

# The Power of the 'The'

---

The Most Frequently Used Word in the English Language

---



Want to learn a new English word every day? If so, check out: [www.owad.de](http://www.owad.de)

Did you know that about 6% of everything you say, read, and write uses the word *the*? *The* accounts for about one out of every 16 words we encounter on a daily basis. This article, for instance, contains the word *the* 37 times. You could therefore predict that the length of this article will be  $37 \times 16 = 592$  words. The actual length is 607 words, an **accuracy** of 97.5%!

Now, you may think there's some **jiggery-pokery** on my part, especially on account of the excessive use of the word *the* in this article. But just consider *Alice in Wonderland*, which contains 1,638 instances of *the*. Thus, the predicted length would be  $1,638 \times 16 = 26,208$  total words. The actual length of *Alice in Wonderland* is 26,669! An accuracy of 98.3%!

The top 20 most common English words in their order of occurrence are *the, of, and, to, a, in, is, I, that, it, for, you, was, with, on, as, have, but, be, they*.

This word order reveals a most unusual pattern concerning the frequency of words in all the languages of the world, and in all the books, reports, and articles written or typed by the human hand. This pattern is called Zipf's Law, named after George Kingsley Zipf (1902–1950), a Harvard University linguist.

Zipf's Law states that the second most frequent word (in English it's *of*) will appear around one half as often as the most used word (*the*), the third

word (*and*) one third as often, the fourth word (*to*) one fourth as often ... and so on, all the way down.

So, the good news in learning any new language is that **roughly speaking**, nearly 50 per cent of any book, article, report, or conversation will be comprised of the same 100 most frequently used words. The other 56% will be words that appear in that selection around once.

For example, *Alice's Adventures in Wonderland* contains 26,669 words. Of these, about 14,000 words (56%) are repetitions of the most frequent 100 words. The remaining 11,000 words (44%) occur about once.

Zipf's Law isn't just used to describe word frequencies – it also predicts city populations, website traffic, earthquake magnitudes, last names, cookbook ingredients, chess move openings, and even the rate at which we forget information.

Some things get most of our attention, some get only a little, and most of what we experience on a day-to-day basis is forgotten. *The Dictionary of Obscure Sorrows* by John Koenig offers a word for this: *Olēka*, the awareness of how few days are memorable. I've been alive for almost 27,000 days, but I couldn't tell you anything about even a fraction of them. Everything we do, see, think, say, hear, and feel is forgotten at a rate quite similar to Zipf's Law, which makes sense. A few things we remember well, but most things barely, or not at all.

Thinking about thinking, the vocabulary of our memories is similar to language. About 100 of our remembrances **comprise** around 50 per cent of what we think about; the remaining 50 per cent are unique feelings and thoughts.

It makes me wonder that so much is forgotten, even things that at the time I thought I could never forget: books I've read, films I've seen, places I've visited. It's interesting that so many memories seem to have disappeared. Or have they? People suffering near-death experiences report seeing their whole life flashing by in cinematographic detail.

And as we grow older, memories of childhood slowly return, suggesting that deep in our minds, those 50 per cent of unique experiences may not be lost after all.

Paul Smith

## Vocabulary Trainer

• <b>accuracy</b>	Genauigkeit/Richtigkeit
• <b>jiggery-pokery</b>	Gemauschel (ugs.)
• <b>roughly speaking</b>	grob gesagt
• <b>to comprise</b>	umfassen