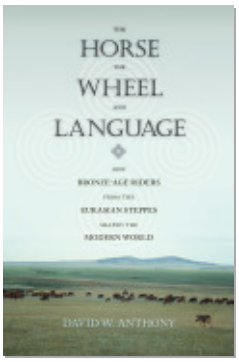
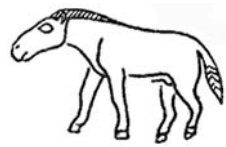
The Horse, the Wheel, and Language

David W. Anthony



Published by Princeton University Press

Chapter Two



How to Reconstruct a Dead Language

Proto-Indo-European has been dead as a spoken language for at least forty-five hundred years. The people who spoke it were nonliterate, so there are no inscriptions. Yet, in 1868, August Schleicher was able to tell a story in reconstructed Proto-Indo-European, called "The Sheep and the Horses," or *Avis akvasas ka*. A rewrite in 1939 by Herman Hirt incorporated new interpretations of Proto-Indo-European phonology, and the title became *Owis ek'woses-k^{w}e*. In 1979 Winfred Lehmann and Ladislav Zgusta suggested only minor new changes in their version, *Owis ekwosk^{w}e*. While linguists debate increasingly minute details of pronunciation in exercises like these, most people are amazed that anything can be said about a language that died without written records. Amazement, of course, is a close cousin of suspicion. Might the linguists be arguing over a fantasy? In the absence of corroborative evidence from documents, how can linguists be sure about the accuracy of reconstructed Proto-Indo-European?[1]

Many archaeologists, accustomed to digging up real things, have a low opinion of those who merely reconstruct hypothetical phonemes—what is called "linguistic prehistory." There are reasons for this skepticism. Both linguists and archaeologists have made communication across the disciplines almost impossible by speaking in dense jargons that are virtually impenetrable to anyone but themselves. Neither discipline is at all simple, and both are riddled with factions on many key questions of interpretation. Healthy disagreement can resemble confusion to an outsider, and most archaeologists, including this author, are outsiders in linguistics. Historical linguistics is not taught regularly in graduate archaeology programs, so most archaeologists know very little about the subject. Sometimes we make this quite clear to linguists. Nor is archaeology taught to graduate students in linguistics. Linguists' occasional remarks about archaeology can sound simplistic and naïve to archaeologists, making some

of us suspect that the entire field of historical linguistics may be riddled with simplistic and naïve assumptions.

The purpose of these first few chapters is to clear a path across the no-man's land that separates archaeology and historical linguistics. I do this with considerable uncertainty—I have no more formal training in linguistics than most archaeologists. I am fortunate that a partial way has already been charted by Jim Mallory, perhaps the only doubly qualified linguist-archaeologist in Indo-European studies. The questions surrounding Indo-European origins are, at their core, about linguistic evidence. The most basic linguistic problem is to understand how language changes with time.[2]

Language Change and Time

Imagine that you had a time machine. If you are like me, there would be many times and places that you would like to visit. In most of them, however, no one spoke English. If you could not afford the Six-Month-Immersion Trip to, say, ancient Egypt, you would have to limit yourself to a time and place where you could speak the language. Consider, perhaps, a trip to England. How far back in time could you go and still be understood? Say we go to London in the year 1400 CE.

As you emerge from the time machine, a good first line to speak, something reassuring and recognizable, might be the opening line of the Lord's Prayer. The first line in a conservative, old-fashioned version of Modern Standard English would be, "*Our Father, who is in heaven, blessed be your name.*" In the English of 1400, as spoken by Chaucer, you would say, "*Oure fadir that art in heuenes, halwid be thy name.*" Now turn the dial back another four hundred years to 1000 CE, and in Old English, or Anglo-Saxon, you would say, "*Fæader ure thu the eart on heofonum, si thin nama gehalgod.*" A chat with Alfred the Great would be out of the question.

Most normal spoken languages over the course of a thousand years undergo enough change that speakers at either end of the millennium, attempting a conversation, would have difficulty understanding each other. Languages like Church Latin or Old Indic (the oldest form of Sanskrit), frozen in ritual, would be your only hope for effective communication with people who lived more than a thousand years ago. Icelandic is a frequently cited example of a spoken language that has changed little in a thousand years, but it is spoken on an island isolated in the North Atlantic by people whose attitude to their old sagas and poetry has been one approaching religious reverence. Most languages undergo significantly more

changes than Icelandic over far fewer than a thousand years for two reasons: first, no two people speak the same language exactly alike; and, second, most people meet a lot more people who speak differently than do the Icelanders. A language that borrows many words and phrases from another language changes more rapidly than one with a low borrowing rate. Icelandic has one of the lowest borrowing rates in the world.[3] If we are exposed to a number of different ways of speaking, our own way of speaking is likely to change more rapidly. Fortunately, however, although the speed of language change is quite variable, the structure and sequence of language change is not.

