

# How many words do I need to know?

## The 95/5 rule in language learning, Part 2/2



Welcome to Part 2 of my post on “How many words do I need to know? The 95/5 rule in [language learning](#)”. If you haven’t done so already, read through [Part 1](#) before continuing!

How many words are there in some of the world’s major languages? As I stated in Part 1 of this article, there is really is no way to answer this question. Languages are evolving and continuously changing, and subject to people’s own creativity and imagination. After all, it is said that Shakespeare himself invented [1,700 new words](#)!

People continuously invent new words, alter some existing ones, or stop using others altogether. Plus, what about medical and scientific terms? Should they be counted as part of our “vocabulary”? And if we look at the English language, for example, what should we do about Latin words used in law, French words used in cooking, German words used in academic writing, Japanese words used in martial arts? Do you count Scots dialect? Teenage slang? Abbreviations?

The most “objective” measure that we have available for counting the number of words contained in a given language, then, is to calculate the number of words contained in its largest dictionary (really, it’s not that objective, but it’s the only measure we have access to!). I thus began to research answers to this question in regards to some of the world’s major

languages, but quite surprisingly, I couldn't find any resource on the net actually listing languages and their associated number of words based on dictionary word count. So after having scoured the net for scattered answers, I'd love to share with you my findings.

So here's a list for 11 of the most spoken languages around the world (sources given as hyperlinks):

LANGUAGE	LARGEST DICTIONARY	NUMBER OF WORDS
Chinese	<a href="#">汉语大词典 (Hanyu Da Cidian. Lit: Comprehensive Chinese Word Dictionary)</a>	370,000 words; 23,000 head Chinese character entries
English	<a href="#">The Second Edition of the 20-volume Oxford English Dictionary</a>	171,476 words in current use, and 47,156 obsolete words; 615,100 definitions
Dutch	<a href="#">Woordenboek der Nederlandsche Taal (Dictionary of the Dutch language)</a>	430,000 words
French	<a href="#">Le Grand Robert de la langue française</a>	100,000 words; 350,000 definitions
German	<a href="#">Der Duden</a>	135,000 words
Italian	<a href="#">Grande dizionario italiano dell'uso (Gradit)</a>	270,000 words
Japanese	<a href="#">(日本国語大辞典)Nihon kokugo daijiten</a>	500,000 words (this includes definitions and etymologies of foreign loan words (gairaigo, 外来語), highly recent words (gendai yōgo, 現代用語), archaic words (kogo, 古語), idiomatic compound phrases (jukugo, 熟語), words that can be written using

		more than one possible Chinese character to produce subtle differences in meaning (dōkun iji, 同訓異字), and Chinese characters that are written differently but have the same pronunciation (iji dōkun, 異字同訓), some slang (ingo, 隱語), and words used only in regional dialects (hōgen, 方言))
Korean	<a href="#">표준국어대사전 (Korean Standard Unabridged Dictionary)</a>	500,000 words (this includes 190,000 technical words (전문어), 70,000 North Korean words (북한어), 20,000 regionalisms (방언), and 12,000 old sayings (옛말))
Russian	<a href="#">Толкóвый словáрь живóго великорúсского я́зыка</a> (Explanatory Dictionary of the Living Great Russian Language; AKA Dahl's Explanatory Dictionary)	200,000 words
Spanish	<a href="#">Diccionario de la Real Academia Española</a>	100,000 words
Portuguese	<a href="#">Vocabulário Ortográfico da Língua Portuguesa</a>	Nearly 390,000 words

## Is it true that English has the most words of any language?

The first thing that will probably jump to your eyes, here, is the apparently low word count for English. If you do a quick Google search, you will find

out easily enough that many claim that English has “the most number of words of any language” out there, with several hailing the “millionth word” milestone recently reached in the English language.

So why only 171,476 words in current use? Well, again, when comparing the largest dictionaries out there, we have to keep in mind several important points: Which country has the best-developed dictionary industry? The best archives? Do you count obsolete words? Dialectal ones? How many scientific words are included?

In Korean, for example, the largest dictionary ever compiled was the [result of 8 years of work](#), through the collaboration of over 500 scholars, for a total cost surpassing 11.2 billion Korean won (~\$11.2M). The dictionary includes nearly 200,000 technical words in itself, and thousands of old sayings no longer in usage.

Specialized vocabulary used in sciences is most notably very large and growing constantly. The French “*Dictionnaire de la chimie de Duval*” (Duval Chemistry dictionary), far from being exhaustive since we already distinguish over 100,000 coloring matters, already contained [26,400 entries](#) in 1935, and more than 70,000 in 1977.

Therefore the reason why English has 171,476 words in current use in its largest dictionary is partly because the dictionary excludes inflections, does not cover several technical and regional vocabularies, and does not, obviously, include words not yet added to the published dictionary. If distinct senses were counted, according to the Oxford Dictionary, the total word count would probably approach three quarters of a million.

## **Which language has the biggest vocabulary, then?**

As you can see, the list I compiled does not necessarily tell us which language has the “biggest vocabulary”. It simply tells us which dictionary was made to include the most words.

In any case, if I had to give a short answer to this question, I'd say "Who cares?" Each and every language is amazingly rich and interesting in its own way. Each language has its own genius and its own personality. Arabic has apparently over [fifty different words for "camel"](#). In Korean, there are over five different words for each color equivalent in English (i.e. red, blue, yellow, etc.) and several thousands of words have both a pure Korean and a Sino-Korean (한자어) equivalent.

