

The One Language Nobody Has Ever Heard That Spawned Over 400 Others
<https://www.youtube.com/watch?v=Ml4nyhCBaQk>
Transcript: <https://dontveter.com/ec/pie.pdf>

Proto-Indo-European. No one alive has seen it written down. No one has ever heard it spoken. It's been dead for over 5,000 years.

But this language, this ghost, is the ancestor of English, Hindi, Spanish, Russian, Greek, Persian, and over 400 others.

Languages spoken today by more than 3 billion people, nearly half the world.

And we only know of its existence because one man sitting in a courthouse in Calcutta in 1786 noticed something he couldn't explain.

Welcome to the Airlearn Language Show.

What I love about this story is that it starts exactly the way learning a language starts with that feeling of recognizing something you can't quite place.

A word that sounds familiar in a language you've never studied.

A pattern that shouldn't be there but is. That feeling is real and it's pointing at something.

Airlearn is built around chasing that feeling, understanding why beneath the words, not just the words themselves.

Sir William Jones was a British judge, a scholar, a man who by most accounts was almost aggressively brilliant.

He reportedly spoke 13 languages fluently and had a working knowledge of 28 others.

The British East India Company had posted him to Calcutta to serve on the Supreme Court of Bengal and in his spare time he was doing what William Jones always did: learn another language.

This time Sanskrit, the ancient classical language of India, the language of the Vedas, the Hindu scripture, of an entire literary and philosophical tradition stretching back thousands of years.

And as he studied it, he started feeling something strange, a sense of recognition he couldn't account for.

Sanskrit, deva, god, Latin deus, god.

Sanskrit, nava, nine. Latin, novem, nine. Greek ennea Nine.

Sanskrit pitr father Latin pater father Greek pater father

Not similar, not coincidentally close, structurally, systematically, repeatedly the same.

On February 2nd 1786, Jones stood before the Asiatic Society in Calcutta and delivered a lecture that would change the study of language forever.

He said, and I'm paraphrasing, that Sanskrit bore a stronger affinity to Greek and Latin than could possibly have been produced by accident.

So strong, he argued, that no one examining them could fail to believe they had all descended from some common source which no longer existed.

He was right. And that common source was what we now call Proto-Indo-European.

Now think about what that claim actually means. Jones wasn't just saying these languages share some vocabulary.

He was saying there was an original language, a real language spoken by real people that split and spread and evolved over thousands of years into Sanskrit, into Latin, into Greek, into eventually English, Spanish, Russian, Hindi, Persian, and hundreds more.

A single ancestral language spoken by someone, somewhere at some point.

But there was a problem. Nobody had written it down. There were no manuscripts, no inscriptions, no records of any kind.

The language had died thousands of years before writing reached most of the people who spoke its descendants.

So how do you find a language that left no trace? You work backwards.

In the 19th century, a new generation of linguists, German scholars mostly, building on Jones's insight, realized something profound.

Languages don't change randomly. They change according to rules. Predictable, consistent, almost mathematical rules.

One of the most famous examples is called Grimm's law. Yes, the same brothers Grimm who collected fairy tales.

Jacob Grimm was also a serious linguist. Grimm noticed that sounds shifted in systematic ways between related languages.

The Latin P consistently became an F in Germanic languages.

So Latin pater became English father. Latin piscis fish becomes English fish. Latin pes foot becomes English foot.

Every time the rule is consistent. This meant something extraordinary.

If you know the rules of how sounds change and you have the descendants, you can reverse engineer the ancestor, you can run the clock backwards.

Linguists started collecting thousands of words from hundreds of related languages, mapping the sound correspondences and reconstructing what the original must have looked like.

Word by word, sound by sound. Like archaeologists shifting through rubble to reconstruct a building they've never seen.

The language they rebuilt, this ghost assembled from its own descendants, is Proto-Indo-European.

We mark reconstructed words with an asterisk because we can't fully be certain. But the reconstruction is so internally consistent, so well supported across so many languages that most linguists are confident we're close to what was actually spoken.

And here's where it gets extraordinary. Because what we reconstructed didn't just give us a language. It gave us a world.

The words that exist in Proto-Indo-European, the ones we can confidently reconstruct, tell us what these people knew, what they saw, and what they lived with.

And from that vocabulary, we can start to figure out where and when they lived.

They had a word for snow. They had a word for wolf. They had words for birch tree, for salmon, for otter, for beaver.

These are all cold climate northern hemisphere species.

This rules out the Middle East, rules out South Asia, rules out Africa.

Whoever spoke this language lived somewhere temperate and cold.

They had a word for horse. And crucially, they had a word for wheel. This matters enormously because the wheel was only invented around 3,500 BCE.

So Proto-Indo-European couldn't have been spoken before that date.