Language change is not random; it flows in the direction of accents and phrases admired and emulated by large numbers of people. Once a target accent is selected, the structure of the sound changes that moves the speaker away from his own speech to the target is governed by rules. The same rules apparently exist in all our minds, mouths, and ears. Linguists just noticed them first. If rules define how a given innovation in pronunciation affects the old speech system—if sound shifts are predictable—then we should be able to play them backward, in effect, to hear earlier language states. That is more or less how Proto-Indo-European was reconstructed.

Most surprising about sound change is its regularity, its conformation to rules no one knows consciously. In early Medieval French there probably was a time when *tsent'm* 'hundred' was heard as just a dialectical pronunciation of the Latin word *kentum* 'hundred'. The differences in sound between the two were *allophones*, or different sounds that did not create different meanings. But because of other changes in how Latin was spoken, [ts-] began to be heard as a different sound, a phoneme distinct from [k-] that could change the meaning of a word. At that point people had to decide whether *kentum* was pronounced with a [k-] or a [ts-]. When French speakers decided to use [ts-], they did so not just for the word *kentum* but in every word where Latin had the sound *k-* before a front vowel like *-e-*. And once this happened, *ts-* became confused with initial *s-*, and people had to decide again whether *tsentum* was pronounced with a [ts-] or [s-]. They chose [s-]. This sequence of shifts dropped below the level of consciousness and spread like a virus through all pre-French words with analogous sequences of sounds. Latin *cera* 'wax', pronounced [kera], became French *cire*, pronounced [seer]; and Latin *civitas* 'community', pronounced [kivitas], became French *cité*, pronounced [seetay]. Other sound changes happened, too, but they all followed the same unspoken and unconscious rules—the sound shifts were not idiosyncratic or confined to certain words; rather, they spread systematically to all similar sounds in the language. Peoples'

ears were very discriminating in identifying words that fit or did not fit the analogy. In words where the Latin *k-* was followed by a *back* vowel like *-o* it remained a *k–*, as in Latin *costa* > French *côte*.

Sound changes are rule-governed probably because all humans instinctively search for order in language. This must be a hard-wired part of all human brains. We do it without committee meetings, dictionaries, or even literacy, and we are not conscious of what we are doing (unless we are linguists). Human language is defined by its rules. Rules govern sentence construction (syntax), and the relationship between the sounds of words (phonology and morphology) and their meaning. Learning these rules changes our awareness from that of an infant to a functioning member of the human tribe. Because language is central to human evolution, culture, and social identity, each member of the tribe is biologically equipped to cooperate in converting novel changes into regular parts of the language system.[4]

Historical linguistics was created as a discipline in the nineteenth century, when scholars first exposed and analyzed the rules we follow when speaking and listening. I do not pretend to know these rules adequately, and if I did I would not try to explain them all. What I hope to do is indicate, in a general way, how some of them work so that we can use the "reconstructed vocabulary" of Proto-Indo-European with some awareness of its possibilities and limitations.

We begin with phonology. Any language can be separated into several interlocking systems, each with its own set of rules. The vocabulary, or *lexicon*, composes one system; *syntax*, or word order, and sentence construction compose another; *morphology*, or word form, including much of what is called "grammar" is the third; and *phonology*, or the rules about which sounds are acceptable and meaningful, is the fourth. Each system has its own peculiar tendencies, although a change in one (say, phonology) can bring about changes in another (say, morphology).[5] We will look most closely at phonology and the lexicon, as these are the most important in understanding how the Proto-Indo-European vocabulary has been reconstructed.

Phonology: How to Reconstruct a Dead Sound

Phonology, or the study of linguistic sounds, is one of the principal tools of the historical linguist. Phonology is useful as a historical tool, because the sounds people utter tend to change over time in certain directions and not in others.

The direction of phonetic change is governed by two kinds of constraints: those that are generally applicable across most languages, and those specific to a single language or a related group of languages. General constraints are imposed by the mechanical limits of the human vocal anatomy, the need to issue sounds that can be distinguished and understood by listeners, and the tendency to simplify sound combinations that are difficult to pronounce. Constraints within languages are imposed by the limited range of sounds that are acceptable and meaningful for that language. Often these language-specific sounds are very recognizable. Comedians can make us laugh by speaking nonsense if they do it in the characteristic phonology of French or Italian, for example. Armed with a knowledge of both the *general* tendencies in the direction of phonetic change and the *specific* phonetic conventions within a given language group, a linguist can arrive at reliable conclusions about which phonetic variants are early pronunciations and which come later. This is the first step in reconstructing the phonological history of a language.

We know that French developed historically from the dialects of Latin spoken in the Roman province of Gaul (modern France) during the waning centuries of the Roman Empire around 300–400 CE. As late as the 1500s vernacular French suffered from low prestige among scholars, as it was considered nothing more than a corrupt form of Latin. Even if we knew nothing about that history, we could examine the Latin *centum* (pronounced [kentum]), and the French *cent* (pronounced [sohnt]), both meaning "hundred," and we could say that the sound of the Latin word makes it the older form, that the Modern French form could have developed from it according to known rules of sound change, and that an intermediate pronunciation, [tsohnt], probably existed before the modern form appeared—and we would be right.

