

The Linguistic Diversity of Aboriginal Europe

January 6, 2009 @ 4:42 pm · Filed by [Mark Liberman](#) under [Linguistic history](#)

What was Europe like, linguistically speaking, between the end of the last ice age and the coming of the Indo-European languages? This question has been in the background of many Language Log posts over the years. Not long ago, in the hallway between our offices, I asked [Don Ringe](#) for a summary of the state of knowledge on this issue. His response was so interesting — as conversations with Don generally are — that I asked him if he'd write something for Language Log on the topic. The result is below.

[Guest post by Don Ringe]

How to solve the problem.

What the languages of prehistoric Europe might have been like has recently become a focus of renewed interest. Most people seem to assume that we can do no better than speculate. But we can do better, because all the historical sciences, including historical linguistics, have a basic tool for investigating the unobservable past: the Uniformitarian Principle (UP).

The basic idea behind the UP is that the unobservable past must have been like the observable present, *insofar as relevant conditions have not changed in the meantime*. (The proviso is very important; see further below.) For linguistics the UP can be stated more precisely as follows:

Unless we can demonstrate significant changes in the conditions of language acquisition and use between some time in the unobservable past and the present, we must assume that the same types and distributions of structures, variation, changes, etc. existed at that time in the past as in the present.

We use the UP to interpret the linguistic documents of the past, which are always less precise, detailed, and comprehensive than modern data, in terms of what we know about modern languages that have been described scientifically; and we also use the UP to extrapolate from the documented past into prehistory, and across gaps in the historical record.

But the qualification in the UP—that conditions have to remain comparable—is crucial, and never more so than in the case of Europe. Conditions in Europe have obviously changed radically over the past three millennia. True, one's native language is still

learned in the first few years of life, and is still learned from parents, caregivers, and older playmates; that's clearly been true since language became a universal characteristic of our species, probably at least 100,000 years ago. And it's still true that language is most often used to communicate with and express solidarity with one's family, friends, and neighbors. But everything else has changed. Most Europeans now live in powerful states that insist on the use of one or two languages throughout their territories; that's the main reason why most local languages like Occitan and Breton are slowly dying out. Moreover, most of those states officially recognize only one dialect of each of their official languages—a "standard" dialect. Government ministers, people in the media, and other important people all speak the same dialect in public, and that is the one dialect that everyone is exposed to and taught to think of as "correct". Many Europeans talk regularly to people outside their own communities, and when they do, they typically use one of the standard dialects. In some countries there are educated people whose native dialect *is* the standard dialect. All these phenomena have contributed to a dramatic linguistic homogenization of the European continent.

None of these state-related conditions can possibly have existed in the pre-state societies of the European Iron Age and earlier. It follows that modern Europe is *not an appropriate model* for applying the UP to prehistoric Europe; we need a model that approximates prehistoric conditions in Europe much better.

In fact such a model already exists. Johanna Nichols' groundbreaking article on linguistic diversity (Nichols 1990) is based on an exhaustive worldwide survey of languages and families at the time of first sustained European contact during the colonial expansion of European empires. At that time much of the world—most of the Americas and sub-Saharan Africa, Australasia, Oceania, the Southeast Asian highlands, Siberia—harbored only pre-state societies, or states which were too small or too weak to impose any significant linguistic homogenization on extensive populations. Nichols' results, used with care, thus provide a suitable model from which to deduce the linguistic situation in prehistoric Europe by use of the UP.

Pre-state patterns of language diversity.

The basic fact of pre-state language distribution is that no single language can occupy, for more than a few centuries, an area too large for all its native speakers to communicate with each other regularly. The reasons for that are simple and obvious. All languages change, slowly but steadily, over time. Each change originates in a small part of the speaking population and spreads outward through the speech community. (See e.g. Labov 2001 for an in-depth discussion of this process as actually observed in contemporary language communities.) Many changes either spread through the entire

community over two or three generations or are suppressed by social “stigmatization”; some are accepted by some parts of the community but not by others, creating “dialect” differences within the broader speech community. But if parts of the speech community cease to communicate altogether, or communicate so rarely that they have no incentive to imitate each others’ speech, changes cannot spread from one to another; different changes will accumulate on either side of the linguistic barrier, and within a thousand years, at most, a single language will have become two or more. (For a discussion of this process in detail see e.g. Ross 1997.)

