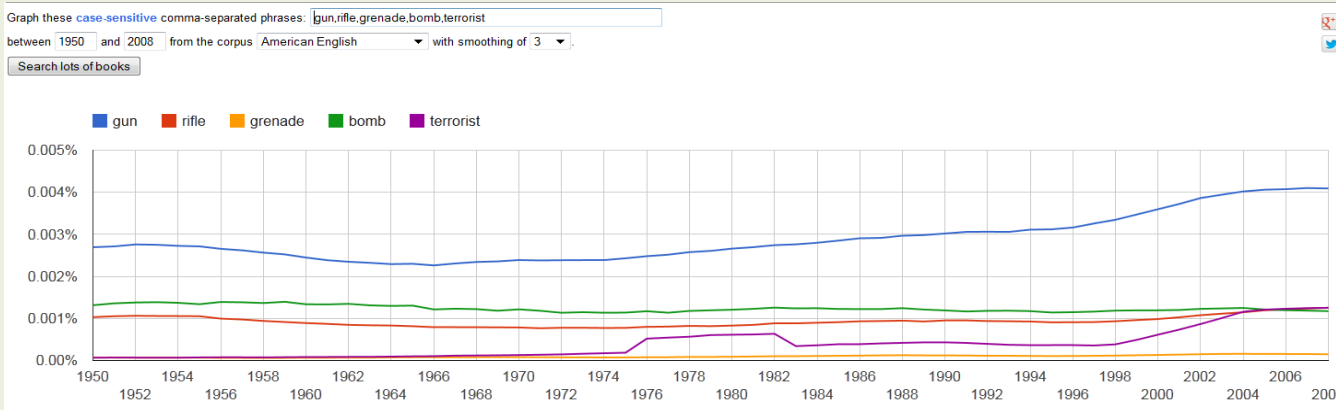


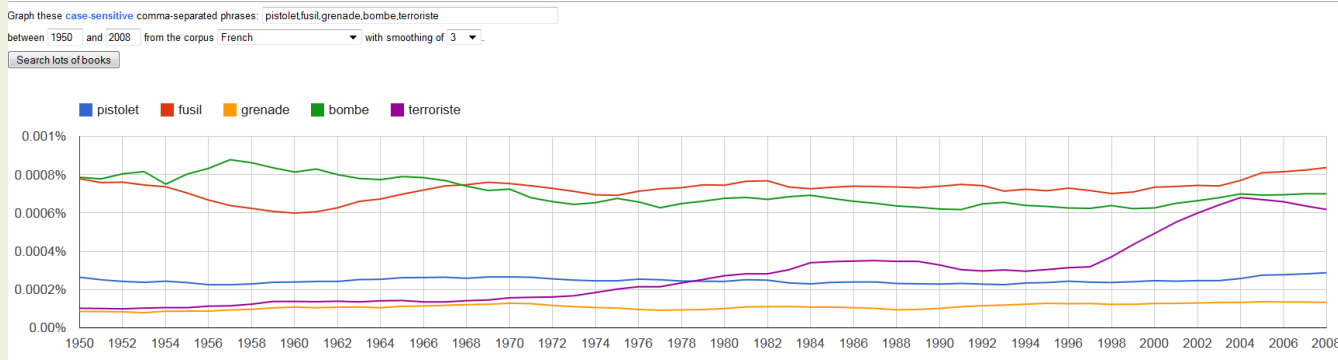
# Ngram google search

- Search in millions of books
- in different languages and
- In different corpuses.
- Can restrict to nouns or verbs etc by appending the syntax `_NOUN` etc
- Reference: <http://books.google.com/ngrams/info>
- To search, click on <http://books.google.com/ngrams/>
- Some examples:

# Ngram google: Frequency of appearance of weapons in millions of books in USA English versus French



- The frequency of the word **gun** in millions of book has grown in USA from 0.003% in 1950 to 0.004% in 2008. The words **rifle** and **bomb** are steady with a frequency of 0.0012%



