THE HISTORY OF ENGLISH PODCAST TRANSCRIPTS

EPISODE 3: THE INDO-EUROPEAN FAMILY TREE

Presented by Kevin W. Stroud

EPISODE 3: THE INDO-EUROPEAN FAMILY TREE

Welcome to the History of English podcast – a podcast about the history of the English language and the people who contributed to that history. In the last episode, we looked at how a British judge in India helped to discover the oldest known ancestor of English – the ancient Indo-European language. In this episode we will look more closely at the Indo-European family of languages and how English fits into that family.

Before we look at the Indo-European language family in detail, let me emphasize the importance of beginning a history of English with this ancient language – the language which eventually led to English and most of the other European languages. You might be surprised how similar many of the words in the original Indo-European language were to the words we use in Modern English.

Now no one knows for certain how the original Indo-European words were pronounced, but some of the words which have been reconstructed in this language appear to be very similar to their Modern English equivalents:

oxen – *uks-en
mother – *mater
one – *oinos
six – *sweks
seven – *septm
bear – *behr
apple – *abel

But this is about more than just some similar words. A large portion of the base vocabulary of English came from this source. But just as importantly, it is the parent language of not only English, but also Latin, Greek, the Celtic languages, and all of the Germanic languages including the Scandinavian languages. In other words, all of the languages which have melded together to form modern English derived from the same source language. So as we move forward in our look at the history of English, it helps to see how interconnected these various influences are. Ultimately, all of these languages are cousins and they share a substantial amount of vocabulary.

About 50% of the world's population speaks an Indo-European language as their native language. That's about 3 billion native speakers of Indo-European languages.

So what about English? It is an Indo-European language, but it is a hybrid language that pulls words and other influences from a variety of Indo-European languages. As I have said, English has at its core the original Germanic language known as Anglo-Saxon or Old English. But it has lost a significant portion of the original Anglo-Saxon vocabulary and replaced it with borrowed words. Most of these sources of borrowed words are also Indo-European languages. So Indo-European roots find their way into English in many different (and sometimes redundant) ways. For example, you may recall in the last podcast we talked about words like ***father*** and ***foot*** and how their Latin equivalents – ***pater*** and ***ped*** – not only also found their way into English meaning

essentially the same thing, but they also ultimately came from the same original source word as the English words did. ***Father*** and ***pater*** were once the same word. And ***foot*** and ***ped*** were once the same word. It has been estimated that almost 50% of the entire reconstructed vocabulary of the original Indo-Europeans is represented in some form in modern English. In other words, almost half of the known words of the original Indo-European language, spoken thousands of years ago, can be found in English today. They come in either as a direct inheritance from the Germanic languages or as words borrowed from one of the other Indo-European languages like Latin or Greek, or sometimes from both as in the examples of ***father*** and ***foot***.

In the last podcast, I talked about Sir William Jones, the British judge in India who is credited with discovering the connections between many European languages and the Persian and Sanskrit languages of Central Asia. Jones basically recognized that these languages were related. In other words they were part of the same family of languages. And he ventured some reasonable guesses about how these various languages were related, but subsequent research has given us a much more complete understanding of how the languages fit together. We can now look at the whole family tree of Indo-European languages and see which branch English belongs to – which by the way is the Germanic branch. You can check out the family tree diagram at the website for this podcast – historyofenglishpodcast.com. Just click on the link for Episode 3.

So at this point I want to describe the family tree and introduce you to the Indo-European languages. This will serve as a helpful primer as we move forward. And in the next few podcasts we will be looking at the specific languages within this family which directly impacted the development of English.

Let me begin by noting that the family tree on the podcast website features 12 branches of the Indo-European language family. It should be noted that other illustrations or diagrams use 10 or 11 branches. There a couple of reasons for this variation. First, some linguists combine some of the branches into a single branch initially and then separate them into separate branches later, while other linguists prefer to represent each of these branches separately from the beginning.

For example, the languages of Eastern Europe are represented by two separate branches of the family tree. The Baltic languages include Lithuanian and Latvian. The Slavic languages include Russian, Ukrainian, Polish, Czech, Bulgarian, Bosnian, Croatian and Serbian. All of these languages have some fundamental similarities, but they are typically divided into separate Baltic and Slavic language families because of the similarities within those groupings. But some Indo-European family trees will combine them in a common Balto-Slavic language group initially and then separate them into separate groups after that point. For our purposes, it doesn't really matter. But it explains why the number of branches can sometimes vary. Some charts will also separate the Indian languages and the Iranian languages into separate branches as well, even though the family tree which I use follows the more conventional approach of grouping those together initially and then separating them. There are reasons why linguists argue over whether some branches of the family tree should be combined or not. A lot of that debate concerns certain assumptions about the earliest speakers of those languages and whether they initially represent a single tribe who spoke a common language which later divided into separate groups, or whether

they were in fact separate groups all along. Since none of those particular languages have any significant impact on the history of English, I will not dwell on those debates here.

