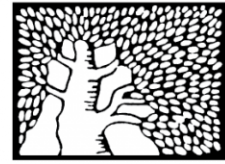


מכון ויצמן למדע

WEIZMANN INSTITUTE OF SCIENCE



## The History of Question Marks and Exclamation Marks across Languages

### Document Version:

Early version, also known as pre-print

### Citation for published version:

Pilpel, Y 2022, 'The History of Question Marks and Exclamation Marks across Languages: Punctuation Culturomics', *Default journal*, pp. 1-6. <https://doi.org/10.34933/wis.000523>

*Total number of authors:*

1

### Digital Object Identifier (DOI):

[10.34933/wis.000523](https://doi.org/10.34933/wis.000523)

### Published In:

Default journal

### License:

Other

### General rights

@ 2020 This manuscript version is made available under the above license via The Weizmann Institute of Science Open Access Collection is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognize and abide by the legal requirements associated with these rights.

### How does open access to this work benefit you?

Let us know @ [library@weizmann.ac.il](mailto:library@weizmann.ac.il)

### Take down policy

The Weizmann Institute of Science has made every reasonable effort to ensure that Weizmann Institute of Science content complies with copyright restrictions. If you believe that the public display of this file breaches copyright please contact [library@weizmann.ac.il](mailto:library@weizmann.ac.il) providing details, and we will remove access to the work immediately and investigate your claim.



# The History of Question Marks and Exclamation Marks across Languages: Punctuation Culturomics

Yitzhak Pilpel  
Department of Molecular genetics  
Weizmann Institute of Science  
[pilpel@weizmann.ac.il](mailto:pilpel@weizmann.ac.il)

**Abstract** What can punctuation tell us about cultural historical trends? Here I analyze the change in frequency of the punctuation marks '?' and '!' in six languages over the last two centuries by a culturomics study of the Google Books online repository. I found that in German, Italian and Spanish the ratio of usage of question marks to exclamation marks sharply declines towards the Second World War, and steadily increases thereafter, whereas in English '?' was always more heavily used. A common trend in all languages is a rise in '?' compared to '!' in the second half on the 20th century. Furthermore, over the last decades, the usage in English of 'Why', 'How' and 'What' - open ended questions, with often no definitive answer - has tended to increase in frequency more sharply than 'Where', 'When', and 'Who'. I propose that the relative usage of question marks and the type of questions asked may serve as a meaningful dynamic measure of the cultural state of societies.

Culturomics is a form of computational lexicology that studies human behavior, language, cultural and historical trends through the quantitative analysis of texts (Michel et al., 2011). A major source of culturomics data has been Google Books, a Google service that allows to search the full texts of books and magazines scanned and converted to digitized text and stored in a database. This database contains books printed in nine different languages from the year 1500, though predominantly from 1800 until 2009. The main output of an inquiry is a temporal representation profile of each word (the so-called "*n-gram*"), depicting the number of its appearances in the corpus of scanned books in any given year, normalized to the total number of words (or *n-grams*) scanned during that year. The Ngram Viewer enables browsing of the data online.

In the first Culturomics publication, Michel et al. (2011) analyzed trends of linguistic, cultural and historical significance on Google Books. Among their interesting findings was the discoveries that the usage of the English lexical item *God* has steadily decreased over the last century, and that the effect of censorship in Nazi Germany could be deduced from the sudden disappearance of ‘*Marc Chagall*’ specifically from German literature during the Second World War (WWII).

It stands to reason that the Culturomics not only of words and names but also of punctuation marks could be a reflection of cultural change and historical trends. To explore this possibility, I followed the usage of the question mark (‘?’) and exclamation point (‘!’) over the course of the 19<sup>th</sup> and 20<sup>th</sup> centuries in six languages: UK and USA English, French, German, Italian and Spanish. While the usage of ‘?’ obviously denotes the tendency to ask questions, the usage of ‘!’ is somewhat more complex, as it ends imperative sentences, but may also be used to express sentiments, such as astonishment or excitement.

This Culturomics punctuation exploration revealed both similarities and differences among the usage of ‘?’ and ‘!’ in the various languages. In all the languages examined, the usage of ‘?’ increased during the second half of the 20<sup>th</sup> century (Fig. 1A).