Thus in pre-state communities every language spread automatically results in language fragmentation. Of course not all the fragments survive; pre-state language communities sometimes gradually abandon their native language and adopt the language of another community with which they are in intimate contact, as linguists working in the highlands of New Guinea have observed (Foley 1986:24-5). But the fragments that do survive continue to diverge, century after century, until the original connections between them can no longer be discovered with any certainty. Extensive structural convergence of languages, as opposed to mere word-borrowing or the adoption of a few superficial traits, turns out to be rare, evidently because of the way native languages are learned (see e.g. Fantini 1985, Meisel 1989 with references, Bhatia and Ritchie 1999:574-5); divergence or death is the normal fate of languages. (By contrast, dialects that are mutually intelligible can and do merge; but that doesn’t decrease linguistic diversity very much.) The result is a diversity not only of languages, but also of language families, within a pre-state geographical area of any significant size.

But not all pre-state areas are equally diverse linguistically; that was one of the many interesting findings of Nichols 1990. Her discussion of the patterns and causes of linguistic diversity (Nichols 1990:477-94) is worth reading in detail, but some of the basic principles underlying the patterns are most relevant here. In the following discussion note that “lineages” refers to genetic units of any size—languages, obvious families (such as Germanic or Romance or Slavic), or “stocks” (differentiated families of the largest size discoverable by scientific methods, such as Indo-European); it is assumed that comparisons will be made between comparable units in evaluating an area’s linguistic diversity. The following general principles hold:

- “Other things being equal, density of lineages is substantially greater at low latitudes than at high latitudes.” (Nichols 1990:484)
- “Other things being equal, the coastal area of a continent will generally have substantially greater lineage density than the interior. Not every coastal area is

high in lineage density, but the extensive areas of high density are all on or near coastlines. ... [Because of its richer resources, the] seacoast offers the possibility of economic self-sufficiency for a small group occupying a small territory." (ibid. pp. 484-5)

- "The discrepancy in the lineage density of coastline and interior is most pronounced where the interior is relatively dry (ibid. p. 485)
- "The cause of high lineage density in mountain areas is generally attributed to the fact that mountainous geography naturally isolates populations, resists large-scale economic integration, and creates refuge zones." (ibid. p. 485)
- "Density of lineages is low in areas dominated by large-scale economies, higher in areas with smaller-scale economies. ... Reduction of lineage density in response to increased scale of economy is not immediate, as shown by the ancient Near East." (ibid. p. 486)

As Nichols herself notes (p. 488), it all boils down to scale of economy: in areas where a small group can support itself in a small area, small groups do exactly that, and over time their languages steadily diverge; in areas in which populations must range over a large area in order to survive, we find lineages occupying correspondingly larger areas—though the languages in question are not necessarily spoken by larger populations. Not surprisingly, "[w]hen a language or family is distributed over an area favoring high density and one favoring low density, we can expect to see corresponding changes in the rate or geographical scale of differentiation" (Nichols 1990:488-9), and Nichols adduces some examples of language families which are "compressed" near a coast but "elongated" in the interior (ibid. p. 489).

In prehistoric Europe, then, we should expect to find the following pattern of languages and families, roughly speaking:

- numerous languages, belonging to many families not provably related to each other, in the Mediterranean coastal zone, including virtually all of Greece and Italy;
- somewhat less, but still notable, diversity along the cooler Atlantic coast, including the British isles;
- still less diversity in the interior of the continent (though not markedly less, given the adequate rainfall that Europe enjoys)—except probably for the Alps and the mountainous parts of the Balkan peninsula, which are likely to have been refugia for small and linguistically diverse populations, much like the modern Caucasus;

- fairly little diversity in Scandinavia—though probably not less than exists today, with two different language families belonging to different stocks (!).

