still evolving core. Germanic has some archaic traits that suggest an initial separation at about the same time as Pre-Celtic and Pre-Italic, but then later it was strongly affected by
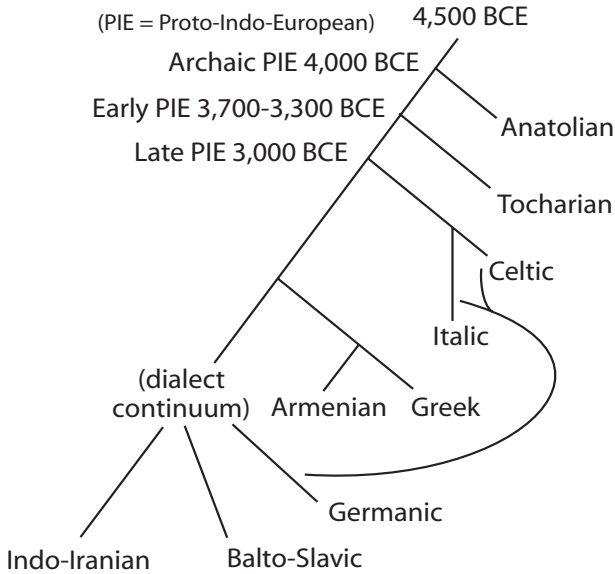
Figure 3.2 The best branching diagram according to the Ringe–Warnow–Taylor (2002) cladistic method, with the minimal separation dates suggested in this chapter. Germanic shows a mixture of archaic and derived traits that make its place uncertain; it could have branched off at about the same time as the root of Italic and Celtic, although here it is shown branching later because it also shared many traits with Pre-Baltic and Pre-Slavic.

borrowing from Celtic, Baltic, and Slavic, so the precise time it split away is uncertain. Pre-Greek separated after Italic and Celtic, followed by Indo-Iranian. The innovations of Indo-Iranian were shared (perhaps later) with several language groups in southeastern Europe (Pre-Armenian, Pre-Albanian, partly in Pre-Phrygian) and in the forests of northeastern Europe (Pre-Baltic and Pre-Slavic). Common Indo-Iranian, we must re-member, is dated *at the latest* to about 1700 BCE. The Ringe-Tarnow branching diagram puts the separations of Anatolian, Tocharian, Italic, Celtic, German, and Greek before this. Anatolian probably had split away before 3500 BCE, Italic and Celtic before 2500 BCE, Greek after 2500 BCE, and Proto-Indo-Iranian by 2000 BCE. Those are not meant to be exact dates, but they are in the right sequence, are linked to dated inscriptions in three places (Greek, Anatolian, and Old Indic), and make sense.

By 2500 BCE the language that has been reconstructed as Proto-Indo-European had evolved into something else or, more accurately, into a variety of things,—late dialects such as Pre-Greek and Pre-Indo-Iranian that continued to diverge in different ways in different places. The Indo-European languages that evolved after 2500 BCE did not develop from Proto-Indo-European but from a set of intermediate Indo-European languages that preserved and passed along aspects of the mother tongue. By 2500 BCE Proto-Indo-European was a dead language.