The reason I compiled a list of the number of words in the dictionaries of some of the world's most widely spoken languages is simply out of sheer curiosity, not to stir up a debate over which language has the most words. This question, once again, has no definite answer.

What does matter to you as a language learner, though, is to know the approximate number of words needed in order to reach conversational [fluency in a language](#). Of course, you could very well learn a language without ever asking this question and, frankly, it wouldn't matter the least. But it's still nice to know. And this number is the approximate amount of words you will actually have to more or less "deliberately" memorize before reaching a point where you can essentially learn almost only through context and good guesswork. For more on that, see [Part 1](#) of this article.

## **How many words does a native speaker use in daily life?**

"[Green Eggs and Ham](#)," is a book written by Dr. Seuss (a pen-name of Theodor Seuss Geisel), whose vocabulary famously consists of just fifty different words. It was the result of a bet between Seuss and his publisher, Bennett Cerf, that Seuss (after completing *The Cat in the Hat* using 225 words) could not complete an entire book using so few words.

Obviously, if one can write a book using as few as 50 words, it makes no doubt that having a vocabulary of 40,000 words is not necessary for communicating. For your information, though, according to Susie Dent, lexicographer and expert in dictionaries, the average active vocabulary of

an adult English speaker is of around [20,000 words](#), with a passive one of around 40,000 words.

What is the difference between an active and a passive vocabulary? Simply put, an active vocabulary is comprised of words that you can recall and use in a sentence yourself. A passive vocabulary, on the other hand, is a vocabulary that you can recognize and know the definition of words, but are not able to use yourself.

Now, here's where it gets interesting: although an average adult native English speaker has an active vocabulary of about 20,000 words, the *Reading Teachers Book of Lists* claims that the first 25 words are used in 33% of everyday writing, the first 100 words appear in 50% of adult and student writing, and the first 1,000 words are used in 89% of every day writing! Of course, as we progressively move to a higher percentage, the number of words starts to dramatically increase (especially after 95% of comprehension), but it has been said that a vocabulary of just 3000 words provides coverage for around 95% of common texts (such as news items, blogs, etc.). [Liu Na and Nation \(1985\)](#) have shown that this is the rough **amount of words necessary before we can efficiently learn from context with unsimplified text.**

When it comes to Chinese, approximately [3,000 characters are required to read a Mainland newspaper](#). The PRC government defines literacy amongst workers as a knowledge of 2,000 characters, though this would be only functional literacy. Of course, given the nature of the Chinese language, 3000 characters equals to many, many more words. Nevertheless, the highest level (VI) of the new Hànyǔ Shuǐpíng Kǎoshì (HSK), also known as the Chinese Proficiency Test, is a [vocabulary of 5000 words](#) (2633 characters).

Finally, in French, the 600 most common words [apparently account for 90% of words](#) found in common texts, although I cannot verify the veracity of this claim. But I think you can see from the numbers here that really, in

order to understand the biggest part of a language, it is not necessary to know tens of thousands of words. Generally speaking, a vocabulary of about 3000 words (not counting for inflexions, plurals, etc.), then, would be the number necessary to efficiently learn from context with unsimplified text.

## Do the Math

We have seen that the Oxford English Dictionary contains 171,476 words in current use, whereas a vocabulary of just 3000 words provides coverage for around 95% of common texts. If you do the math, that's 1.75% of the total number of words in use! That's right, **by knowing 1.75% of the English dictionary, you'll be able to understand 95% of what you read.** That's still just **7.5% of the average passive vocabulary of a native speaker (3000 vs. 40,000 words).** Isn't that great news?

Let's repeat the math for Chinese. The Hanyu Da Cidian contains 370,000 words, whereas 2500 words (1710 characters)

are necessary in order to “read Chinese newspapers and magazines and watch Chinese films”, according to the [HSK test](#) (level 5). That's 0.68% of the total number of words contained in the Hanyu Da Cidian! Knowing 5000 words, the minimum number required to pass the highest HSK test (level 6), would mean knowing 1.35% of the total number of words contained in the Hanyu Da Cidian.



## Pareto's Law and Language Learning

We will end this already lengthy article by once more taking a look at Pareto's Law, also known as the 80-20 rule. If you've already forgot, the law states that for many events, roughly 80% of the effects come from 20%

of the causes. In other words, in the context of work or study, 20% of the efforts bring in 80% of the results.

If we drop the unrealistic figures of the number of words in the largest dictionaries out there, and we instead count the number of words an average educated native speaker knows, which is around 30 to 40 thousand for many languages, we will find out that Pareto's Law works on steroids! In many cases, knowing just 5-7% of the total number of words that a native speaker knows will allow you to understand anywhere from 90 to 95% of the vocabulary found in common texts! That's right, 5 to 7% of the effort brings you 95% of the results. That is great news for you my friend.



Italian Economist Vilfredo Pareto

So yes, languages contain fabulous numbers of words, and for many, learning a foreign language seems like an insurmountable barrier, something that takes dozens of years to accomplish. But the fact is, by learning from the very beginning words in context (I highly recommend the [Assimil method](#)), and by gradually building your vocabulary to around 2500-3000 words, it is possible to reach quite rapidly a level at which you will be able to read common texts in the language and understand anywhere from 90 to 95% of it. This is essentially the “golden” number, since this amount of understanding is enough not to make reading in the language a frustrating experience. More importantly, though, **this is the rough amount of words necessary before you'll be able to efficiently learn from context.**

Pdf converted using <http://converttifftopdf.com>