

# The Historical Linguistic Record and the Limits of Language Reconstruction

*by Jessi Warthen*

## Introduction

The question of how human language originated and diversified has occupied linguists, anthropologists, theologians, and historians for centuries. The dominant secular framework positions language as an emergent product of biological and cultural evolution, diversifying gradually over tens of thousands of years through geographic isolation, population migration, and social drift. The biblical framework, by contrast, attributes the origin of human linguistic diversity to a single discrete event: the confusion of languages at Babel, recorded in Genesis 11, following the genealogical record of nations in Genesis 10.

This paper does not argue from sentiment or tradition. It argues from documented data alone, holding both the biblical framework and the secular alternative to the same evidentiary standard. The central contention is that when the actual count of documented language families and isolates is measured against a population model derived from the Genesis genealogies, the numbers are not only compatible with a Babel origin for human linguistic diversity but are more consistent with that framework than with the assumptions underlying secular linguistic prehistory. Furthermore, the existence of language isolates, particularly Basque and Sumerian, and the documented universality of flood accounts across isolated populations, constitute independent lines of evidence that intersect with and support the biblical historical framework.

The analysis proceeds in three stages. First, the documented linguistic data is examined on its own terms, stripping out extrapolated estimates and working only with confirmed counts. Second, a population model based on post-flood genealogical data and documented fertility rates is applied to establish the number of family heads present at Babel under a young-earth chronology. Third, the resulting ratio of original language groups to currently documented languages is evaluated for internal consistency and compared against what the secular framework requires but cannot demonstrate.

---

## Part One: The Documented Linguistic Record

### 1.1 How Many Languages Actually Exist

The first discipline required in this analysis is the removal of estimated figures treated as factual data. Secular linguistics frequently cites a figure of approximately 20,000 languages as having existed prior to the Neolithic or Agrarian Revolution, a transition period whose dating remains unsettled across a range of approximately 10,000 to 3,500 BC depending on geographic region and methodological approach. This figure is not documented. It is a backward projection derived from current extinction rates applied to an assumed timeline. No such languages have been

recorded, catalogued, or verified. The number exists within a secular framework as a modeling assumption, not as an empirical finding.

The two primary reference databases for documented human languages are Ethnologue, published by SIL International, and Glottolog, maintained by the Max Planck Institute for Evolutionary Anthropology. These databases differ in scope and methodology in ways that are directly relevant to this analysis. Ethnologue's 2025 edition catalogues 7,159 living languages, meaning languages with documented active speaker communities, however small. Extinct languages are tracked separately by Ethnologue through the ISO 639-3 standard, with at least 633 extinct languages catalogued under that system, though the database itself acknowledges this figure substantially undercounts the total number of extinct languages for which documentation exists. Glottolog takes a broader scope, cataloguing 8,612 documented human languages in total across living, endangered, and extinct categories, including languages like Selk'nam and Sumerian for which no living native speakers remain but for which sufficient recorded data exists for classification and study.

The variance between these figures results from differing methodological standards regarding the boundary between a language and a dialect, the treatment of extinct languages with limited documentation, and the inclusion or exclusion of categories such as sign languages, pidgins, and creoles. Neither database claims finality. For the purposes of this analysis, Glottolog's total documented figure of 8,612 is the more appropriate reference because the argument concerns the full scope of linguistic diversity originating at Babel, not merely languages with living speakers. An even working figure of approximately 8,000 to 8,600 documented languages captures the full range of both databases without overstating either source, and this paper uses 8,000 as a conservative baseline for subsequent calculations.

## 1.2 Language Families and Isolates

A language family is a grouping of languages demonstrably descended from a common ancestor through the comparative method, requiring regular sound correspondences, shared morphological patterns, and systematic lexical relationships across the proposed member languages. A language isolate is a language for which no such demonstrable relationship to any other language or family can be established. It stands alone.

Campbell's count of 406 independent language families includes 129 confirmed language isolates. A separate count places the figure at 159 when extinct isolates with sufficient documentation are included. These isolates constitute approximately 32 to 39 percent of all language families, meaning that over one third of the world's linguistic diversity consists of languages with no proven relatives anywhere on the language tree.

This is a remarkable statistic that the secular framework handles poorly. If languages diversified gradually through population migration and cultural drift from a common ancestral tongue over hundreds of thousands of years, the expectation would be a relatively continuous tree structure in which most languages cluster into demonstrable families. Instead, over a third of the world's linguistic families stand in complete isolation from one another. The secular explanation is that these isolates are the last surviving members of once-larger families whose relatives have gone

extinct without documentation. This is possible. It is also untestable. It is an inference dressed as an explanation.