How does this compare with what we actually know about the distribution of languages in Europe at the dawn of history?

Attested languages in early Europe.

Most of the information that we possess about European languages before the Roman period comes from the Mediterranean area, simply because writing was adopted much earlier in the Mediterranean basin than elsewhere. The following languages are securely attested (see e.g. Vetter 1953, Buck 1955, Lejeune 1971, 1974, Prosdocimi 1978, Poultney 1979, Untermann 1980, 2001, Duval 1985, Marinetti 1985, Eska 1989, Rix 1998, Bonfante and Bonfante 2002, Wallace 2007, and many of the articles in Woodard 2004).

1. Indo-European (IE) languages:

- Greek, splintered into about two dozen (known) dialects, in Greece, the Aegean islands, and areas further east (the Asia Minor seaboard, Pamphylia, Cyprus); clearly one language, very different from all others;
- Messapic, in southeastern Italy, largely uninterpretable but with proper names exhibiting IE nominal morphology;
- Venetic, in the lowlands of northeastern Italy;
- the Italic subfamily, divided into two divergent branches:
 - Latino-Faliscan, including
 - Latin, originally confined to Latium, and
 - Faliscan, spoken at Falerii on the upper Tiber (surrounded by Etruscan territory);
 - Sabellian, including
 - South Picene, east of the Appennines, and
 - a dialect continuum from Oscan in the south through the hill dialects east of Rome to Umbrian, as well as
 - a poorly attested dialect spoken in Campania before the Samnite invasion which might or might not have been part of the same dialect continuum;

- Sikel, poorly attested from Sicily, might also have been an Italic language (Vetter 1953:359-60); it was clearly an IE language;
- the Celtic subfamily, represented by
 - Hispano-Celtic (Celtiberian) in northeastern Spain;
 - Cisalpine Celtic (Leponitic) around the lakes of northern Italy, and
 - Trans-alpine Celtic (Gaulish) in what is now France, which may or may not have been dialects of a single language.
- 2. The language of the Linear A script, uninterpretable but clearly neither IE nor Semitic (Packard 1974), sometimes called "Minoan".
- 3. The language of some uninterpretable inscriptions in the Greek alphabet from eastern Crete (Guarducci 1942:137-42), sometimes called "Eteocretan".
- 4. Elymian, known from some coins and epigraphical fragments from Sicily, with apparently non-IE nominal morphology.
- 5. Tyrrhenian languages (Rix 1998):
 - Etruscan, which is fairly well attested but of which only some words and some points of grammar are understood;
 - Lemnian, attested chiefly on a 6th-century BCE stele from Lemnos in the north-eastern Aegean, which resembles Etruscan in some detail;
 - possibly Raetic, in the valley of the upper Adige in northeastern Italy, which bears a less obvious resemblance to Etruscan.
- 6. The language of the stele of Novilara (east of San Marino) and a few other fragments (see Poultney 1979, whose suggestion that the language is IE is not very convincing).
- 7. Iberian, attested in inscriptions throughout southern and eastern Spain; attempts to link it with any other language, including Basque, have not succeeded (cf. Untermann 2001:27).
- 8. Tartessian, known from 78 stelae of unknown function found in southwestern Iberia (see Untermann 2001:28-32).

To these we must add Basque, which, though it happens not to have been written so early, must have occupied an area including that in which it is still spoken.

On the level of languages, and even on the level of obvious language families (Italic, Celtic, Tyrrhenian), this is the kind of diversity we expect to have existed in the

prehistoric Mediterranean: one substantial family (Italic), two smaller ones, and nine or ten languages that do not belong to any of them. The only other observation that needs to be made is that the record is certainly incomplete: we might reasonably expect there to have been quite a few other contemporary languages that were never recorded, for instance in the Alps, in Liguria, along the Dalmatian coast, and on Corsica, Sardinia, Malta, the Balearic islands, etc.