Some Basic Rules of Language Change: Phonology and Analogy

Two general phonetic rules help us make these decisions. One is that initial hard consonants like *k* and hard *g* tend to change toward soft sounds like *s* and *sh* if they change at all, whereas a change from *s* to *k* would generally be unusual. Another is that a consonant pronounced as a stop in the back of the mouth (*k*) is particularly likely to shift toward the front of the mouth (*t* or *s*) in a word where it is followed by a vowel that is pronounced in the front of the mouth (*e*). Pronounce [ke-] and [se-], and note the position of your tongue. The *k* is pronounced by using the back of the tongue and both *e* and *s* are formed with the middle or the tip of the tongue,

which makes it easier to pronounce the segment *se-* than the segment *ke-*. Before a front vowel like *-e* we might expect the *k-* to shift forward to [ts-] and then to [s-] but not the other way around.

This is an example of a general phonetic tendency called *assimilation*: one sound tends to assimilate to a nearby sound in the same word, simplifying the needed movements. The specific type of assimilation seen here is called *palatalization*—a back consonant (*k*) followed by a front vowel (*e*) was assimilated in French toward the front of the palate, changing the [k] to [s]. Between the Latin [k] (pronounced with the back of the tongue at the back of the palate) and the Modern French [s] (tip of the tongue at the front of the palate) there should have been an intermediate pronunciation *ts* (middle of the tongue at the middle of the palate). Such sequences permit historical linguists to reconstruct undocumented intermediate stages in the evolution of a language. Palatalization has been systematic in the development of French from Latin. It is responsible for much of the distinctive phonology of the French language.

Assimilation usually changes the quality of a sound, or sometimes removes sounds from words by slurring two sounds together. The opposite process is the *addition* of new sounds to a word. A good example of an innovation of this kind is provided by the variable pronunciations of the word *athlete* in English. Many English speakers insert [-uh] in the middle of the word, saying [ath-uh-lete], but most are not aware they are doing so. The inserted syllable always is pronounced precisely the same way, as [-uh], because it assimilates to the tongue position required to pronounce the following *-l*. Linguists could have predicted that some speakers would insert a vowel in a difficult cluster of consonants like *-thl* (a phenomenon called *epenthesis*) and that the vowel inserted in *athlete* always would be pronounced [-uh] because of the rule of assimilation.

Another kind of change is *analogical* change, which tends to affect grammar quite directly. For example, the *-s* or *-es* ending for the plural of English nouns was originally limited to one class of Old English nouns: *stān* for *stone* (nominative singular), *stānas* for *stones* (nominative plural). But when a series of sound changes (see note 5) resulted in the loss of the phonemes that had once distinguished nouns of different classes, the *-s* ending began to be reinterpreted as a *general* plural indicator and was attached to all nouns. Plurals formed with *-n* (oxen), with a zero change (sheep), and with a vowel change in the stem (women) remain as relics of Old English, but the shift to *-s* is driving out such "irregular" forms and has been doing so for eight hundred years. Similar analogical changes have affected verbs: *help/helped* has replaced Old English *help/holp* as the *-ed* ending has been

reinterpreted as a general ending for the past tense, reducing the once large number of strong verbs that formed their past with a vowel change. Analogical changes can also create new words or forms by analogy with old ones. Words formed with *-able* and *-scape* exist in such great numbers in English because these endings, which were originally bound to specific words (*measurable, landscape*), were reinterpreted as suffixes that could be removed and reattached to any stem (*touchable, moonscape*).

Phonological and analogical change are the internal mechanisms through which novel forms are incorporated into a language. By examining a sequence of documents within one language lineage from several different points in the past—inscriptions in, say, classical Latin, late vulgar Latin, early Medieval French, later Medieval French, and modern French—linguists have defined virtually all the sound changes and analogical shifts in the evolution of French from Latin. Regular, systematic rules, applicable also to other cases of language change in other languages, explain most of these shifts. But how do linguists replay these shifts "backward" to discover the origins of modern languages? How can we reconstruct the sounds of a language like Proto-Indo-European, for which there are *no* documents, a language spoken before writing was invented?

"Hundred": An Example of Phonetic Reconstruction

Proto-Indo-European words were not reconstructed to create a dictionary of Proto-Indo-European vocabulary, although they are extraordinarily useful in this way. The real aim in reconstruction is to prove that a list of daughter terms are cognates, descended from the same mother term. The reconstruction of the mother term is a by-product of the comparison, the proof that every sound in every daughter word can be derived from a sound in the common parent. The first step is to gather up the suspected daughters: you must make a list of all the variants of the word you can find in the Indo-European languages (table 2.1). You have to know the rules of phonological change to do even this successfully, as some variants of the word might have changed radically in sound. Just recognizing the candidates and making up a good list can be a challenge. We will try this with the Proto-Indo-European word for "hundred." The Indo-European roots for numbers, especially 1 to 10, 100, and 1,000, have been retained in almost all the Indo-European daughters.

