

## Summer Pulls Ahead of Winter in Eight Languages!

A search for the words "summer" and "winter" in online linguistic corpora reveals that references to "summer" (in published books) have come from behind to surpass those of "winter" during the course of the nineteenth and twentieth centuries, *in eight different languages*. This reversal in the relative frequencies of "summer" and "winter" is shown in the graphs of the [Google Books Ngram Viewer](#) for Chinese, English, French, German, Italian, Russian, and Spanish, and likewise for Portuguese in Mark Davies's [Corpus do Português](#).

The Ngram Viewer also represents books published in Hebrew, but the summer/winter reversal is not clear-cut in that language. "Summer" (קיץ) is more frequent than "winter" (חורף) from the 1840s to the present consistently, as well as prior to that period generally. I am not qualified to say how faithfully these words correspond to English *summer* and *winter*, in a language that presumably developed in a climate quite different from that of England. (You can [view the graph](#) for Hebrew online, remembering that the red line is for "summer" and blue for "winter".)

Nor am I prepared to speculate at this moment on any cultural implications that the summer/winter reversal may support—I will leave that exercise to the [culturomicists](#). My aim here is simply to present the data on the frequency of "summer" and "winter" over time in the eight corpora.

### English

For English generally, the Ngram graphs for "summer" and "winter" are never *far* apart from the early 1800s to the present, and in fact beginning around 1810 they follow similar curves, going up together and down together. But they do cross: In 1820 the frequency of "summer" is about five sixths that of "winter"; then, for a period centered around 1880, they are nearly equal; and finally by 2000, the frequency of "summer" has risen to more than 1½ times that of "winter". You can [view the graph](#) online.

Prior to 1800 the picture is very different: "winter" is even more frequent, compared to "summer". In fact, from about 1720 to about 1790, summer is hardly mentioned at all, compared to winter. But *caveat lector*: the makers of the Ngram Viewer point out that relatively few books were published before the nineteenth century, and this makes it easy for small differences to be drastically magnified, even if they are not statistically significant.

English, unlike the other languages counted, can be subdivided into American, British, fictional, and "English One Million", which is a modified form of the English corpus that equalizes the number of books representing each year, in order to reduce the "magnifying" effect mentioned above. In each of these subdivisions of English, "winter" starts out more frequent than "summer" and ends up less so, with the crossover period in the mid-nineteenth century.

### French

French for "summer" is *été*, but that is also the past participle of the extremely frequent

verb *être* "to be". In order to count only the noun, I searched for it (and for French "winter") in combination with the definite article: *l'été* and *l'hiver*, respectively. Again, "winter" is more frequent than "summer" all through the nineteenth century and up until the early 1940s, when both lines on the graph take a slight, temporary bump up in popularity, after which "summer" is more frequent—about 70% more so by 2000. ([View the graph.](#))

## German

The German words for "summer" and "winter" are *Sommer* and *Winter*. The nineteenth century begins with winter somewhat more frequent than summer. From around 1860 to 1900 they are approximately equal, and then in the twentieth century "summer" gradually pulls ahead, rising to about one and two thirds the frequency of "winter" by 2000. ([View the graph.](#))

## Italian

The Italian words for "summer" and "winter" are *estate* and *inverno*. The frequency of "winter" is almost twice that of "summer" for most of the years from about 1750 to about 1850. Summer finally surpasses winter in the mid-1950s, and is exceeding it by some 30% at the end of the 20th century.

## Russian

The Russian words for "summer" and "winter" are *лето* and *зима*, respectively, which might be transliterated as *léto* and *zimá*. Prior to 1840, "winter" is far more frequent than "summer". Beginning around 1840 the two words fall into a pattern of near equality until the 1910s and '20s, at which point "summer" takes a huge surge in frequency, so that by 2000, "summer" is more than 2½ times as frequent as "winter". ([View the graph](#), with summer red and winter blue.)

Russian nouns alter their form according to the grammatical case in which they are being used. "In summer" and "in winter" are expressed in the instrumental case: *летом* and *зимой* (*letom*, *zimoi*) respectively, and these forms are much more frequent than the nominative forms given above. With these forms, summer's move to pass up winter is even more striking: Summer is hardly ever mentioned before the 1910's, at which time it surges to more than twice the frequency of "winter". ([View the graph](#)).

## Spanish

On the Ngram graph for Spanish, the lines of *verano* and *invierno* ("summer" and "winter" respectively) are extremely close together, especially after 1840, but "winter" remains slightly more frequent than "summer" up until the mid-1950s, when "summer" pulls slightly ahead. By 2000, the instances of "summer" outnumber those of "winter" by some 40%. ([View the graph.](#))

## Portuguese

Books in Portuguese are not (yet) accessed by the Ngram Viewer, but word-frequency

data for this language can be gathered from the [Corpus do Português](#) (produced by Mark Davies at Brigham Young University). Here, searches are focused century by century. It is not possible to pinpoint the year or decade when the graphs for *verão* ("summer") and *inverno* ("winter") cross, but we can observe that the summer/winter ratio in the *nineteenth* century is 426/453 (94%), while in the *twentieth* century, the ratio is 1,305/1,073 (122%)—summer has moved from slightly less frequent to somewhat more frequent.

## Chinese

For Chinese books, the Ngram Viewer refers only to those using the [simplified characters](#) adopted in the People's Republic of China in the 1950s, so our comparison of "summer" and "winter" will not reach back anywhere near the nineteenth century. Nevertheless, in the years since 1950 the graphs show "winter" (冬季 *dōngjì*) slightly more frequent up until about 1980, and then, as with the other languages mentioned here, "summer" (夏季 *xiàjì*) pulls slightly ahead in frequency thereafter. ([View the graph](#); as with the other graphs linked above, the blue line is for "winter" and the red line for "summer".)

## Conclusion

Corpus data for eight languages—Chinese, English, French, German, Italian, Portuguese, Russian, and Spanish—suggest that earlier books published in each language mentioned winter more frequently than summer, but that at various dates (or rather periods) between the 1830s and the 1980s, the books shifted to the reverse order: mentioning summer more than winter. In chronological order, the crossover dates are as follows:

1830s	Russian (nominative case)
1880s	English
1900	German
1900?	Portuguese
1910	Russian (instrumental case)
1940s	French
1950s	Spanish, Italian
1980	Chinese