

"English Has A Word For Everything"

<https://www.youtube.com/watch?v=sVT2btZjYIM>

Transcript: <https://dontveter.com/ec/wordforeverything.pdf>

From my experience, most people who learn English later in their lives say one of three things about the language.

English is a language of simple grammar. They say English is full of exceptions and they say English has a word for everything.

While the first two comments are full of interesting implications, today we'll focus on the third and in my opinion the most underrated idea.

So why does English have so many more words than the other major languages?

The short answer is, it doesn't. It's just unbelievably good at describing the situations that surround what are otherwise very common words.

For the sake of comparison, we'll use throughout this video five major world languages across multiple linguistic families and subfamilies.

English, Italian, and Spanish, Hindi, Chinese, and Arabic.

Of those, all five have between 200,000 and 300,000 different words, excluding obsolete and technical terms.

They also are all estimated to have between 30,000 and 40,000 colloquial words in circulation with the average native speaker of each language being able to recognize around 40,000 words.

These are averages, of course, and there is implied variation. Some very well-educated native English speakers can reach 70,000 passively recognized words, while toddlers and MMA promoters have both been known to reach a thousand words.

Mixed Martial Arts (**MMA**) has grown into one of the most popular combat sports worldwide, with numerous promotions hosting elite fighters and delivering thrilling events.

Now, a niche nifty little rule called Zip's law states that the frequency of a word's use is inversely proportional to its place in the comprehensive word count list of a language.

That means, if the most common word in a language is used at a certain frequency, the second most common will have 1/2 the frequency of the first.

This pattern continues nicely to follow the curve f of $i = 1 / i$.

The important bit here is that even among people who can recognize 40,000 words in a language with hundreds of thousands, 50% of regular speech involves only a hundred different words and 80% involves a thousand.

The point is that English doesn't have more words. It has more colloquial words across a broader spectrum of linguistic origins.

And its linguistic history pushes speakers to use more of those very terms.

The key lies in a linguistic phenomenon that quantifies the specificity of the environment that words are used within.

It's called lexical collocation, but nobody really cares about the name, and I'm only going to use it like 30 more times throughout the video.

Don't worry about it. Don't worry about it.

Okay, here's what it means. In this scene from *Pride and Prejudice*, Mr. Darcy briefly notices the book Elizabeth is holding.

And of course, you must improve her mind by extensive reading.

Does he watch the book, spot the book, glance at the book, or does he peer at the book?

¿Mira el libro, localiza el libro, mira de reojo el libro o mira fijamente el libro?

You most likely said that he glances at it, which is defined as to take a brief or hurried look.

The word implies that the person glancing is only looking for a short time.

But then in this scene from *Batman Begins*, does the criminal glance at Batman?

Most people will intuitively tell you no, because the criminal wasn't the one who looked away. Batman exited his field of view.

As such, what he did was glimpse the Batman defined as a momentary or partial view.

This type of specificity in the niche used for each word extends across the whole language.

And there are a plethora of examples. Here's a good one. This is walking. This is strolling. This is sauntering, strutting, striding, marching, hobbling, waddling, shuffling, trudging. This is hiking, wandering, pacing. This is plotting, trekking.

Esto es caminar. Esto es pasear. Esto es pasear, pavonearse, caminar a grandes zancadas, marchar, cojear, contonearse, arrastrar los pies, caminar con dificultad. Esto es hacer senderismo, vagar, caminar de un lado a otro. Esto es planear, hacer trekking.

Now, I want to be fully clear with my nonnative angophone audience. These are not rare synonyms. They're all words that the majority of native English speakers know and will understand in their specific context without fail.

It isn't necessarily that any other use of the word sounds wrong, but there is one use that intuitively sounds most right because each word conjures up a specific image in the mind.

I want to also be fully clear with my monolingual native angophone audience.

This is an exceptionally rare quality among world languages. But why?

Now, at this point, you're probably asking yourself why you haven't clicked off this video yet.

Please allow me to make it even less interesting for you with some statistics and ancient linguistic history. But don't worry, here's a split screen.

Among the top 10,000 most commonly used words in English, about 40% are of Germanic origin, 40% of French origin, about 15% of Latin origin, and the rest are scattered across many languages.

While this might give us one reason why there are so many synonyms, we need to dig deeper to find why they're so consistently localized.

English's strict lexical collocations sit in the collective subconscious of speakers purely through repeated learned experience.

It's almost as if each synonym is a piece of culture. And like any piece of culture, it's also a piece of history.

The Germanic languages originated from the hypothetical reconstructed mother language, Proto-Indo-European, which was spread throughout much of Eurasia by various Neolithic farming cultures.

These peoples, called the Yamnaya by archaeologists, started migrating around 5,000 years ago from the steppes of what is now Ukraine, Russia and Kazakhstan.