Our list includes Latin *centum*, Avestan *satəm*, Lithuanian *šimtas*, and Old Gothic *hunda-* (a root much like *hunda-* evolved into the English word *hundred*). Similar-looking words meaning "hundred" in other

Table 2.1
Indo-European Cognates for the Root "Hundred"

Branch	*Language*	*Term*	*Meaning*
Celtic	Welsh	cant	hundred
	Old Irish	cēt	hundred
Italic	Latin	centum	hundred
Tocharian	TochA	känt	hundred
	TochB	kante	hundred
Greek	Greek	ἑκατόν	hundred
Germanic	Old English	hund	hundred
	OldHighGerm.	hunt	hundred
	Gothic	hunda	100, 120
	OldSaxon	hunderod	(long) hundred
Baltic	Lithuanian	šimtas	hundred
	Latvian	simts	hundred
Slavic	OldChurchSlav.	sŭto	hundred
	Bulgarian	sto	hundred
Anatolian	Lycian	sñta	unit of 10 or 100
Indo-Iranian	Avestan	satəm	hundred
	OldIndic	śatám	hundred

Indo-European languages should be added, and I have already referred to the French word *cent*, but I will use only four for simplicity's sake. The four words I have chosen come from four Indo-European branches: Italic, Indo-Iranian, Baltic, and Germanic.

The question we must answer is this: Are these words phonetically transformed daughters of a single parent word? If the answer is yes, they are cognates. To prove they are cognates, we must be able to reconstruct an ancestral sequence of phonemes that could have developed into all the documented daughter sounds through known rules. We start with the first sound in the word.

The initial [k] phoneme in Latin *centum* could be explained if the parent term began with a [k] sound as well. The initial soft consonants ([s] [sh])

in Avestan *satəm* and Lithuanian *šimtas* could have developed from a Proto-Indo-European word that began with a hard consonant [k], like Latin *centum,* since hard sounds generally tend to shift toward soft sounds if they change at all. The reverse development ([s] or [sh] to [k]) would be very unlikely. Also, palatalization and sibilation (shifting to a 's' or 'sh' sound) of initial hard consonants is expected in both the Indic branch, of which Vedic Sanskrit is a member; and the Baltic branch, of which Lithuanian is a member. The general direction of sound change and the specific conventions in each branch permit us to say that the Proto-Indo-European word from which all three of these developed could have begun with 'k'.

What about *hunda*? It looks quite different but, in fact, the *h* is expected—it follows a rule that affected all initial [k] sounds in the Germanic branch. This shift involved not just *k* but also eight other consonants in Pre-Germanic.[6] The consonant shift spread throughout the prehistoric Pre-Germanic language community, giving rise to a new Proto-Germanic phonology that would be retained in all the later Germanic languages, including, ultimately, English. This consonant shift was described by and named after Jakob Grimm (the same Grimm who collected fairy tales) and so is called Grimm's Law. One of the changes described in Grimm's Law was that the archaic Indo-European sound [k] shifted in most phonetic environments to Germanic [h]. The Indo-European *k* preserved in Latin *centum* shifted to *h* in Old Gothic *hunda-*; the initial *k* seen in Latin *caput* 'head' shifted to *h* in Old English *hafud* 'head'; and so on throughout the vocabulary. (C*aput*>*hafud* shows that *p* also changed to *f,* as in *pater*>*fater*). So, although it looks very different, *hunda-* conforms: its first consonant can be derived from *k* by Grimm's Law.

The first sound in the Proto-Indo-European word for "hundred" probably was *k*. (An initial [k] sound satisfies the other Indo-European cognates for "hundred" as well.)[7] The second sound should have been a vowel, but which vowel?

The second sound was a vowel that does not exist in English. In Proto-Indo-European resonants could act as vowels, similar to the resonant *n* in the colloquial pronunciation of *fish'n'* (as in *Bob's gone fish'n'*). The second sound was a resonant, either **m̥* or **n̥*, both of which occur among the daughter terms being compared. (An asterisk is used before a reconstructed form for which there is no direct evidence.) *M* is attested in the Lithuanian cognate *šimtas*. An *m* in the Proto-Indo-European parent could account for the *m* in Lithuanian. It could have changed to *n* in Old Indic, Germanic, and other lineages by assimilating to the following *t* or *d*, as both *n* and *t* are articulated on the teeth. (Old Spanish *semda* 'path'

changed to modern Spanish *senda* for the same reason.) A shift from an original *m* to an *n* before a *t* is explicable, but a shift from an original *n* to an *m* is much less likely. Therefore, the original second sound probably was ṃ. This consonant could have been lost entirely in Sanskrit *satam* by yet another assimilative tendency called total assimilation: after the *m* changed to *n*, giving **santam*, the *n* was completely assimilated to the following *t*, giving *satam*. The same process was responsible for the loss of the [k] sound in the shift from Latin *octo* to modern Italian *otto* 'eight'.

