

Research reconciles two dominant hypotheses of Indo-European language origin

October 25 2023, by Kim Schulte



Credit: Pixabay/CC0 Public Domain

The languages in the Indo-European family are spoken by almost half of the world's population. This group includes a huge number of languages, ranging from English and Spanish to Russian, Kurdish and Persian.

Ever since the discovery, over two centuries ago, that these languages belong to the same family, philologists have worked to reconstruct the first Indo-European [language](#) (known as Proto-Indo-European) and establish a "language family tree," where branches represent the evolution and separation of languages over time. This approach draws on phylogenetics—the study of how biological species evolve—which also provides the most appropriate model for describing and quantifying the historical relationships between languages.

Despite numerous studies, many questions still remain as to the origin of Indo-European: where was the original Indo-European language spoken in prehistoric times? How long ago did this language group emerge? How did it spread across Eurasia?

Anatolia or the Pontic Steppe?

There are two main, though apparently contradictory, established hypotheses. On one side we have the [Anatolian Hypothesis](#), which traces the origins of the Indo-European people to Anatolia, in modern day Turkey, during the Neolithic era. According to this hypothesis, created by British archaeologist [Colin Renfrew](#), Indo-European languages began to spread towards Europe around 9,000 years ago, alongside the expansion of agriculture.

On the other side we have the [Steppe Hypothesis](#), which places the origin of Indo-European languages further north, in the [Pontic Steppe](#). This theory states that Proto-Indo-European language emerged somewhere north of the Black Sea around 5,000 or 6,000 years ago. It is linked to [Kurgan culture](#), known for its distinctive burial mounds and horse breeding practices.

DNA comparison

In order to decide which of these two hypotheses is correct, [genetic studies](#) have been carried out to compare DNA found at prehistoric sites with that of modern humans. However, this type of research can only provide indirect clues as to the origins of Indo-European languages, since language, unlike, for example, blood type, is not inherited through genes.

A [new study](#) published in *Science* has approached the question from a different angle by using direct linguistic data to assess the timelines put forward by both hypotheses.

In this project, carried out by over 80 linguists under the direction of Paul Heggarty and Cormac Anderson from the [Max Planck Institute for Evolutionary Anthropology](#) in Leipzig, we applied a [new methodology](#) that allowed us to obtain more exact results.

More comprehensive sampling

The samples used in earlier phylogenetic studies were taken from a limited pool of languages. Moreover, some analyses had assumed that modern languages are derived directly from ancient written languages, when they actually come from oral variants that were spoken during the same period—Spanish, for example, did not come from the classical Latin found in Virgil's works, but from the popular or "vulgar" Latin which was spoken by ordinary people. These shortcomings and assumptions have distorted age estimates for Indo-European language family subgroups such as Germanic, Slavic or Romance.

The new study addresses these issues, eliminating inconsistencies and taking data from a wider range of sources (from 161 languages, to be exact), to provide a more balanced and complete sample set. This data then underwent a [Bayesian phylogenetic analysis](#), a statistical method for establishing the most probable relationships between languages and branches of the family tree.

The study showed, for example, that an Italo-Celtic language family cannot exist, since the Italic and Celtic languages separated several centuries before the separation of the Germanic and Celtic languages, which took place around 5,000 years ago.

An eight thousand year old language family

Regarding the question of the origin of Indo-European languages, calculations based on the new data show that they were first spoken approximately 8,000 years ago.

The results of this research do not line up neatly with either the Anatolian or the Kurgan hypotheses. Instead they suggest that the birthplace of Indo-European languages is somewhere in the south of the Caucasus region. From there, they then expanded in various directions: westward towards Greece and Albania; eastward towards India, and northward towards the Pontic Steppe.

Around three millennia later there was then a second wave of expansion from the Pontic Steppe towards Europe, which gave rise to the majority of the languages that are spoken today in Europe. This hybrid hypothesis, which marries up the two previously established theories, also aligns with the results of the [most recent studies](#) in the field of genetic anthropology.

In addition to bringing us closer to solving the centuries-old enigma of the origin of our languages, this research illustrates how disciplines as disparate as genetics and linguistics can complement each other to provide more reliable answers to questions of human prehistory. It is hoped that the same methodology will also serve, in future research, to expand our understanding of how languages and populations spread to other continents.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Research reconciles two dominant hypotheses of Indo-European language origin (2023, October 25) retrieved 27 April 2024 from <https://phys.org/news/2023-10-dominant-indo-european-language.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.