The biblical framework handles isolates differently and more parsimoniously. If God created hundreds of distinct languages at Babel simultaneously, each constituting its own original linguistic system rather than a modification of a prior common tongue, then language isolates are exactly what the model predicts. Languages whose sibling families died out entirely would leave a surviving isolate. Languages whose descendants never diversified enough to produce recognizable daughter families would also register as isolates. The presence of 129 to 159 confirmed isolates among 406 total families is consistent with an origin event producing many discrete, unrelated language systems at a single point in time.

### 1.3 The Chonan and Chuvash Case Study

The Chonan language family of Tierra del Fuego and Patagonia includes Selk'nam, also known as Ona, and the closely related Tehuelche and Haush languages. Selk'nam is classified as extinct, with the last fully fluent native speakers having died in the 1980s. It used an object-verb-subject word order, found in only approximately one percent of the world's languages, and employed agglutinative morphology, building complex meaning through the stacking of affixes onto roots. The family name derives from the Selk'nam word for "man."

Chuvash is a Turkic language spoken in the Chuvash Republic of Russia, currently classified as vulnerable with approximately one million speakers. It is the sole surviving member of the Oghuric or Bulgaric branch of the Turkic family, which diverged from Proto-Turkic approximately 2,000 to 2,500 years ago. It is not mutually intelligible with any other Turkic language, including Tatar and Bashkir, and has been shaped by extensive contact with Uralic, Mongolic, and Slavic languages over centuries. Its roots trace to the Volga Bulgars, a medieval steppe civilization.

A surface-level comparison of Chonan and Chuvash finds no demonstrable connection. They belong to entirely different branches of what would be, under the biblical model, the post-Babel dispersion. Chonan peoples descended from migration waves originating in the Altai-Siberian corridor into the Americas, representing some of the earliest and most isolated movements out of the Babel dispersion. By the time Chuvash emerged as a distinct branch from its Oghuric ancestor, Selk'nam ancestors had been isolated in the southern cone of South America for an unknown but substantial period. The two languages are useful precisely because the absence of connection between them illustrates the scope and completeness of post-Babel dispersal. They are on opposite ends of the earth because the dispersion from Babel was complete and total, not gradual or limited.

---

## Part Two: Population at Babel and the Heads of Families Problem

### 2.1 The Table of Nations

Genesis 10 records seventy named descendants of Noah's three sons: Shem, Ham, and Japheth. Twenty-six are attributed to Shem, thirty to Ham, and fourteen to Japheth. This record is commonly referred to as the Table of Nations, and it has historically been interpreted in

creation research as a near-complete inventory of language groups originating at Babel. The Institute for Creation Research has cited it in this way, noting that the number of major language families is "of the same order of magnitude" as the seventy recorded names.

The Table of Nations is not presented in the text as a census of all persons or language groups at Babel. It is a divinely revealed genealogical and covenant record, documenting the named patriarchal lines through which God established the origins of nations and the framework of human history from creation forward. The text itself acknowledges further spread beyond those named through phrases indicating that additional peoples descended from these lineages. It is covenant and genealogical history written with eternal purpose, not a linguistic inventory, and the quantitative limits of its seventy named lines do not constrain what God accomplished at Babel. They reflect what He chose to reveal through Moses for the specific purposes that record serves across all of human history.

## 2.2 Population Modeling from the Genealogies

The question of how many people were actually present at Babel is a demographic one, and it can be addressed with reasonable precision using the genealogical data of Genesis and a documented fertility parameter.

Dr. Nathaniel Jeanson, a geneticist affiliated with Answers in Genesis whose Y-chromosome research has been published in peer-reviewed creation science journals, has estimated the population at the time of the Babel dispersion at approximately 18,000 people. This figure derives from applying a post-flood fertility rate and exponential growth model beginning from three reproductively viable founding couples: the sons of Noah, Shem, Ham, and Japheth, and their wives. Noah and his wife are excluded from the reproductive baseline as the text does not document children born to them after the flood. Jeanson places the Babel event at approximately 150 years post-flood, consistent with the Masoretic genealogical timeline placing Peleg's birth within the first several generations after the flood, with Peleg's name meaning "division" interpreted as a reference to the dispersion.

The fertility rate of 7.5 children per woman is not drawn from secular demographic assumptions about ancient agrarian societies. It is derived directly from the Table of Nations itself. Jeanson calculated this figure by averaging the number of children documented per patriarch across the Genesis 10 genealogical record, producing a fertility parameter grounded entirely in the biblical text rather than in external demographic modeling. Beginning from three founding couples and applying a generational interval of twenty to twenty-five years consistent with the marriage patterns documented in the Genesis genealogies, the following population trajectory emerges at a 7.5 fertility rate.