In one respect, however, the situation just described cannot be the same as it was in, say, 3000 BCE: far too many of the obvious families and more isolated languages belong to a single stock, IE. The multiple branching structure of the IE family is very untypical. Nichols' assessment of the structure of linguistic "family trees" worldwide is worth quoting at some length:

... most branchings are binary, and the usual result of branchings over time is the survival of one to three families per stock; isolates and one-branch stocks are common, representing over half of the total lineages. Groupings with three or more branchings are not uncommon, but for the most part they represent relatively recent splits at the family level or lower. ... [M]ore elaborate splits may be fairly common, but over time some consolidation and/or extinction takes place to reduce the survival rate. ... There are two conspicuous counterexamples in my database, ancient groups with elaborate first-order branchings: Indo-European and Afro-Asiatic. ... These are groups whose breakup and spread were precipitated by the development of nomadic or seminomadic stock breeding, which rapidly increased the scale of the economy. (Nichols 1990:489)

We will need to discuss the IE family at length below. For the moment the important point is that the spread of IE languages cannot be expected to have any parallels in the older prehistory of Europe. Before the arrival of speakers of IE languages in the Mediterranean, the linguistic situation must have been *even more* diverse; a reasonable estimate would be more than thirty languages—possibly many more—grouped into more than twenty families belonging to at least fifteen different stocks.

Of course the rest of the continent can be expected to have been somewhat less diverse linguistically, but only somewhat less. Given the number of areas that should have promoted modest diversity—the Atlantic coast, the Alps, the Balkans—it would be no surprise if the rest of the continent together exhibited a linguistic diversity similar to that of the Mediterranean region, with little overlap of families or stocks between the Mediterranean and the rest of the continent: perhaps sixty languages in Europe altogether, representing some forty families and thirty stocks. This is not an extreme estimate. Note that the archaeologist David Anthony, who is willing to contemplate

"language communities" (isolated languages or families of closely related languages) spread over territories the size of Yugoslavia or even France, estimates that there must have been between twenty and forty such communities in Europe in the late Neolithic period (Anthony 1991:196-8).

In the most general terms, aboriginal Europe should have exhibited a degree of linguistic diversity comparable to that of western North America, with the Mediterranean region comparable to aboriginal California, the Atlantic coast comparable to the northwest coast of North America, and the hinterlands very roughly comparable. (Europe might be expected to show less diversity because it is less mountainous than the interior of western North America, but more diversity because it is much less dry; probably the two factors would more or less cancel each other out.)

Of course that is not what we find today, and that is not what Europe was like even two millennia ago; already at that time large areas were occupied by people speaking Celtic and Germanic dialects that were not very diverse, showing that they had spread over the areas they occupied relatively recently. We need to discuss how that happened.

The spread of Indo-European languages.

Proto-Indo-European (PIE) was a single language for which a complex grammar and an extensive vocabulary can be reconstructed; for a sketch of PIE see e.g. Ringe 2008:4-66. It follows that the speakers of PIE *must* have occupied a comparatively small territory. The idea that PIE might once have been spoken over most of Europe isn't just unlikely; it's *impossible*, because it's so extravagant a violation of the UP. Exactly when and where PIE was spoken will probably be debated forever, but David Anthony presents an overwhelmingly strong case that it must have been somewhere in the steppes north of the Black and Caspian Seas around 4000 BCE (Anthony 2007; see also Mallory 1989). Much of his case is based on incontrovertible linguistic evidence. For instance, the fact that a word for 'horse' is solidly reconstructable for PIE (with reflexes in *all* the earliest-attested branches of the family, including Anatolian) rules out Mesopotamia, Anatolia, and any forested part of Europe as the area where PIE was spoken; the fact that words for 'wool', 'yoke', and 'thill' are also reconstructable for PIE, and that a word for 'wheel' is reconstructable for the last common ancestor of all the non-Anatolian branches of the family, eliminates any date much earlier than 4000 BCE. Later dates are eliminated by the fact that, by the time we have records of them, Hittite, Vedic Sanskrit, and Greek are so different from each other that they must have been diverging for two millennia or so.