I will stop here, with an ancestral **k'ṃ-*, in my discussion of the Proto-Indo-European ancestor of *centum*. The analysis should continue through the phonemes that are attested in all the surviving cognates to reconstruct an acceptable ancestral root. By applying such rules to all the cognates, linguists have been able to reconstruct a Proto-Indo-European sequence of phonemes, **k'ṃtom*, that could have developed into all the attested phonemes in all the attested daughter forms. The Proto-Indo-European root **k'ṃtom* is the residue of a successful comparison—it is the proof that the daughter terms being compared are indeed cognates. It is also likely to be a pretty good approximation of the way this word was pronounced in at least some dialects of Proto-Indo-European.

The Limitations and Strengths of Reconstruction

The comparative method will produce the *sound* of the ancestral root and confirm a genetic relationship *only* with a group of cognates that has evolved regularly according to the rules of sound change. The result of a comparative analysis is either a demonstration of a genetic connection, if every phoneme in every cognate can be derived from a mutually acceptable parental phoneme; or no *demonstrable* connection. In many cases sounds may have been borrowed into a language from a neighboring language, and those sounds might replace the predicted shifts. The comparative method cannot force a regular reconstruction on an irregular set of sounds. Much of the Proto-Indo-European vocabulary, perhaps most of it, never will be reconstructed. Regular groups of cognates permit us to reconstruct a Proto-Indo-European root for the word *door* but not for *wall*; for *rain* but not for *river*; for *foot* but not for *leg*. Proto-Indo-European certainly had words for these things, but we cannot safely reconstruct how they sounded.

The comparative method cannot prove that two words are *not* related, but it can fail to produce proof that they *are*. For example, the Greek god Ouranos and the Indic deity Varuna had strikingly similar mythological attributes, and their names sound somewhat alike. Could Ouranos and

Varuna be reflexes of the name of some earlier Proto-Indo-European god? Possibly—but the two names cannot be derived from a common parent by the rules of sound change known to have operated in Greek and Old Indic. Similarly Latin *deus* (god) and Greek *théos* (god) look like obvious cognates, but the comparative method reveals that Latin *deus*, in fact, shares a common origin with Greek *Zéus*.[8] If Greek *théos* were to have a Latin cognate it should begin with an [f] sound (*festus* 'festive' has been suggested, but some of the other sounds in this comparison are problematic). It is still possible that *deus* and *théos* were historically related in some irregular way, but we cannot prove it.

In the end, how can we be sure that the comparative method accurately reconstructs undocumented stages in the phonological history of a language? Linguists themselves are divided on the question of the "reality" of reconstructed terms.[9] A reconstruction based on cognates from eight Indo-European branches, like **k'm̥tom-*, is much more reliable and probably more "true" than one based on cognates in just two branches. Cognates in at least three branches, including an ancient branch (Anatolian, Greek, Avestan Iranian, Old Indic, Latin, some aspects of Celtic) should produce a reliable reconstruction. But how reliable? One test was conceived by Robert A. Hall, who reconstructed the shared parent of the Romance languages using just the rules of sound change, and then compared his reconstruction to Latin. Making allowances for the fact that the actual parents of the Romance languages were several provincial Vulgar Latin dialects, and the Latin used for the test was the classical Latin of Cicero and Caesar, the result was reassuring. Hall was even able to reconstruct a contrast between two sets of vowels although none of the modern daughters had retained it. He was unable to identify the feature that distinguished the two vowel sets as length—Latin had long vowels and short vowels, a distinction lost in all its Romance daughters—but he was able to rebuild a system with two contrasting sets of vowels and many of the other, more obvious aspects of Latin morphology, syntax, and vocabulary. Such clever exercises aside, the best proof of the realism of reconstruction lies in several cases where linguists have suggested a reconstruction and archaeologists have subsequently found inscriptions that proved it correct.[10]

For example, the oldest recorded Germanic cognates for the word *guest* (Gothic *gasts*, Old Norse *gestr*, Old High German *gast*) are thought to be derived from a reconstructed late Proto-Indo-European **ghos-ti-* (which probably meant both "host" and "guest" and thus referred to a relationship of hospitality between strangers rather than to one of its roles) through a Proto-Germanic form reconstructed as **gastiz*. None of the

known forms of the word in the later Germanic languages contained the *i* before the final consonant, but rules of sound change predicted that the *i* should theoretically have been there in Proto-Germanic. Then an archaic Germanic inscription was found on a gold horn dug from a grave in Denmark. The inscription *ek hlewagastiz holitijaz* (or *holtingaz*) *horna tawido* is translated "I, Hlewagasti of Holt (or Holting) made the horn." It contained the personal name Hlewagastiz, made up of two stems, *Hlewa-* 'fame' and *gastiz* 'guest'. Linguists were excited not because the horn was a beautiful golden artifact but because the stem contained the predicted *i*, verifying the accuracy of both the reconstructed Proto-Germanic form and its late Proto-Indo-European ancestor. Linguistic reconstruction had passed a real-world test.