At generation one, approximately twenty years post-flood, the total population reaches approximately thirty individuals with roughly fourteen cumulative male heads of family. At generation two, roughly forty to fifty years post-flood, the total reaches approximately 115, with cumulative male heads of family approaching fifty-six. At generation three, approximately sixty to seventy-five years post-flood, the total population reaches approximately 431 with 215 cumulative male heads of family. At generation four, approximately eighty to one hundred years post-flood, the total population reaches approximately 1,618 with over 800 cumulative male heads of family. At generation six, corresponding to approximately 120 to 150 years post-flood,

the total population reaches approximately 18,000 with over 3,000 cumulative male heads of family.

The range of proposed timelines for the Babel event produces meaningfully different head of family counts and both are worth noting. At 100 years post-flood, which this paper takes as the more accurate chronological position based on the genealogical data, the model produces approximately 800 cumulative male heads of family. At 150 years post-flood using Jeanson's timeline, the figure rises to over 3,000. Under either timeline the number of male heads of family present at Babel substantially exceeds seventy by a considerable margin.

The seventy names in Genesis 10 therefore represent a documented fraction of the actual family heads present at Babel, recorded through divine revelation to Moses as the covenant-relevant patriarchal lines through which God established the traceable origins of nations. This is not a problem for the biblical account. It is exactly what a divinely structured genealogical record written for eternal covenant purposes would look like.

### 2.3 The Ratio of Original Languages to Current Languages

If the Babel event at approximately 100 years post-flood produced approximately 800 distinct language groups corresponding to the heads of families present, and if approximately 8,000 languages are documented today, the ratio is approximately 10 documented languages per original Babel language group.

This is a defensible and internally consistent figure. The Indo-European language family alone, descending from a single reconstructed ancestor language, produced hundreds of daughter languages over a comparable timeframe. A rate of ten documented descendants per original Babel language over roughly 4,000 years is conservative by comparison. It requires no extraordinary linguistic multiplication after the event. God created approximately 800 distinct language systems at Babel. Natural divergence through geographic isolation, population splitting, and cultural differentiation over the subsequent millennia produced the documented 8,000.

This framework is transparent about its assumptions. The 800 family heads derive from a fertility parameter drawn directly from the Table of Nations and applied to the genealogical record of Genesis. The 8,000 language baseline represents a working average between Ethnologue's catalogue of 7,159 living languages and Glottolog's total documented count of 8,612 languages spanning living, endangered, and extinct varieties. That averaged baseline is itself a modeling choice, not a precise count, and is acknowledged as such. The resulting ratio of ten documented languages per original Babel family is a model with stated assumptions whose output is consistent with observable rates of linguistic diversification in known language families. The secular alternative requires 20,000 or more prehistoric languages for which no documentation exists, built entirely on extrapolation from assumed evolutionary timelines. On this specific question, documented data supports the biblical model more directly than the secular one.

### 3.1 Basque

The Basque language, known to its speakers as Euskara, is spoken in the Basque Country spanning the western Pyrenees across northern Spain and southwestern France. It is the oldest documented language in Europe and is unrelated to any other language on earth, including all members of the Indo-European family that surrounds it. The only demonstrated genetic connection in the scholarly literature is to Aquitanian, an ancient language known from Roman-era inscriptions dating between the first and third centuries AD, which is itself merely an older form of Basque. The connection to Aquitanian does not extend Basque into a broader family. It pushes the same isolate further back in time.

The Hand of Irulegi, a bronze artifact discovered in 2021 near Navarre and dated to the first century BC, bears the oldest known Basque inscription, including the word "sorioneku" meaning "fortunate" or "blessed," directly cognate with the modern Basque "zorioneko." Archaeological evidence indicates continuous human occupation of the Basque region extending to the Aurignacian period, assigned under secular dating conventions to approximately 40,000 years ago, a figure this paper does not adopt as it rests on radiometric dating methodologies whose foundational assumptions are contested within creation science literature. What the archaeological record does confirm, independent of the dating dispute, is demonstrable cultural continuity in the region across multiple occupation layers extending well into the pre-Roman period. The Basque people demonstrate the highest concentration of Rh-negative blood type of any population on earth, a genetic anomaly with no fully satisfying secular explanation, and their own oral traditions and legends include accounts of descent from a sunken or submerged civilization.