Also, another language group known as Phrygian is listed on the family tree on the website, but some illustrations omit that language group altogether. Frankly, so little is known about the language at this time that some linguists don't even include it at all. Again, for our purposes Phrygian had no significant impact on English, so I will not mention it further.

I spent a large portion of the last podcast talking about Sanskrit and Persian. So let me mention a couple of things about those languages before we move on to the other languages in the Indo-European family tree. Persian is part of the branch which is typically called the Iranian branch since it represents languages which are native to modern-day Iran. This includes the ancient Persian language spoken within the Persian Empire during the times of the ancient Greeks and Romans. It also includes most of the modern Iranian languages including Farsi. Remember that despite the common religion, Iran is not an Arabic country. Ethnically and linguistically it is different. Arabic is not an Indo-European language. It is a Semitic language of the Middle East.

To the south and east of the Persian language family, we have the Indian language family which includes languages spoken primarily in northern and central India. Of course this includes Sanskrit which we have previously discussed, as well as modern Indian languages like Hindi and Urdu.

The family tree also includes the Albanian language spoken in Albania. Modern Albanian is the only language in that branch of the family tree.

Another branch of the tree which only contains one language is Armenian. It contains the modern Armenian language which is spoken in Armenia which is located south of the Caucasus mountains between the Black Sea and Caspian Sea.

There are two other branches which represent languages which are long extinct as actively spoken languages. The Hittites of modern Turkey are mentioned in the Old Testament of the Bible. In the early 1900s, a set of cuneiform inscriptions from the Hittite Empire were discovered and deciphered. It turned out – much to the surprise of many linguists – that the Hittite language was in fact an Indo-European language. It is in fact the oldest attested Indo-European language.

Another long-extinct Indo-European language was also discovered in the early 1900s in the deserts of the Xinjiang region in northwestern China. Several manuscripts were discovered by archaeologists in this region. They were translated and, once again to everyone's surprise, the language was an Indo-European language spoken in the second millennium BC. The area in which the manuscripts were found lies along the Silk Road which connected East Asia with Central Asia, the Near East and Europe. This is the easternmost discovered Indo-European language and probably represents an early Indo-European-speaking group from Eurasia which migrated along the Silk Road and eventually settled in northwestern China.

That leaves four remaining branches of the Indo-European family: the Hellenic (or Greek) branch, the Italic branch (which includes Latin), the Celtic branch, and the Germanic branch. These four language groups represents the Indo-European languages spoken in Western Europe – and consequently the four branches which directly impacted the history of English. So these four language groups will be the focus of much of this first volume of the podcast.

The Hellenic branch represents Greek from its earliest form spoken in Mycenean Greece at least by 1600 BC if not earlier. It also includes the various regional Greek dialects of the classical Greek period and the modern Greek language spoken in Greece today. Greek has had a major impact on English. Some Greek words came directly into English, but most of the Greek words in English came into English via Latin. As you probably know, the Romans were heavily influenced by Greek culture and adopted a tremendous number of Greek words. As we will see shortly, the Latin language has consistently influenced English and its predecessors since the time of the Roman Empire. As a result, many English words can be traced back to ancient Greece. The Greeks also borrowed an alphabet from the Phoenicians which is the same basic alphabet we use today – again as modified by the Romans.

The Italic branch represents the languages spoken in Italy after various Indo-European-speaking tribes settled there. Over time, the Latin dialect spoken in Rome won out as the Roman Empire came to dominate Italy and eventually the entire Mediterranean and Western Europe. Over time Latin became a dead language in the sense that people stopped speaking it as their native or first language – though it continued to be studied and learned as the language of the Church and later as a language of science and academia. The original Latin language fractured into various regional dialects, and those dialects evolved into the modern languages we know today as the Romance languages – French, Italian, Spanish, Portuguese, and a few others. So all of those languages also belong to the Italic branch, even though most of them are not spoken in Italy (except of course as learned languages). By the way, many people are under the impression that these Latin-derived languages are called Romance languages because they sound 'romantic' when they are spoken. They may very well sound romantic, but that is not why they are called Romance languages. They are called that because they are spoken in areas that were once Roman and thus spoke Latin. But the word ***romance*** or ***romantic*** actually comes from the French word for ***Roman*** as well. The term originally described a type of French literature which involved themes of chivalry. The term eventually came to refer to French literature involving a love story. So ***Rome*** is the root of ***romance*** in its conventional sense and in its use to describe Latin-based languages in Europe.