Similarly linguists working on the development of the Greek language had proposed a Proto-Indo-European labiovelar $*k^w$ (pronounced [kw-]) as the ancestral phoneme that developed into Greek *t* (before a front vowel) or *p* (before a back vowel). The reconstruction of $*k^w$ was a reasonable but complex solution for the problem of how the Classical Greek consonants were related to their Proto-Indo-European ancestors. It remained entirely theoretical until the discovery and decipherment of the Mycenaean Linear B tablets, which revealed that the earliest form of Greek, Mycenaean, had the predicted k^w where later Greek had *t* or *p* before front and back vowels.[11] Examples like these confirm that the reconstructions of historical linguistics are more than just abstractions.

A reconstructed term is, of course, a phonetic idealization. Reconstructed Proto-Indo-European cannot capture the variety of dialectical pronunciations that must have existed more than perhaps one thousand years when the language was living in the mouths of people. Nevertheless, it is a remarkable victory that we can now pronounce, however stiffly, thousands of words in a language spoken by nonliterate people before 2500 BCE.

The Lexicon: How to Reconstruct Dead Meanings

Once we have reconstructed the *sound* of a word in Proto-Indo-European, how do we know what it *meant*? Some archaeologists have doubted the reliability of reconstructed Proto-Indo-European, as they felt that the original meanings of reconstructed terms could never be known confidently.[12] But we can assign reliable meanings to many reconstructed Proto-Indo-European terms. And it is in the meanings of their words that we find the best evidence for the material culture, ecological environment,

social relations, and spiritual beliefs of the speakers of Proto-Indo-European. Every meaning is worth the struggle.

Three general rules guide the assignment of meaning. First, look for the most ancient meanings that can be found. If the goal is to retrieve the meaning of the original Proto–Indo–European word, modern meanings should be checked against meanings that are recorded for ancient cognates.

Second, if one meaning is consistently attached to a cognate in all language branches, like *hundred* in the example I have used, that is clearly the least problematic meaning we can assign to the original Proto-Indo-European root. It is difficult to imagine how that meaning could have become attached to all the cognates unless it were the meaning attached to the ancestral root.

Third, if the word can be broken down into roots that point to the same meaning as the one proposed, then that meaning is doubly likely. For example, Proto-Indo-European **k'm̥tom* probably was a shortened version of **dek'm̥tom*, a word that included the Proto-Indo-European root **dek'm̥* 'ten'. The sequence of sounds in **dek'm̥* was reconstructed independently using the cognates for the word *ten*, so the fact that the reconstructed roots for *ten* and *hundred* are linked in both meaning and sound tends to verify the reliability of both reconstructions. The root **k'm̥tom* turns out to be not just an arbitrary string of Proto-Indo-European phonemes but a meaningful compound: "(a unit) of tens." This also tells us that the speakers of Proto-Indo-European had a decimal numbering system and counted to one hundred by tens, as we do.

In most cases the meaning of a Proto-Indo-European word changed and drifted as the various speech communities using it became separated, centuries passed, and daughter languages evolved. Because the association between word and meaning is arbitrary, there is less regular directionality to change in meaning than there is in sound change (although some semantic shifts are more probable than others). Nevertheless, general meanings can be retrieved. A good example is the word for "wheel."

"Wheel": An Example of Semantic Reconstruction

The word *wheel* is the modern English descendant of a PIE root that had a sound like **k^{w}ék^{w}los* or **k^{w}ek^{w}lós*. But what, exactly, did **k^{w}ék^{w}los* mean in Proto-Indo-European? The sequence of phonemes in the root **k^{w}ék^{w}los* was pieced together by comparing cognates from eight old Indo-European languages, representing five branches. Reflexes of this word survived in Old Indic and Avestan (from the Indo-Iranian branch), Old Norse and

Old English (from the Germanic branch), Greek, Phrygian, and Tocharian A and B. The meaning "wheel" is attested for the cognates in Sanskrit, Avestan, Old Norse, and Old English. The meaning of the Greek cognate had shifted to "circle" in the singular but in the plural still meant "wheels." In Tocharian and Phrygian the cognates meant "wagon" or "vehicle." What was the original meaning? (table 2.2).