The wheel word gives us a hard floor no earlier than about 4,000 BCE.

They had words for cattle, for sheep, for pigs. They were pastoral people, herders.

They had words for yoke, for harnessing animals to vehicles. They were mobile. They moved with their hearts.

And they had words for a sky god, *dyeus*, the shining father of the heavens.

You can hear him still in the Latin *Jupiter* from *dyeus pater* sky father, in the Greek *Zeus* in the Sanskrit *dyaus*.

Three civilizations that had no recorded contact with each other worshiping variations of the same god with the same name inherited from the same source.

So where were they? The leading theory today supported by both linguistics and genetics is called the Steppe Hypothesis.

And it points to a specific place, the Pontic Caspian Steppe, a vast green land stretching north of the Black Sea and the Caspian Sea in what is now Ukraine, southern Russia, and Kazakhstan.

Around 4,000 B.C.E., a people we call the Yamnaya were living there. They were semi-nomadic pastoralists.

They herded cattle and sheep. They were among the first humans to domesticate the horse, not just for meat, but for riding.

And they had wheels. The earliest wheeled vehicles ever found come from their burial mounds.

They match. They match the vocabulary. They match the ecology. And genetics has confirmed what linguistics predicted.

When ancient DNA from Yamnaya burial sites was analyzed, it showed a massive expansion.

Between about 3000 and 2,000 BCE, Yamnaya descendants populations spread into Europe into the west and into central Asia and South Asia to the east.

The genetic signature of these steppe people shows up in ancient European populations in South Asian populations across an enormous geographic range.

A small group of people from a specific place at a specific moment in history and their language became the seed of half the world's languages.

Now think about what that expansion looked like in practice. These weren't conquerors in the traditional sense.

There was no single army, no single migration, no dramatic movement. It was something slower and stranger.

Groups of mobile pastoralists on horseback with wagons spreading gradually over centuries and millennia into territories with existing populations.

Some of those populations adopted the newcomers language. Some resisted. Languages merged, diverged, evolved.

One branch moved west into Europe and eventually became the ancestor of Latin and from Latin, Spanish, French, Portuguese, Italian, Romanian.

Another branch moved into what is now Iran and gave rise to old Persian and eventually modern Persian.

Another moved into South Asia and became Sanskrit, the ancestor of Hindi, Urdu, Punjabi, Bengali and dozens of others.

Another spread north and east into what became the Slavic world, Russian, Polish, Czech, Ukrainian.

Another became the ancestor of the Germanic languages, German, Dutch, Swedish, Norwegian, Danish, and the ancestor of English.

Every time a group split off and lost contact with the others, the shared language started drifting.

Sounds shifted. New words were borrowed from local populations. Old words changed meaning. Grammar simplified or complexified in different ways.

And over thousands of years, what started as one language became hundreds.

The part of the story I find the most remarkable is not the languages themselves. It's the detective work.

Consider what linguists actually did. They took hundreds of languages living and dead and found the signal hiding inside the noise.

They identified the systematic patterns underneath the surface variation.

They reconstructed words for concepts, snow, wheel, wolf, god that had never been written down, in a language spoken by people who left no cities, no monuments, no manuscripts. And from those words, they deduced a time, a place, an ecology, a culture.

It's as if someone handed you 500 photocopies of photocopies, each one degraded, distorted, slightly different, and he had to reconstruct the original documents that no one alive had ever seen. And they did it.

We know what Proto-Indo-European words sounded like. We know roughly when and where they were spoken. We know what the people who spoke them herded and worshiped and built, all from the patterns in their linguistic descendants.

Here's something to sit with. The next time you say the word father in English, think about the chain of transmission that word traveled to reach you.

It was *pitr* in Sanskrit, *pater* in Latin, *fater* in old high German. And before any of that, it was *ph₂ter* in Proto-Indo-European, spoken by someone in a grassland north of the Black Sea, probably on horseback 6,000 years ago.

A word for the most fundamental human relationship, passed from mouth to mouth to mouth across six millennia and an entire hemisphere.

You are speaking an echo of that person every time you use it.

Sir William Jones noticed that Sanskrit looked like Latin and Greek. He had no idea what he had opened.

What followed was 200 years of the greatest linguistic detective story in history. A reconstruction of a ghost from its fingerprints.

A portrait of a civilization assembled entirely from the words it left behind in languages it never knew would exist.

And the ghost spoke not in manuscripts or inscriptions, but in the mouths of three billion people who had no idea they are still carrying it.

If this kind of thing, how languages work, where they come from, what they tell us about history, makes you want to actually dive in and learn one of these 400 descendants, the Airlearn app teaches several of them.

It's contextual, practical, learning the way language actually works in real conversations. links in the description. See you in the next one.