It follows that the appearance of IE languages in much of Europe at an early date must reflect a considerable spread of IE languages from their point of origin. Many commentators, for a great variety of reasons, would like to believe that that spread occurred without any significant population movements; but that, too, violates the UP. It has to be remembered that those IE languages spread not as trade languages for some specialized use, but as *native* languages; and all our contemporary experience shows that a language can acquire new populations of native speakers only if already existing native speakers are in intimate contact with communities speaking other languages. One can imagine an IE language spreading from village to village through intermarriage, but if that's what happened, the spread must have been slow; the frontier of IE-speaking territory might have advanced, say, a depth of six villages per century by such a process—and in the meantime the IE language that was spreading in that area would have been diversifying into dialects and eventually fragmenting into two or more languages (see above). In some areas that could be what happened. But there is no way that such a process could have resulted in a few closely related Celtic dialects or languages being spoken over a large continuous territory from the Atlantic seaboard to Bohemia and beyond—a situation that clearly existed around 500 BCE (cf. Mallory 1989:95-107). We cannot avoid the inference that there were substantial migrations of people speaking IE languages into Europe in the prehistoric period.

On the other hand, it seems impossible that the populations speaking IE languages could suddenly have become large enough to overrun vast territories and crowd out the earlier inhabitants; that scenario probably violates the UP too. But we do not need to posit vast folk migrations to explain the spread of IE languages. History shows that comparatively small groups of immigrants can induce their new neighbors to adopt their language — gradually over several generations, of course — if (1) they have enough economic or political power, and (2) they can offer some important advantage to those who are willing to adopt their language. Anthony 2007:117-9 gives some interesting examples of such “elite recruitment” from the historical record. Though the details will have differed from case to case (and of course are unrecoverable), it seems clear that that was how many prehistoric IE languages spread. The result should be that, while most Europeans' linguistic ancestors were speakers of PIE, many or even most of their biological ancestors at the same time depth were speakers of non-IE languages already residing in Europe.

There is no point in reviewing the spread of IE languages in detail here. Anthony 2007:123-457 reconstructs the early stages in detail on the basis of archaeological evidence; Mallory 1989:24-109 summarizes the later stages, bringing evidence of several kinds to bear. For those who are primarily interested in IE studies, Fortson 2004

(which is about to appear in a second, revised and expanded edition) is an excellent introduction, and specialist works on the subject constitute an entire library.

But if you're interested in what Europe was like six millennia ago, that's all irrelevant.

What it all means.

If you're not used to the kind of discussion that I've posted, I'd like to make sure you take away some very general messages; that's far more important than any of the details.

In the first place, if you want "reality-based" answers, take a scientific approach. Science may or may not reveal the existence of an objectively real world out there, but it does give results that can be replicated and answers that can be *proved* by anyone who knows how the system works. That's good enough for me because I think it's the best we can hope for. If people find specific scientific conclusions ideologically inconvenient, that's their problem.

Secondly, keep constantly in mind that all human society and all human language are single phenomena with multiple instantiations that are only superficially different. In the present context, this means that there is nothing special about Europe, and there is nothing special about IE either. I've adduced parallels for aboriginal Europe from elsewhere in the world because they are genuinely comparable. I attribute the spread of IE languages not to any innate superiority of the languages or their speakers, but to the fact that they had more cattle, better horses, and probably better weapons. (It's the same with the European colonial expansion of the 15th and 16th centuries. Forget the supposed superiority of European ideas; Europeans had better ships and better artillery, and some of the people they encountered *didn't* have immunity to Eurasian diseases. No other explanation for European success is necessary.)

I find it hard to see what relevance anything much earlier than the Roman Empire can have for modern Europe; but if you're a European and you see things differently, maybe you should think about the following. Unless you speak Basque, your native language was brought to where you live by immigrants — and unless you speak Greek or Irish Gaelic or Welsh, or are a native of one of a few selected provinces of Italy (such as Tuscany or Lazio), they weren't the first known immigrants, either. Your ancestry is almost certainly mixed, possibly as mixed as mine. (I have known ancestors from Ireland, Spain, France, the Kingdom of Hannover, the Rhineland, southern Germany, the Italian Alps, Croatia, and Serbia. God only knows what mixture lies behind each of

those lines of ancestry) You are the product of diversity because Europe has always been diverse.