Latin is the biggest influence on English outside of English's native Germanic language family. Estimates are that as many as 1/4 of the words in a full-sized Latin dictionary have made their way into English in some form.

Latin words have found their way into English almost without interruption since English was a discernable language – and even before that.

First, during the time of the Roman Empire – before English was English when the ancestors of the earliest Anglo-Saxons were still living among other Germanic tribes in northern Europe – Latin was seeping into these early Germanic languages. The Romans traded with the Germanic tribes, they regularly employed them as mercenaries in the Roman army, and several Germanic tribes became protectorates of the Roman Empire living within the Empire and some of them eventually became Roman citizens.

Of course southern Britain was also part of the Roman Empire for about four centuries. So Latin was being spoken in Britain among many of the native Romanized Britons when the Anglo-Saxons arrived in the 5th century. These Latin influences impacted English after the Anglo-Saxons arrived in Britain.

Shortly after that, the Church centered in Rome expanded throughout western Europe and into Britain itself. From the early Middle Ages until the Renaissance, the church was the dominant unifying factor in Western Europe, not only religiously, but also politically, socially and culturally. For much of this period, most of the literacy in Western Europe was confined to the church monks. Latin was also a lingua franca which enabled travelers throughout western Europe to communicate in something close to a common tongue. Not surprisingly, Latin continued to seep into English during this period as well.

Of course in 1066, the biggest impact of all occurred when the Norman French invaded and conquered England under William the Conqueror. As I said, French is a Romance language having evolved from the original Latin spoken by the Romans, so the massive number of French words which entered English after the Norman Invasion resulted in one of the biggest deposits of Latin words into English.

Even as late as the Renaissance and thereafter, Latin continued to influence English as the language of scholarship, science and medicine.

The net result of all of the Latin influence on English is that we have a language today that is really a blended language with the two biggest influences being the Germanic languages and Latin. You can think of English in its most basic terms as a blend of these two language groups. English is not a Romance language, but it bears certain similarities to those languages due to this massive borrowing. If you've ever studied French, Spanish, Italian or another Romance language, you will constantly notice similarities in vocabulary between English and the Romance words. Some of these similarities are due to their common Indo-European roots, but most of it is a result of borrowing.

Some linguists think of English as a massive oak tree. The roots and trunk of the tree are the original Germanic Anglo-Saxon words. These are the core words of the language. While they are relatively few in number compared to the non-Germanic words which have been borrowed into English, they are the ones we use the most in day-to-day speech. Many of the first words that small children learn to speak, and later read and write, in English are Germanic words – numbers, body parts, family relations, basic verbs and pronouns. These words represent the core vocabulary of English. That is why those words don't change very much from one generation to

the next. You learn them early on and you use them everyday. So that is why they represent the roots and the trunk of the tree. They hold the tree up and they rarely change. But all of the limbs, branches and leaves of that oak tree represent the borrowed words and most of those limbs, branches and leaves are from Latin. They give the tree is shape and color. They fill up a dictionary. But if you flip through an English dictionary, you will see many words that you don't know – and many that you recognize, but hardly every use. Most of these words are the Latin and other borrowed words. We have them at our disposal if we need them, but we really tend to use them to supplement our core vocabulary which is very small by comparison and dominated by Germanic terms.

To give you some actual numbers to illustrate this point, let me read to you the most commonly used words in the English language. This is my Top-25 list. These are the Top-25 most commonly used words in the English language in order from number 1 to 25:

I, the, and, a, to, is, you, that, it, he, of, in, was, for, on, are, as, with, his, they, at, be, this, have and ***from***. All 25 of those words are from Germanic origins – either Old English or other Germanic languages which have worked their way into Modern English. None of those come from any other source.

Now let me read you the next 25. These are numbers 26 through 50:

or, one, had, by, word, but, not, what, all, were, we, when, your, can, said, there, use. And that's it. At number 42 on our list, the word ***use*** is the first non-Germanic word. It's actually an Old French word that comes into English, and that's the first word on the list that is of a non-Germanic origin. But then, if we pick up with the next word, number 43 is ***an, each, which, she, do, how, their*** and ***if***. That's the remainder of the second group of 25 words.

So out of the top 50 words, we only have one word that is not from the Germanic language family. And if we continued this out, we'd basically see the same thing. Out of the 200 most commonly used words in English, 183 of them come from the Germanic language family. Only a small handful come from other language families, and again most of those come from Latin. Yet here is the key. When the entire vocabulary of English is taken into account – when one looks at entire English dictionary for example – the Anglo-Saxon words represent a tiny fraction of the total. They are very few, but they represent the core of the vocabulary. They are the words we use most often. So again, the Anglo-Saxon words are the roots and the trunk of our oak tree using the analogy I gave, while the other languages – especially Latin – represent the branches and leaves. Keep that oak tree analogy in the back of your mind as we discuss Latin and the Germanic languages throughout this podcast series.

This leaves us with two language groups – the Celtic and the Germanic. Once upon a time, before the Romans expanded into Europe, the Celtic languages dominated much of Europe and were spoken throughout Britain when the Anglo-Saxons arrived. Of course, they are still spoken in much of the British Isles outside of England. This includes Modern Irish, Welsh and Gaelic. Unfortunately, the Celtic languages in many of these areas are slowly dying out with fewer and fewer speakers There is also one area outside of the British Isles where Celtic languages are

spoken today – in the French province of Brittany in northwestern France. The Celtic language spoken there is called Breton. The Celtic ancestors of modern Brittany actually came from Britain. As the Anglo-Saxons poured into Britain in the fifth and sixth centuries, many native Celts moved northward into Scotland and westward into Wales, but some fled southward out of Britain and across the English Channel into northwestern France where they helped to found Brittany.

Today, when we think of Celtic culture, we typically think of places like Scotland, Ireland and Wales. But these are merely the modern remnants of a culture that once dominated most of Europe. In fact the ancient Celts were the first iron age people of Europe and they are often called 'The First Masters of Europe.' The extent to which we can think of all of these Celtic-speaking people who emerged in central Europe around 800 BC as a specific group of people is the subject of ongoing debate among historians. But these people who we call Celts did share certain cultural characteristics – most notably a common language or a common group of languages.

But the ultimate story of the Celts in continental Europe is the story of being caught between a rock and a hard place. To the south in Italy and the Mediterranean was the Roman Empire with intentions of expanding northward. To the north in Scandinavia were the Germanic tribes with intentions of expanding southward. The Celts occupied the vast expanse of Europe in between. But over time the Celts would be caught in between these expanding forces and by the second century AD they were completely overtaken by the Latin-speaking Romans and Germanic tribes. And by the 6th century, they were overtaken in the area we would come to know as England by another group of Germanic tribes – the Anglo-Saxons.

Since the Anglo-Saxons emerged as the conquerors, they borrowed very little from the defeated Celts. This included the Celtic language which had some, but relatively little, impact on Modern English.

That leaves us with the ancestors of the Anglo-Saxons – the Germanic tribes. As I have mentioned, English is part of the Germanic branch of the Indo-European family tree. When we say English is a Germanic language, what does that mean? Well, it doesn't mean English came from German. Modern German and English each evolved separately from an ancient common shared language which linguists call Proto-Germanic. In that sense, English and German are related – you might say cousins – within a larger Germanic family of languages. Other modern languages within that larger Germanic family include Dutch, Danish, Swedish, and Norwegian.

The Germanic language family began with a group of early Germanic speakers in Scandinavia. Over time they migrated southward into the heart of continental Europe. Early on, these tribes divided into three distinct groups which are represented by the three branches of the Germanic language family. The tribes who remained in Scandinavia are known as the northern tribes and their language developed into Old Norse – the language of the Vikings – and eventually into modern Danish, Swedish, Norwegian and Icelandic.

The tribes which moved southward divided into two separate groups. One group moved eastward into eastern Europe, and these linguistic groups are known as the East Germanic tribes. These dialects eventually died out as these tribes became assimilated into other tribes and territories. The most notable of these tribes were the Goths who played a large role in the fall of the Roman Empire.

While some Germanic tribes migrated southeastward into eastern Europe, others migrated southward and westward in the areas of modern Germany and eventually modern France. These tribes are known as the West Germanic tribes. And this is where we find English, as well as modern German.

The Western branch of the Germanic family tree is often subdivided into two separate groups – High German and Low German. 'High' and 'Low' refers to the altitude and can be easily confused. The southern part of Germany is mountainous so the Southern part represents High German. As you move north towards the North Sea and the Baltic Sea, the elevation drops to sea level and this represents the area Low German. So High German is in the south. Low German is in the north.

The modern standard German language is spoken within the higher elevations and is considered part of High German. This also includes Bavarian and Austrian.

However, there are also German dialects spoken closer to the coast as I mentioned. So Low German includes these German dialects as well, which are usually just called the Low German dialects. Low German also include Dutch and Flemish which are spoken in the Netherlands and Belgium.

Now let's focus more closely on the Low German dialects because that is also where we find English. This is our little family within the larger Germanic family and within the much larger Indo-European family. Again, the Low German territory includes the modern Netherlands, the coastal lowlands of northern Germany and parts of Denmark. By the second or third century AD, there were a variety of Germanic tribes in these regions. The most prominent of these tribes were Saxons, Angles, Jutes and Frisians. All of these tribes participated in the general migration to Britain beginning in the fifth century, but we know these groups today as the Anglo-Saxons. The Jutes and Frisians represented a smaller portion of the total number of immigrants. Since they did not get their names in the label 'Anglo-Saxon,' we tend to ignore them, but they were definitely included. The Jutes lived in modern Denmark in the peninsula which we still now today at Jutland which basically means 'land of the Jutes.' The Jutes primarily settled in eastern England in the area known as Kent.

The Frisians lived along the North Sea coast from part of the modern Netherlands northward along the German coast. If you look at a map, you will notice that the Angles, Saxons and Jutes lived east of Frisia, and Britain is across the North Sea to the west. In the process of migrating to the west to Britain, the Angles, Saxons and Jutes would have passed through Frisia. In the process, some of these Frisians joined in the migration, but they did not represent enough people to carve out their own distinct territory in Britain like the Angles and Saxons and Jutes.

Frisian territory still exists in the Netherlands and Northern Germany. And a modern Frisian language is still spoken there and in parts of coastal Denmark.

In fact, if German is a cousin of English, Frisian is probably the closest thing we have to a sibling of English. In fact, there are remarkable similarities between Frisian and Old English. A common poem reads "Bread, butter and green cheese is good English and good Friese." This phrase is read almost identically in English and Frisian. Dutch speakers would pronounce bread, butter and cheese quite differently, but the Frisian pronunciation is very close to English and so the phrase is literally true.

The circumstances which led to the migration of the Low German tribes to Britain beginning in the fifth century will be covered in an upcoming podcast on the Anglo-Saxon migrations and the development of Old English. But for now, we will just note that these tribes settled in southern Britain. The Jutes established a territory in the east which is modern Kent. The Angles tended to settle in the northern part of this territory in central Britain. The Saxons settled in the south.

The Saxons in the east were known as the 'East Saxons' and their territory eventually became known as Essex from 'East Saxons.' The Saxons in the middle were known as the South Saxons and their territory became known as Sussex from 'South Saxons.' In the west, the Saxons were called the West Saxons and their territory was called 'Wessex' from West Saxons. And it is here that we can finally pinpoint Old English. Remember that at this early stage, there were still regional variations among these tribal groups. But most of their territory would eventually fall to Viking invaders a couple of centuries later. It was only the West Saxons – under their leader Alfred – known as Alfred the Great – that were able to withstand the Viking invasions. They eventually fought back and preserved the culture and the language of the Anglo-Saxons. Since they emerged as the dominant territory in the aftermath of the Viking disruptions, it is their dialect – the West Saxon dialect – that would come to be used in most Old English documents.

And there you have it – English within the Indo-European family of languages.

You will probably have noticed by this point that all of the major European languages are included within the Indo-European family. However, there are a few European languages which are not considered Indo-European languages. This includes Finnish in Finland, Estonian in Estonia, Saamin in northern Scandinavia, and Hungarian (also known as Magyar) in Hungary – all of which are considered part of a separate language family known as Uralic. The Basque language spoken in northern Spain is also a unique language with no ties to the original Indo-European language. And the modern Turkish languages also represent a separate family of languages.

But with these few exceptions, all of the other languages of Europe are descended from the original Indo-European language.

So what about that original Indo-European language. Where was it spoken? When was it spoken? And who were the people who spoke it? In the next podcast we will look at how linguists have actually been able to reconstruct significant portions of this ancient language. And one of the key

figures in this process was a collector of fairy tales – one of the famous Brothers Grimm. So next time we will explore Grimm's Law and look at how the Germanic languages, including English, have changed over time. And with this knowledge, we will see how Grimm's Law allows us to reconstruct a significant portion of that original Indo-European language, and in the process, see many of original sources of Modern English words and grammar.

So until next time, thanks for listening to the History of English Podcast.