The Co-Evolution of Agriculture, Language and Human Genetic Mutations

Abstract: The effort to harness global Y-Chromosome variation as a tool for linguistic research has produced an interesting observation. Recent genetic evidence supports Bellwood’s 2005 early farming dispersal hypothesis. Thus, the standard approach to Indo-European origins, the so-called steppe hypothesis, may not be a persuasive model of language prehistory.

Keywords: Indo-European languages, the steppe hypothesis, the early farming dispersal hypothesis, the co-evolution hypothesis, Y-chromosome single nucleotide polymorphisms.

Introduction

Among linguists, the most accepted approach to the pre-history of Indo-European languages is the steppe hypothesis. In 2005, Peter Bellwood presented an alternative approach, the early farming dispersal hypothesis. This paper introduces a new concept related to Bellwood’s hypothesis, which is the co-evolution hypothesis. Utilizing recent genetic data published after 2005, co-evolution hypothesis strengthens the early farming dispersal hypothesis by asserting that language typology, the archaeological record and human genetic mutations converge to provide one explanation for global language variation. This approach to language prehistory offers a model of the prehistory for Indo-European language family that is strikingly similar to the prehistory of the Afro-Asiatic, Niger-Congo, Austro-Asiatic, Sino-Tibetan, Austronesian, Dravidian, Trans-New Guinea and Uto-Aztecan language families. Consequently, advocates of the steppe hypothesis must now overcome a major obstacle and explain why they provide an atypical explanation of language prehistory.

The Standard Approach to Indo-European Prehistory

The so-called steppe hypothesis of Indo-European origins stands as perhaps the received explanation among linguists who explore the geographic origins and expansion of the Indo-European language family. This approach finds its origins in the nineteenth century when linguists began searching for an ancestral culture that spread Indo-European languages throughout a vast geographic expanse that extends from Icealand to India. In the 1950s, Marija Gimbutas, a Lithuanian archaeologist, undertook this project. In a paper published in 1997, she presents the culmination of her work in this area. Gimbutas takes the position that the Kurgan people, a prehistoric nomadic culture of the Pontic Steppes north of the Caspian and Black Sea, invaded Europe during the Bronze Age, about 4,000 years ago. Following the invasion, according to Gimbutas, the Kurgans imposed their Indo-European language on the indigenous agrarian inhabitants of the continent. Her conclusions are partly archaeological, but are mainly supported by phonological reconstructions. Later, the anthropologist David Anthony substantially reworked Gimbutas’ Kurgan theory and has arguably become the current Kurgan crusader. Like Gimbutas, he still adheres to the idea that nomads from the Pontic Steppes disseminated Indo-European languages throughout the European continent. Furthermore, like Gimbutas, he uses phonological reconstruction as the primary support for his ideas. Anthony recently co-published a paper with Don Ringe, a linguist, that further explores and endorses the Kurgan hypothesis based on linguistic reconstructions. Moreover, he recently published a paper with several geneticists who endorse the Kurgan hypothesis based on a small dataset of ancient DNA samples.
An Alternative Approach to European Prehistory.

The language-farming hypothesis, or the alternate explanation for language variation in Europe, was initially presented by Colin Renfrew in his 1987 book *Archaeology and Language: the Puzzle of Indo-European Origins*. He identifies the Middle East and Fertile Crescent as the putative homeland of this language family. According to this approach, farming and language co-expanded from Southwest Asia to Europe during the Neolithic. Rather than a cultural expansion, this hypothesis suggests that the expansion of Indo-European languages represents a demographic process. The basic idea is that agricultural supports far more people per square kilometer than the alternate form of food production that it replaces, which involves hunting and gathering techniques. Consequently, the Neolithic transformation in Europe, which began about 8,000 years ago, produced a population explosion, which caused early agriculturalist to migrate throughout the European continent in search of new farmland.

In his 2005 book, the archaeologist Peter Bellwood expanded the language farming hypothesis to examine language variation from a global perspective, which he labels the early farming dispersal hypothesis. According to the hypothesis, shortly after the initial domestication of crops and animals by human cultures, which occurred independently in several regions of the world, population pressure forced early agriculturalists to migrate. This population expansion, in turn, dispersed several different language families, such as Indo-European, Afro-Asiatic, Niger-Congo, Sino-Tibetan, Austronesian, and Uto-Aztecan. To support his hypothesis, Bellwood relies heavily on the archaeological record and paleobotanical evidence.