Bibliography.

Anthony, David. 1991. "The archaeology of Indo-European origins." *Journal of Indo-European Studies* 19.193-222.

—. 2007. *The horse, the wheel, and language*. Princeton: Princeton U. Press.

Bhatia, Tej, and William Ritchie. 1999. "The bilingual child: some issues and perspectives." Ritchie and Bhatia (edd.), *Handbook of child language acquisition* (San Diego: Academic Press) 569-643.

Bonfante, Giuliano, and Larissa Bonfante. 2002. *The Etruscan language: an introduction*. Revised ed. Manchester: Manchester U. Press.

Buck, Carl. 1955. *The Greek dialects*. Chicago: U. of Chicago Press.

Duval, Paul-Marie (ed.). 1985. *Recueil des inscriptions gauloises*. Paris: CNRS.

Eska, Joseph. 1989. *Towards an interpretation of the Hispano-Celtic inscription of Botorrita*. Innsbruck: IBS.

Fantini, Alvino. 1985. *Language acquisition of a bilingual child: a sociolinguistic perspective (to age ten)*. San Diego: College-Hill Press.

Foley, William. 1986. *The Papuan languages of New Guinea*. Cambridge: CUP.

Fortson, Benjamin. 2004. *Indo-European language and culture: an introduction*. Oxford: Blackwell.

Guarducci, Margherita. 1942. *Inscriptiones creticae. III. Tituli Cretae orientalis*. Rome: Libreria dello Stato.

Labov, William. 2001. *Principles of linguistic change. Vol. 2. Social factors*. Oxford: Blackwell.

Lejeune, Michel. 1971. *Lepontica*. Paris: Les Belles Lettres.

—. 1974. *Manuel de la langue vénète*. Heidelberg: Winter.

- Mallory, J. P. 1989. *In search of the Indo-Europeans*. London: Thames and Hudson.
- Marinetti, Anna. 1985. *Le iscrizioni sudpicene*. Florence: Olschki.
- Meisel, Jürgen. 1989. "Early differentiation of languages in bilingual children." Hyltenstam, Kenneth, and Loraine Obler (edd.), *Bilingualism across the lifespan* (Cambridge: CUP) 13-40.
- Nichols, Johanna. 1990. "Linguistic diversity and the first settlement of the New World." *Language* 66.475-521.
- Packard, David. 1974. *Minoan Linear A*. Berkeley: U. of California Press.
- Poultney, James. 1979. "The language of the northern Picene inscriptions." *Journal of Indo-European Studies* 7.49-64.
- Prodocimi, Aldo (ed.). 1978. *Popoli e civiltà dell' Italia antica. Vol. 6. Lingue e dialetti*. Rome: Biblioteca di Storia Patria.
- Ringe, Don. 2008. *From Proto-Indo-European to Proto-Germanic*. Revised ed. Oxford: OUP.
- Rix, Helmut. 1998. *Rätisch und Etruskisch*. Innsbruck: IBS.
- Ross, Malcolm. 1997. "Social networks and kinds of speech-community event." Blench, Roger, and Matthew Spriggs (edd.), *Archaeology and language I: theoretical and methodological orientations* (London: Routledge) 209-61.
- Untermann, Jürgen. 1980. *Trümmersprachen zwischen Grammatik und Geschichte*. Opladen: Westdeutscher Verlag.
- . 2001. *Die vorrömischen Sprachen der iberischen Halbinsel*. Wiesbaden: Westdeutscher Verlag.
- Vetter, Emil. 1953. *Handbuch der italischen Dialekte. I. Band*. Heidelberg: Winter.
- Wallace, Rex. 2007. *The Sabellic languages of ancient Italy*. Munich: Lincom Europa.
- Woodard, Roger (ed.). 2004. *The Cambridge encyclopedia of the world's ancient languages*. Cambridge: CUP.