Every serious attempt to connect Basque to a known language family has failed. Proposed connections have been advanced to ancient Iberian, ancient Egyptian, Celtic, Latin, Sanskrit, Pictish, Berber, Sudanic, Semitic, Etruscan, Minoan, Sumerian, the Munda languages of India, Uralic, Dravidian, Sino-Tibetan, Eskimo, and the Na-Dene languages of North America. The Caucasian languages have been considered the strongest candidates among living languages. None of these proposals has met the evidentiary standard of the comparative method, requiring regular and systematic sound correspondences across a broad range of basic vocabulary and shared irregular morphology. The conclusion of mainstream linguistics is that Basque cannot be demonstrated to be related to any other known language, living or dead.

The Vasconic Substrate Theory, proposed by the German linguist Theo Vennemann, argues that Basque is the last surviving member of a large family of related languages spoken across Western Europe before the arrival of the Indo-Europeans, conventionally dated to around 3,000 BC under secular linguistic chronology. Under this theory, Basque is not simply an old language. It is the sole living remnant of an entire pre-Indo-European linguistic world that was displaced and absorbed by Indo-European expansion. This proposal has not achieved consensus but has not been definitively refuted.

Under the framework developed in this paper, Basque occupies a significant position. It is a language whose sibling families have entirely disappeared, leaving no demonstrable relatives anywhere in the documented record. It has preserved its isolation through geography, cultural cohesion, and resistance to conquest, surviving Roman, Visigothic, Frankish, and later

pressures that absorbed every surrounding population. The extraordinary methodological caution required before identifying any language as an original Babel family is acknowledged. The absence of a connection to any other language family does not constitute proof of original status. It is, however, consistent with original status and inconsistent with the secular narrative of gradual divergence from a common ancestral tongue, which would predict traceable connections even over long timescales.

### 3.2 Sumerian

Sumerian is the oldest language for which a substantial written record exists. Administrative records from the city of Uruk represent the first sustained writing system in human history, with the earliest texts dated by secular chronology to approximately 3,200 BC. Sumerian was spoken in southern Mesopotamia and was the dominant language of the region until displaced by Akkadian as a spoken language, after which it continued as a literary and liturgical language for centuries.

Sumerian is classified as a language isolate. No demonstrated genetic relationship to any other language family has been established despite extensive comparison. Proposed connections have included the Dravidian languages of South Asia, the Munda languages, Uralic languages, and Basque itself. The Basque-Sumerian connection has attracted particular attention from researchers, with one scholar cataloguing over eighty apparent vocabulary similarities between the two languages. However, the structural problem is significant. Sumerian is characterized by a high proportion of monosyllabic roots, while Basque has very few monosyllabic roots. Root structure is one of the deepest markers of linguistic identity. Vocabulary overlap, particularly in small samples, is readily explained by chance, borrowing, or the operation of universal tendencies in how human languages name basic concepts. Vocabulary overlap in the absence of systematic structural correspondence does not constitute a genetic connection under the comparative method.

Sumerian and Basque are therefore both confirmed language isolates with no proven connection to each other or to any other language family. Under the biblical model this is expected. Two language groups originating at Babel whose sibling languages died out entirely would each appear as an isolate in the documented record, and there is no reason to expect them to be connected to each other. Two distinct Babel language families whose siblings all went extinct would produce exactly this pattern.

### 3.3 The Universal Flood Account as Corroborating Evidence

A body of comparative mythology research, most prominently represented by Harvard linguist Michael Witzel in his paper "Pan-Gaeon Flood Myths: Gondwana Myths and Beyond," published in 2010 in the proceedings of the International Association for Comparative Mythology, documents the presence of flood narratives across the full geographic range of human population, including populations with no documented contact with one another or with Near Eastern literary traditions.

Witzel classifies world mythologies into two broad categories: Gondwana, representing the older Out-of-Africa populations and their geographic descendants, and Laurasian, representing the

northern Eurasian and American populations. Flood myths appear in both categories with consistent core elements: a universal inundation, divine causation as retribution for human transgression, a small number of survivors, escape by boat or elevation, and a subsequent covenant or warning. The Andaman Island account, associated with the Jarawa and Onge peoples whose mitochondrial DNA places them in the Asian haplogroup M rather than the African haplogroup L, includes a creator figure named Puluga who sends a flood in response to human disobedience, from which only two men and two women in a canoe survive, followed by a direct covenant warning that such punishment would follow further transgression. Witzel himself describes the biblical echoes in this account as striking.

