How many words do you know in your native language?

Answer:

My native language is Mandarin. My Chinese Vocabulary Size is around **4000 ~ 5000**, or more specifically 4000~5000 types of words (**entry in a vocabulary**, uncorrelated) in Simplified Chinese, and know how to understand compound of these 4000~5000 types of words to read Newspaper, web pages, medical records, financial filings, product reviews, discussion forums, social media, etc.

Corpora (a collection of writing) and Vocabulary (unique, the amount of recognized character that you can read, write, speak, and listen). A Chinese high school student knows 20000-30000 **Corpora** and 4000 **Vocabulary**. By the way, the requirement of Vocabulary in Chinese high school is 3500.

From the research in Chinese education department, when you "know" 500 Chinese words, then you can "read" 78% of Chinese Newspaper; when you "know" 1000 Chinese words, then you can "read" 90% of Chinese Newspaper; when you "know" 3000 Chinese words, then you can "read" 99.99% of Chinese Newspaper. Here, the **know** means you need to understand the morpheme of words and read means understand the article meaning.

In a Chinese **dictionary** book called "通用规范汉字表" (sampling with a dictionary), they categorize common used characters into three levels: the first level is 3500 words, which is very common for daily use, you should know every words to communicate with no trouble. The second level is 3000 words, a general high school student in China should know 30~40 % of these words to able to write and read complex content. The third level is 1000+ words, if you do not major in Chinese, then you probably know less than 10 words from them.

People do not need use **lemmatization** to reduce inflected forms to base forms such as "龍" (Traditional Chinese) to "龙" (Simplified Chinese), which all means Dragon in Chinese.

There is **no** inflectional morphology (actually very little so people ignore them in daily use) or derivational morphology in mandarin, all it matters is **word segmentation** (Compounds). For example, "内 nei" means inside, "卷 juan" means reel. However, the combination of two "内卷 nei juan" means peer competition or peer pressure, which is irregular expression compared to original meaning independently.

In general, most of meaning of compounds can be found through their original meaning of word individually. For example, "吃 chi" means eating, "饭 fan" means rice, then "吃饭 chi fan" means having meal.

Another classical example for Chinese vocabulary is meaning of "Me". In English, we can use "I, Me, Mine" to express the meaning of myself. In Chinese, it can be "某，鄙人，区区，微臣，下官，咱家，老夫，老衲，臣妾，贱妾，奴婢，贫僧，老朽，小的,余,予,愚,朕,洒家，奴才，草民，罪臣，本官，本宫，哀家，小可，本座，小女子，为师，为父，为夫，

为妻，为兄，愚兄，愚弟，私，己", and many more words to express meaning of myself in Chinese based on your age, sex, job, and etc. However, most of time, we just use "我 wo" to express myself although we know other words to express the meaning of myself.

Chinese words can generally be divided into two categories: one is the monosyllabic word, and the other, the polysyllabic. Monosyllabic word usually single character to express only one meaning, like "人 ren" means people only and there is no other morphemes of this character.

Polysyllabic is more complex because is composed one or two or more morphemes, which is hard to test them and categorize them into Chinese vocabulary. "Due to the fact that the permissible patterns of the monosyllabic structure are very restricted in Chinese, there are a very large number of monosyllabic homophones in modern Chinese; hence, word formation tends to become more and more polysyllabic. Nowadays roughly two thirds of the basic lexicon of everyday Chinese can be considered consisting of polysyllabic words" (Lu 1965; Comrie 1987).

I use a Chinese testing website (an online resource): https://www.arealme.com/chinese-vocabulary-size-test/en/ About the word vocabulary size, different institutions have different definitions. The website have made references from 《语言文字论辩集》，《汉语与拼音文字》，《54 种中西文化比较》 to build their own model and have created this quiz.

I think this website do not reflect very accurate result because it uses Chinese Slang and phrase to format some question in the quiz which may be suitable to certain people, but a lot of words is not being generally used nowadays.

I do like the website provide antonym, number, subject-verb-object, adjective, Noun, Modifier, Time, Adverbial and even idiom to test the vocabulary, I do feel my native language has been degenerated because I haven't back to China for four years and use English most of time instead of Chinese.