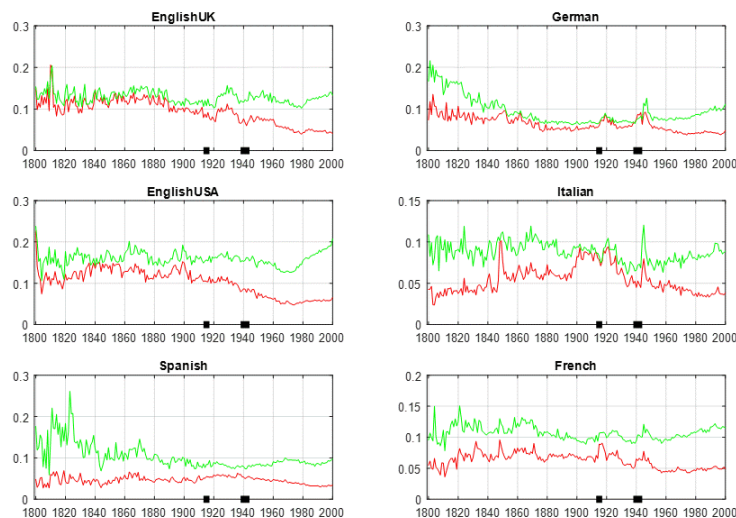


Figure 1A: Relative usage of ‘?’ (green) and ‘!’ (red) in six languages from 1800-2000. The two World Wars are denoted as black rectangles on the x-axis.

However, earlier trends differ between the languages. In German, ‘?’ was almost twice as prevalent as ‘!’ at the beginning of the 19<sup>th</sup> century, but this gap steadily narrowed with time, and either nearly or entirely disappeared around the year 1850, and then again at the end of the second decade of the 20<sup>th</sup> century, at the end of the 1920s, and toward the end of the 1940s (Fig. 1A). Since then, the gap has steadily widened, with ‘?’ increasing and ‘!’ decreasing in usage. The seemingly coincidental decline in usage of ‘?’ in the period between the two World Wars is intriguing. Equally remarkable is the local peak in usage of ‘!’ during WWII, followed immediately thereafter by an increase in the prevalence of ‘?’.

Curiously, the trends in Italian bear resemblance to those observed in German, as in this language, too, is seen a narrowing of the gap between ‘?’ and ‘!’ during the first decades of the 20<sup>th</sup> century, followed by a widening of this gap shortly after WWII, with a steady increase in the usage of ‘?’, and a decline in that of ‘!’, in the ensuing years. Similar to German, Italian also featured a local increase in the relative usage of both punctuation marks toward the end of WWII.

The trends in American and British English are more similar to one another than to the European languages examined here, featuring a rather steady increase in the gap between the usage of ‘?’ and ‘!’ throughout the entire 20<sup>th</sup> century. Notably, though German is phylogenetically closer to English than to Italian, with respect to the differences between the usage proportions of the two punctuation marks over the last two centuries, German resembles Italian much more. This phenomenon is shown in a cluster analysis comparing the similarities and differences between all six languages with respect to the relative usage of the two punctuation marks over the previous two centuries (Fig. 1B). This analysis demonstrates that the two English corpora (USA and UK) show very similar patterns of usage and cluster tightly together, with French more remotely related, while German, Italian and Spanish reside in a different cluster altogether. Common to German, Italian and Spanish is the relative narrowing of the gap in usage between ‘?’ and ‘!’ roughly in the period between the two World Wars.

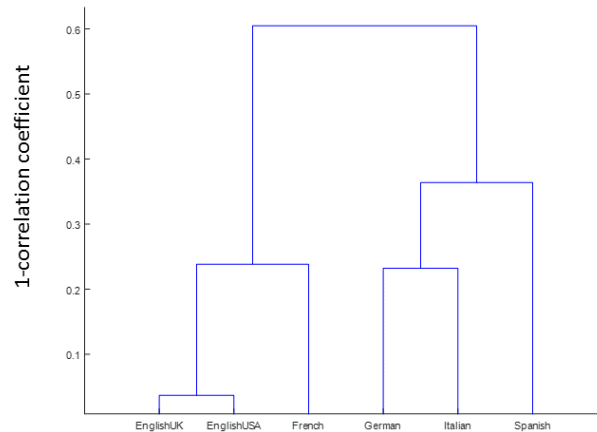


Figure 1B: Cluster analysis of the six languages with respect to the relative usage of ‘?’ and ‘!’ during the second half of the 20<sup>th</sup> century. Hierarchical clustering was run using Matlab’s clustering tool, with a 1-Pearson Correlation coefficient as the distance metric, and the Average Linkage option.