The Selk'nam of Tierra del Fuego, the Inuit of the Arctic, and the Jarawa of the Andaman Islands have been identified as among the most geographically isolated populations in the world, with no documented contact with each other or with the ancient Near East. All three carry flood accounts with structural similarities to the Genesis narrative that are closer than those found in cultures with actual historical contact with the biblical text. The secular response to this pattern is cultural diffusion, the idea that flood stories spread through ancient human population movements. This explanation collapses when applied to populations whose isolation is itself the premise of the secular framework. A population isolated for tens of thousands of years cannot have received a flood narrative through cultural diffusion from the Near East during that isolation. The secular framework is self-defeating on this point.

The biblical framework requires no diffusion mechanism. If a global flood occurred, all surviving human populations would carry a memory of it. The degree of structural similarity in accounts from isolated populations is consistent with common descent from survivors of a single event rather than independent invention of a similar narrative. Witzel, writing from a secular comparative mythology framework, documents the universality of the pattern while being unable to account for it satisfactorily within his own framework. His data constitutes a significant body of hostile-witness evidence in support of the biblical flood account. The parallel to the linguistic evidence examined throughout this paper is direct. The same geographic isolation that defeats the secular diffusion argument for flood myths is precisely what produces language isolates. Populations that carried no borrowable flood narrative from the Near East equally carried no borrowable linguistic relatives from surrounding families. Both lines of evidence converge on a single dispersal event at Babel as the most parsimonious explanation for what the documented record shows.

---

## **Conclusion**

The central contention of this paper has been that the documented linguistic record, when evaluated strictly on confirmed data rather than extrapolated assumptions, is more consistent with the biblical account of Babel than with the secular framework of gradual linguistic evolution. That contention holds across each of the three lines of evidence examined.

The quantitative data presents the secular framework with a problem it cannot resolve through its own methodology. A model of gradual divergence from a common ancestral tongue predicts a relatively continuous family tree structure across the world's languages. The actual documented record produces confirmed isolates approaching or exceeding one third of all documented

language families, with no demonstrable genetic relationship to any other language anywhere in the world. That proportion is not what gradual divergence predicts. It is precisely what a discrete origin event producing hundreds of simultaneous, unrelated language systems would produce.

The population model derived from the Genesis genealogies, applying a fertility parameter drawn directly from the Table of Nations and beginning from the three founding couples of Shem, Ham, and Japheth, produces a ratio of documented languages per original Babel family that is consistent with observable rates of linguistic diversification in known language families and requires no undocumented mechanism to sustain it. The seventy names of Genesis 10 represent a divinely revealed covenant record, not a linguistic census, and the quantitative data is fully compatible with that distinction.

The convergence of language isolate evidence with the universal flood account documentation examined in Part Three strengthens the overall framework considerably. Basque and Sumerian stand without demonstrable connection to any known language family despite exhaustive comparison against every plausible candidate across centuries of scholarly effort. The Selk'nam, the Jarawa, and the Inuit carry flood accounts structurally proximate to the Genesis narrative despite documented isolation from one another and from the ancient Near East. The secular framework must invoke separate explanatory mechanisms for each of these phenomena independently and succeeds with none of them. The biblical framework accounts for all of them through a single dispersal event at Babel from which populations carried both their distinct language systems and their memory of the antecedent flood into geographic isolation simultaneously.

No claim is made here that this constitutes proof in the formal sense. What is claimed is that the biblical model, when admitted as a live hypothesis and tested against documented data with stated assumptions and transparent methodology, performs demonstrably better on this specific body of evidence than a secular alternative that requires undocumented prehistoric language counts, unsettled chronological frameworks, and diffusion mechanisms that collapse precisely where the evidence is most critical. The burden of extraordinary evidence does not rest only on the framework that challenges the consensus. It rests equally on the framework that substitutes assumptions for documentation and calls the result science.

## References

Campbell, L. (Ed.). (2018). *Language isolates*. Routledge.

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2025). *Ethnologue: Languages of the world* (28th ed.). SIL International. <https://www.ethnologue.com>

Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (2025). *Glottolog 5.3*. Max Planck Institute for Evolutionary Anthropology. <https://glottolog.org>

International Organization for Standardization. (2007). *ISO 639-3:2007. Codes for the representation of names of languages: Part 3, Alpha-3 code for comprehensive coverage of languages*. ISO.

Jeanson, N. T. (2022). *Traced: Human DNA's big surprise*. Master Books.

Vennemann, T. (2003). *Europa Vasconica – Europa Semitica* (P. Noel Aziz Hanna, Ed.). Mouton de Gruyter.

Witzel, M. E. J. (2010). Pan-Gaeon flood myths: Gondwana myths and beyond. In W. M. J. van Binsbergen & E. Venbrux (Eds.), *New perspectives on myth: Proceedings of the Second Annual Conference of the International Association for Comparative Mythology* (pp. 225–242). Shikanda.