Five of the eight **k^{w}ék^{w}los* cognates have "wheel" or "wheels" as an attested meaning, and in those languages (Phrygian, Greek, Tocharian A & B) where the meaning drifted away from "wheel(s)," it had not drifted far ("circle," "wagon," or "vehicle"). Moreover, the cognates that preserve the meaning "wheel" are found in languages that are geographically isolated from one another (Old Indic and Avestan in Iran were neighbors, but neither had any known contact with Old Norse or Old English). The meaning "wheel" is unlikely to have been borrowed into Old Norse from Old Indic, or vice versa.

Some shifts in meaning are unlikely, and others are common. It is common to name a whole ("vehicle," "wagon") after one of its most characteristic parts ("wheels"), as seems to have happened in Phrygian and Tocharian. We do the same in modern English slang when we speak of someone's car as their "wheels," or clothing as their "threads." A shift in meaning in the other direction, using a word that originally referred to the whole to refer to one of its parts (using *wagon* to refer to *wheel*), is much less probable.

The meaning of *wheel* is given additional support by the fact that it has an Indo-European etymology, like the root for **k'ṃtom*. It was a word created from another Indo-European root. That root was **k^{w}el-*, a verb that meant "to turn." So **k^{w}ék^{w}los* is not just a random string of phonemes reconstructed from the cognates for *wheel*; it meant "the thing that turns." This not only tends to confirm the meaning "wheel" rather than "circle" or "vehicle" but it also indicates that the speakers of Proto-Indo-European made up their own words for wheels. If they learned about the invention of the wheel from others they did not adopt the foreign name for it, so the social setting in which the transfer took place probably was brief, between people who remained socially distant. The alternative, that wheels were invented within the Proto-Indo-European language community, seems unlikely for archaeological and historical reasons, though it remains possible (see chapter 4).

One more rule helps to confirm the reconstructed meaning. If it fits within a semantic field consisting of other roots with closely related reconstructed meanings, we can at least be relatively confident that such a word

Table 2.2
Proto-Indo-European Roots for Words Referring to Parts of a Wagon

PIE Root Word	*Wagon Part*	*Daughter Languages*
*k^wek^wlos	(wheel)	*Old Norse* hvēl 'wheel'; *Old English* hweohl 'wheel'; *Middle Dutch* wiel 'wheel'; *Avestan Iranian* čaxtra- 'wheel'; *Old Indic* cakrá 'wheel, Sun disc'; *Greek* kuklos 'circle' and kukla (plural) 'wheels'; *Tocharian A* kukal 'wagon'; *Tocharian B* kokale 'wagon'
*rot-eh_2-	(wheel)	*Old Irish* roth 'wheel'; *Welsh* rhod 'wheel'; *Latin* rota 'wheel'; *Old High German* rad 'wheel'; *Lithuanian rātas* 'wheel'; *Latvian* rats 'wheel' and rati (plural) 'wagon'; *Albanian* rreth 'ring, hoop, carriage tire'; *Avestan Iranian* ratha 'chariot, wagon'; *Old Indic* rátha 'chariot, wagon'
*ak*s-, or	(axle)	*Latin* axis 'axle, axis'; *Old English* eax 'axle'; *Old High German* *$h_a$ek*s- ahsa 'axle'; *Old Prussian* assis 'axle'; *Lithuanian* ašís 'axle'; *Old Church Slavonic* osĭ 'axle'; *Mycenaean Greek* a-ko-so-ne 'axle'; *Old Indic* áks*a 'axle'
*ei-/*oi-, or	(thill)	*Old English* ār- 'oar'; *Russian* vojë 'shaft'; *Slovenian* oje 'shaft'; *Hittite* h_2ih_3s or hišša- 'pole, harnessing shaft'; *Greek* oisioi* 'tiller, rudderpost'; *Avestan Iranian* aēša 'pair of shafts, plow-pole'; *Old Indic* īs*a 'pole, shaft'
*wéĝheti-	(ride)	*Welsh* amwain 'drive about'; *Latin* vehō 'bear, convey'; *Old Norse* vega 'bring, move'; *Old High German* wegan 'move, weigh'; *Lithuanian* vežù 'drive'; *Old Church Slavonic* vezọ 'drive'; *Avestan Iranian* vazaiti 'transports, leads'; *Old Indic* váhati 'transports, carries, conveys'. Derivative nouns have the meaning "wagon" in *Greek*, *Old Irish*, *Welsh*, *Old High German*, and *Old Norse*.

could have existed in Proto-Indo-European. "Wheel" is part of a semantic field consisting of *words for the parts of a wagon or cart* (table 2.2). Happily, at least four other such words can be reconstructed for Proto-Indo-European. These are:

1. **rot-eh$_2$-*, a second term for "wheel," with cognates in Old Indic and Avestan that meant "chariot," and cognates that meant "wheel" in Latin, Old Irish, Welsh, Old High German, and Lithuanian.
2. **aks-* (or perhaps **h$_2$eks-*) 'axle' attested by cognates that had not varied in meaning over thousands of years, and still meant "axle" in Old Indic, Greek, Latin, Old Norse, Old English, Old High German, Lithuanian, and Old Church Slavonic.
3. **h$_2$ih$_3$s-* 'thill' (the harness pole) attested by cognates that meant "thill" in Hittite and Old Indic.
4. **wégheti*, a verb meaning "to convey or go in a vehicle," attested by cognates carrying this meaning in Old Indic, Avestan, Latin, Old English, and Old Church Slavonic and by cognate-derived nouns ending in **-no-* meaning "wagon" in Old Irish, Old English, Old High German, and Old Norse.