Today over half of the world's population speaks a language derived from these people ranging from Spanish in Chile to Hindi in India.

Proto-germanic split off from this root language around a thousand years after the initial Yemnaya migrations, somewhere in what is now Denmark and northern Germany.

The West Germanic branch of which English is the major force was fully defined by the first century B.C.E.

Old Saxon, the direct predecessor of old English was the language the Romans encountered in their conquest of the British Isles.

They say no.

The region would continue to be dominated by the early Germanic English until arguably the most pivotal moment in the history of the language, William the Conqueror in 1066.

He led the Christian French kingdom of the Normans to conquer most of what is now southern England, bringing French to completely dominate government and aristocracy.

While colloquial layman's speech was still mostly Germanic, centuries of institutionalized French seeped into the language of the region like the scent of the Parisian subway seeps into your clothes.

At the same time, like all of the contemporary western world, the medieval Christian church used its magical powers of public execution to keep academia almost entirely in Latin.

Ah, my bad.

By the time we get to 1300's, middle English, we can already begin to see a pattern of synonyms that exist broadly within the blossoming language and are deliberately kept in their specific context by lines drawn on the borders of culture and social class.

Today, we speak in English that is maintained about 95% of the Shakespeare era's lexicon and the archaic cultural patterns he absorbed from the Middle Ages are still evident.

To further elucidate how these patterns are visible in modern day English, here's Chris.

Mic check. Mic check.

For the small fraction of Tiberio's viewers that are employed, there are words with different etymological backgrounds to describe what they do.

In colloquial discussion, especially for lower skilled roles, we mostly use the Germanic term working.

This extends to other everyday forms of toil like working out or yard work.

When the government or institutions refer to their occupation or when it's discussed in a broader economic context, the job is French derived labor and the person doing the job is the French-derived employee.

If the work involves more complex science associated elements like machinery or computers, the worker is the Latin derived operator.

Through the niche uses of our words, we subconsciously make judgments on the quality and context of an event.

If I tell you I went out to eat, it was likely a casual and lowcost meal.

If I was to say I dined, you probably imagine an upscale establishment.

The word ingest is instead exclusively reserved for medical and scientific contexts.

And if I describe to you what I ingested, I'm most likely referring to a poisoning of some sort.

Back to Tiberia. What's good? It's me again.

So, we can see how English inherited specific uses for words and how culture affected their specific niches.

But the strength of the language as a concise, descriptive tool goes deeper than just having words added to it.

English also inherited a convenient word forming tool from its original Germanic root that only adds to the strength of lexical collocation in the language: compound words.

All the Germanic languages exhibit the phenomenon of joining two entire root words into an unambiguous single term.

In English, it exists three-fold in noun noun form like bedroom or handshake (dormitorio o apretón de manos) or swamp or swamp ass.

A common condition in which the ass crack and crotch becomes overly moist, sweaty, and stinky from one or all of [the following](#):

- sweating on a hot day
- not bathing enough
- long day of work, sports, play
- incomplete wipe due to rush or laziness

Adjective noun form like blueprint or highway, and most interestingly in verb particle form like hangover or castaway.

Word compounding is not at all unique to the Germanic languages. By the way, Chinese actually uses it far more often.

And if someone speaks to you in Turkish for 10 minutes, they've probably said about 15 total words.

The difference lies in the consistency of English compound formation.

In Chinese, the main root of the compound can come first or second and can be conjugated or unconjugated. It is thus up to the listener to interpret based on context.

In Turkish, the compound words aren't exactly single words. They are strings of several modified roots with conjugations functioning as prepositions that bind the compound word.

The respective subjectivity and complexity of the two languages are the main reason why they're considered among the most difficult major world languages to learn.

In English, the main root is always the second one and always the unstressed syllable in the word.

This makes it intuitive to understand what any compound word means the first time you hear it and literally gives any two root words like 50 times the specificity in one term.

We see these compounds forming and spreading quickly even today because as long as one knows both roots and the very basic structure where the second is the main, it is immediately a hyperspecific word.

Words like doomscrolling and ragebaiting exist in brain rot lingo. Brain rot being another compound.

I mean we're talking about words created and propagated by people who can barely read and write.

They can even be fabricated with no real world context and immediately understandable like George Lucas's lightsaber or Tolken's halfling.

George Orwell invented 24 compound words just in his novel 1984, most famously doublespeak, thoughtcrime, and unperson.

Whether it's the first time someone tells you they're starstruck upon meeting you, or whether that's swampass, whether that same person tells you you have swampass, you're probably going to understand.

So, we see the layers that make up how English came to be this way.

You take three competing languages, give them the ability to create specific words whenever they want, compound that for a thousand years, and you find yourself with a language that has a word for every damn thing you can come up with.