The Co-Evolution Approach to Language Prehistory

In 2013, Michael St. Clair, a linguist, began examining global language variation from the perspective of the non-recombining region Y-chromosome, a form of genetic data used extensively in the field of anthropology. The potential correlation between genetic and linguistic variation stems from their shared means of transmission from one generation to the next. In other words, we inherit the mother tongue and our genes from our parents. In 2016, St. Clair prepared a preliminary report, *Harnessing Global Y-Chromosome Variation as a Tool for Linguistic Research*, which is posted on the Genetic-Linguistic Interface webpage. Among the findings presented is that from a Y-chromosome perspective, climate change, agricultural expansion, and cultural isolation have shaped global language variation. This observation generally conforms to a demographic model whereby human genetic variation was initially shaped by founder effect, bottlenecks, and drift among hunter-gatherer societies. Then, agriculture re-shaped the genetic landscape through population expansions, migration, and admixture between farmers and hunter-gatherers.

One important byproduct of St. Clair’s 2016 paper is that he presents genetic data which provide independent corroboration of Bellwood’s 2005 hypothesis. Taken as a whole, the archaeological record, as presented by Bellwood, and the genetic data presented by St. Clair, strongly suggest the following: Much of the world’s linguistic variation was shaped by the co-evolution of agriculture, language and human genetic mutations. In other words, language typology, the archaeological record and human genetic mutations converge to provide one explanation for global language variation. This pattern is found in East Asia, where the cultivation of rice and foxtail millet co-evolved with the Sino-Tibetan language family and haplogroup O3a1c-002611 mutation. In New Guinea the cultivation of tubers and fruits co-evolved with the Trans New Guinea language family and the haplogroups M and S mutations.
In Oceania, the cultivation of rice and tubers co-evolved with Austronesian languages and the haplogroup O-P164 mutation. In Sub-Saharan Africa, the cultivation of pearl millet and sorghum co-evolved with Niger-Congo languages and the haplogroup E-M2 mutation. In Southwest Asia, Europe and India, the cultivation of emmer, einkorn and legumes, as well as the domestication of sheep and goats, co-evolved with Indo-European languages and the haplogroup J2 mutation. In Southwest Asia and North Africa, the domestication of sheep, goats and cattle co-evolved with Afro-Asiatic languages and the haplogroups J1-M267 and E-M81, and R-V88 mutations. In India, Dravidian co-evolved with the cultivation of millet and haplogroup the T-M184 mutation. In the same region, the cultivation of rice co-evolved with Austro-Asiatic languages and the haplogroup the O-M95 mutation. In the Southwestern United States and Mexico, the cultivation of maize co-evolved with the Uto-Aztecan language family and the haplogroup Q-L54(xM3) mutation.6

Conclusions

As noted earlier, essentially two different approaches are utilized in an effort to explain the origins and the initial dispersal of Indo-European languages, the steppe hypothesis and Bellwood’s early farming dispersal hypothesis. This paper strives to introduce a new concept into the debate, which is the co-evolution hypothesis. In doing so, we find that the prehistory of Indo-European languages is strikingly similar to that of Afro-Asiatic, Niger-Congo, Austro-Asiatic, Sino-Tibetan, Austronesian, Dravidian, Trans-New Guinean and Uto-Aztecan language families. Accordingly, it seems as though the steppe hypothesis has a serious deficiency in that it makes Indo-European an exception to the rule. Perhaps those who investigate language prehistory should examine linguistic, archaeological and genetic data across the full range of global human diversity and make their conclusions based on where the data converge. The steppe hypothesis, and its heavy reliance on a single data source, linguistic reconstructions within a single language family, represents a potentially flawed methodological approach, and as such, does not present a persuasive model of language prehistory.

Endnotes

1 For a more detailed discussion of early efforts to find the Indo-Europeans, the following sources are recommended:


4 Haak, Wolfgang et al. 2015. “Massive migration from the steppe was a source for Indo-European languages in Europe.” Nature 522: 207-211.


7 Chapter 2 of the report describes the Y-chromosome as a research methodology. Chapter 5 of the report presents an analysis of the Y-chromosome data. From a Y-chromosome data perspective, isolation, climate change and agriculture have shaped global language variation.


8 Language classification conforms to the following standard:


The agriculture data are derived from the following sources:


The genetic data are derived from the following sources:


