In the biblical story of the Tower of Babel, all of humanity once spoke a single language until they suddenly split into many groups unable to understand each other. We don't really know if such an original language ever existed, but we do know that the thousands of languages existing today can be traced back to a much smaller number.

So how did we end up with so many?
In the early days of human migration, the world was much less populated. Groups of people that shared a single language and culture often split into smaller tribes, going separate ways in search of fresh game and fertile land. As they migrated and settled in new places, they became isolated from one another and developed in different ways. Centuries of living in different conditions, eating different food and encountering different neighbors turned similar dialects with varied pronunciation and vocabulary into radically different languages, continuing to divide as populations grew and spread out further. Like genealogists, modern linguists try to map this process by tracing multiple languages back as far as they can to their common ancestor, or protolanguage.

A group of all languages related in this way is called a language family, which can contain many branches and sub-families.

So how do we determine whether languages are related in the first place?
Similar sounding words don't tell us much. They could be false cognates or just directly borrowed terms rather than derived from a common root. Grammar and syntax are a more reliable guide, as well as basic vocabulary, such as pronouns, numbers or kinship terms, that's less likely to be borrowed. By systematically comparing these features and looking for regular patterns of sound changes and correspondences between languages, linguists can determine relationships, trace specific steps in their evolution and even reconstruct earlier languages with no written records.

Linguistics can even reveal other important historical clues, such as determining the geographic origins and lifestyles of ancient peoples based on which of their words were native, and which were borrowed. There are two main problems linguists face when constructing these language family trees.

One is that there is no clear way of deciding where the branches at the bottom should end, that is, which dialects should be considered separate languages or vice versa. Chinese is classified as a single language, but its dialects vary to the

point of being mutually unintelligible, while speakers of Spanish and Portuguese can often understand each other.

Languages actually spoken by living people do not exist in neatly divided categories, but tend to transition gradually, crossing borders and classifications. Often the difference between languages and dialects is a matter of changing political and national considerations, rather than any linguistic features. This is why the answer to, "How many languages are there?" can be anywhere between 3,000 and 8,000, depending on who's counting. The other problem is that the farther we move back in time towards the top of the tree, the less evidence we have about the languages there.

The current division of major language families represents the limit at which relationships can be established with reasonable certainty, meaning that languages of different families are presumed not to be related on any level.
But this may change. While many proposals for higher level relationships -- or super families -- are speculative, some have been widely accepted and others are being considered, especially for native languages with small speaker populations that have not been extensively studied.

We may never be able to determine how language came about, or whether all human languages did in fact have a common ancestor scattered through the babel of migration. But the next time you hear a foreign language, pay attention. It may not be as foreign as you think.