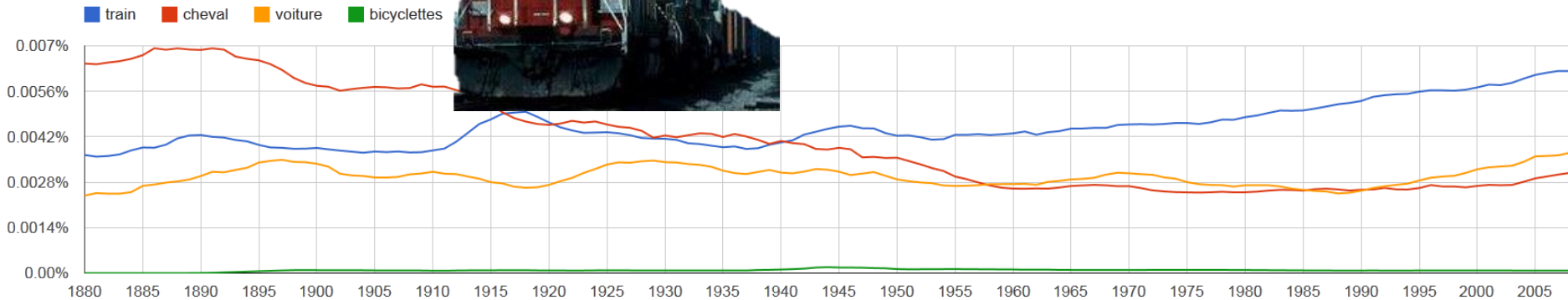
- French: the word corresponding to **gun**, has a steady frequency of 0.0002%, 20 times less than in USA English books. The frequency of **rifle** and **bomb** is around 0.0008%, which is two thirds of the frequency in USA English



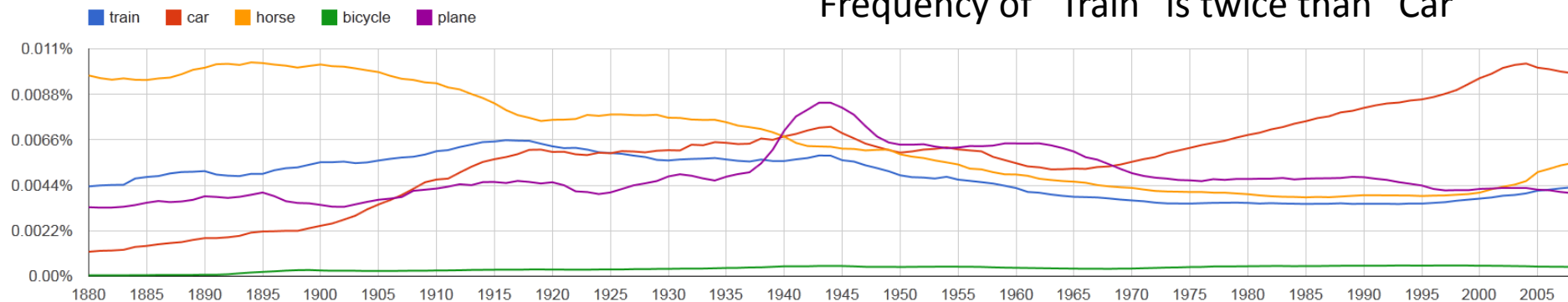
# Public versus private transportation: Comparing with ngram google the fraction of appearance of *trains* and *cars* in millions of books.



English books: **Car** beats **Train** since 1970  
Today, the frequency of car is twice that of train



French books: **Train** beats **Car** since 1950  
Frequency of "Train" is twice than "Car"