To gauge which types of questions drove the increase seen in the relative frequency of the question mark in English during the last decades of the 20<sup>th</sup> century, I examined the relative usage over time of the six common WH question words in English, ‘Who’, ‘What’, ‘Where’, ‘When’, ‘Why’ and ‘How’. I found that while the relative usage of ‘Why’, ‘How’ and ‘What’ clearly increased during the last decades of the 20<sup>th</sup> century, coinciding with, and hence probably contributing to, the increase in the relative usage of ‘?’ during that period, the usage of the three other WH question lexicals, ‘Who’, ‘When’, and ‘Where’, either remained constant or displayed a more modest change (Fig. 2A).

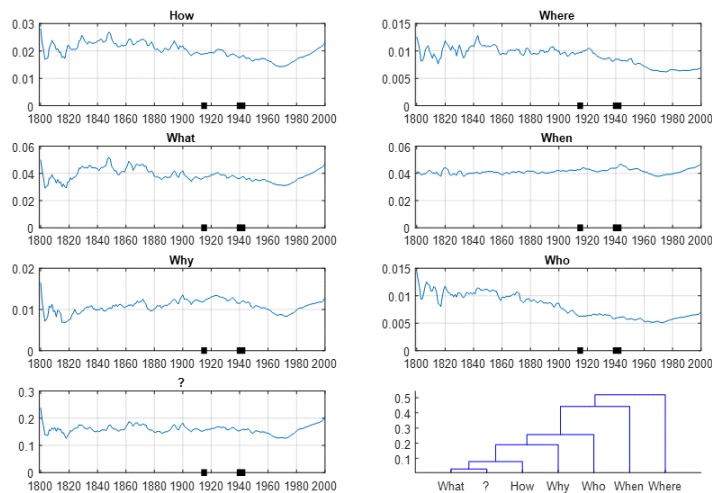


Figure 2. Relative usage patterns in English of the six WH question lexicals and the question mark. Lower right panel shows the cluster analysis of the seven lexicals over the last 5 decades of the 20<sup>th</sup> century; clustering performed as in Figure 1A. All lexicals were searched but only those starting with a capital were counted, e.g., ‘Who’ was counted but not ‘who’, since the former stands for a question, while the latter serves to construct compound sentences.

To more objectively compare the patterns of usage of the six lexicals over the last five decades of the 20<sup>th</sup> century (when ‘?’ displayed a pronounced trend of increased usage), a cluster analysis was conducted on WH question word usage. I found that ‘What’, ‘How’ and ‘Why’ cluster closely with the question mark itself, while ‘Who’, ‘When’ and ‘Where’ are clustered further away, and hence deduced to affect the increased usage of ‘?’ to a lesser extent.

Why did the usage of ‘What’, ‘How’ and ‘Why’ increase at a steeper rate compared to ‘Who’, ‘When’ and ‘Where’? One possible reason could be related to the notion that ‘Why’, ‘How’ and ‘What’ are generally used in open-ended questions, which often lack definitive answers (e.g., ‘Why am I here?’, ‘How do you build a startup company?’, and the ultimate, ‘What is the meaning of life?’) vs. questions that often lead to more well-defined answers (such as ‘Who is the president of the USA now?’ or ‘Where am I’).

What might be the reason(s) underlying our increased tendency to ask open-ended questions? Why do they increase in frequency more sharply? How can we interpret the dynamics in the ratio between the prevalence in usage of ‘?’ and ‘!’ over time; i.e., the “?/! index”? Admittedly, being unable to actually read all the books scanned for this analysis, I can only speculate. Could the ?/!

index serve as a simple, albeit meaningful, means to characterize the nature and state of societies? What can we learn from the coincidence of low index values in German and Italian literature during the period between the two World Wars? And why do all the six languages analyzed feature a steady increase in the use of question marks during the last decades of the 20<sup>th</sup> century? I suggest that this simple lexicology measure might correlate, and may even affect or be affected by, major historical, political and social trends in the life of a nation, and might even serve to predict them.

### **Bibliography**

Jean-Baptiste Michel et al. 2011. 'Quantitative Analysis of Culture Using Millions of Digitized Books', *Science* 331 (6014): 176–182.

### **Acknowledgements**

I thank the Braginski Center for the Interface between the Sciences and the Humanities, and the Minerva Center on Live Emulation of Evolution in the Lab, both at the Weizmann Institute of Science, for grant support, and Ghil'ad Zuckermann and Brian Towers for helpful discussions.