These four additional terms constitute a well-documented semantic field (*wheel*, *axle*, *thill*, and *wagon* or *convey in a vehicle*) that increases our confidence in reconstructing the meaning "wheel" for **k^{w}ék^{w}los*. Of the five terms assigned to this semantic field, all but *thill* have clear Indo-European etymologies in independently reconstructed roots. The speakers of Proto-Indo-European were familiar with wheels and wagons, and used words of their own creation to talk about them.

Fine distinctions, shades of meaning, and the word associations that enriched Proto-Indo-European poetry may be forever lost, but gross meanings are recoverable for at least fifteen hundred Proto-Indo-European roots such as **dekm-* 'ten', and for additional thousands of other words derived from them, such as **kṃtom-* 'hundred'. Those meanings provide a window into the lives and thoughts of the speakers of Proto-Indo-European.

Syntax and Morphology: The Shape of a Dead Language

I will not try to describe in any detail the grammatical connections between the Indo-European languages. The reconstructed vocabulary is most important for our purposes. But grammar, the bedrock of language

classification, provides the primary evidence for classifying languages and determining relationships between them. Grammar has two aspects: *syntax,* or the rules governing the order of words in sentences; and *morphology,* or the rules governing the forms words must take when used in particular ways.

Proto-Indo-European grammar has left its mark on all the Indo-European languages to one degree or another. In all the Indo-European language branches, nouns are declined; that is, the noun changes form depending on how it is used in a sentence. English lost most of these declinations during its evolution from Anglo-Saxon, but all the other languages in the Germanic branch retain them, and we have kept some use-dependent pronouns (*masculine*: he, his, him/*feminine*: she, hers, her). Moreover, most Indo-European nouns are declined in similar ways, with endings that are genetically cognate, and with the same formal system of cases (nominative, genitive, accusative, etc.) that intersect in the same way with the same three gender classes (masculine, feminine, neuter); and with similar formal classes, or declensions, of nouns that are declined in distinctive ways. Indo-European verbs also share similar conjugation classes (first person, second person or familiar, third person or formal, singular, plural, past tense, present tense, etc.), similar stem alterations (run-ran, give-gave), and similar endings. This particular constellation of formal categories, structures, transformations, and endings is not at all necessary or universal in human language. It is unique, as a system, and is found only in the Indo-European languages. The languages that share this grammatical system certainly are daughters of a single language from which that system was inherited.

One example shows how unlikely it would be for the Indo-European languages to share these grammatical structures by random chance. The verb *to be* has one form in the first-person singular ([I] *am*) and another in the third-person singular ([he/she/it] *is*). Our English verbs are descended from the archaic Germanic forms *im* and *ist*. The Germanic forms have exact, proven cognates in Old Indic *ásmi* and *ásti*; in Greek *eimí* and *estí*; and in Old Church Slavonic *jesmĭ* and *jestŭ*. All these words are derived from a reconstructable Proto-Indo-European pair, **h*$_1$*e'smi* and **h*$_1$*e'sti*. That all these languages share the same system of verb classes (first person, second person or familiar, and third person), and that they use the same basic roots and endings to identify those classes, confirms that they are genetically related languages.

Conclusion: Raising a Language from the Dead

It will always be difficult to work with Proto-Indo-European. The version we have is uncertain in many morphological details, phonetically idealized, and fragmentary, and can be difficult to decipher. The meanings of some terms will never be fully understood, and for others only an approximate definition is possible. Yet reconstructed Proto-Indo-European captures key parts of a language that actually existed.

Some dismiss reconstructed Proto-Indo-European as nothing more than a hypothesis. But the limitations of Proto-Indo-European apply equally to the written languages of ancient Egypt and Mesopotamia, which are universally counted among the great treasures of antiquity. No curator of Assyrian records would suggest that we should discard the palace archives of Nineveh because they are incomplete, or because we cannot know the exact sound and meaning of many terms, or because we are uncertain about how the written court language related to the 'real' language spoken by the people in the street. Yet these same problems have convinced many archaeologists that the study of Proto-Indo-European is too speculative to yield any real historical value.

Reconstructed Proto-Indo-European is a long, fragmentary list of words used in daily speech by people who created no other texts. That is why it is important. The list becomes useful, however, only if we can determine where it came from. To do that we must locate the Proto-Indo-European homeland. But we cannot locate the Proto-Indo-European homeland until we first locate Proto-Indo-European in time. We have to know *when* it was spoken. Then it becomes possible to say where.