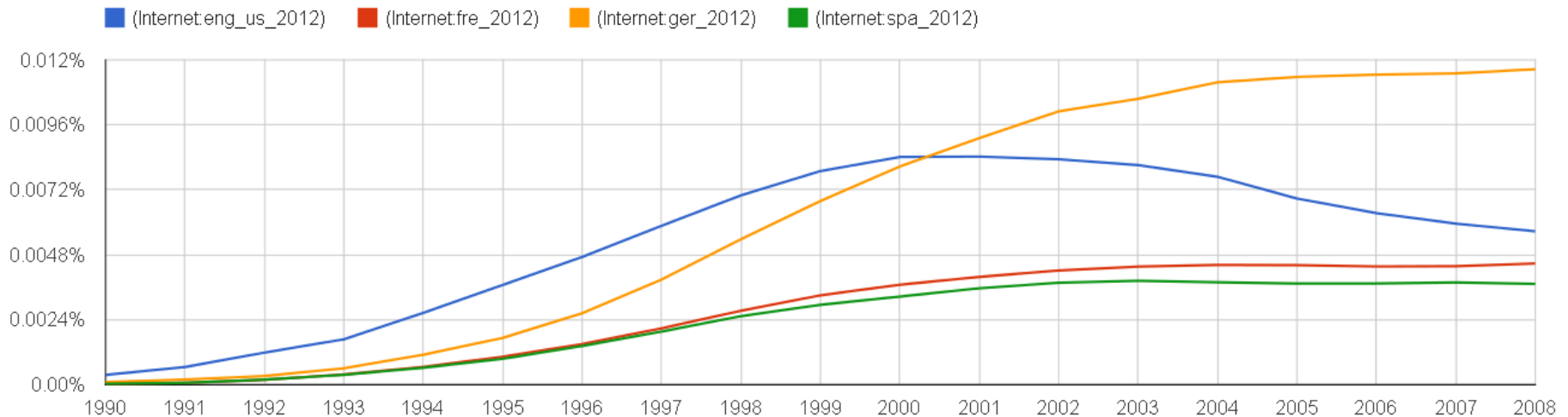


# Comparing the frequency of “Internet” in million of books in USA English, French, German, Spanish.

- In USA the “peak” occurred on 2001, and then there is a decline.
- In the other countries there is still an increase.
- Germany leading.
- We obtained all languages on a single graph by specifying the languages in the search

## Google books Ngram Viewer

Graph these **case-sensitive** comma-separated phrases:  between  and  from the corpus  with smoothing of .



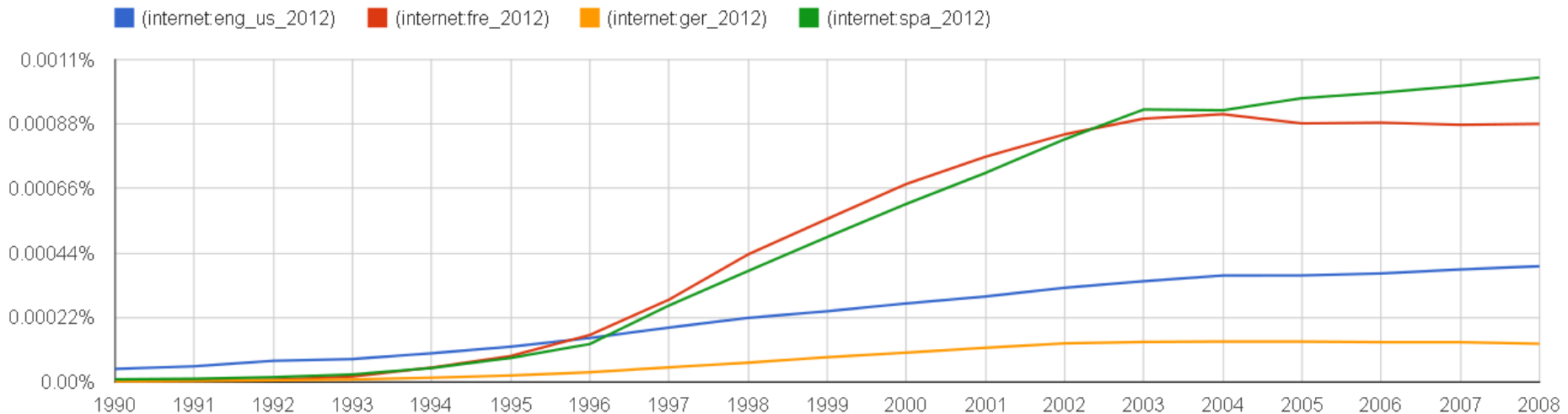
# Correction: search is case sensitive

- We searched for *Internet* . What about *internet*?
- Much less than Internet. But increasing in all languages.
- Very few in German. Why? In German, nouns start with capital letters.

Google books Ngram Viewer

Graph these **case-sensitive** comma-separated phrases:  between  and  from the corpus  with smoothing of .

[Share](#)   
[Tweet](#)



# Adding up to get case insensitive words

- Use + sign to add up (with and without capital letter).
- Minus, division and multiplication are also defined.

Google books Ngram Viewer

Graph these **case-sensitive** comma-separated phrases:  between  and  from the corpus  with smoothing of .

[Share](#)   
[Tweet](#)

■ (Internet:eng\_us\_2012 + internet:e... ■ (Internet:fre\_2012 + internet:fre\_20... ■ (Internet:ger\_2012 + internet:ger\_2... ■ (Internet:spa\_2012 + internet:spa\_...

