# Tracing Antisemitic Language Through Diachronic Embedding Projections
## France 1789-1914

**Rocco Tripodi, Massimo Warglien, Simon Levis Sullam, Deborah Paci**
**Ca' Foscari University - Opinion Dynamics and Cultural Conflict in European Spaces**

### Objective

Study the changes over time in the semantic spaces of words related to the Jewish question.

### Contributions

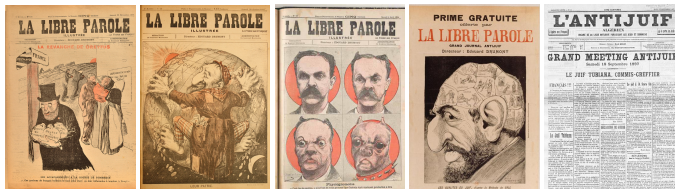
Large French historical corpus

Identification of antisemitic moments[1] and language

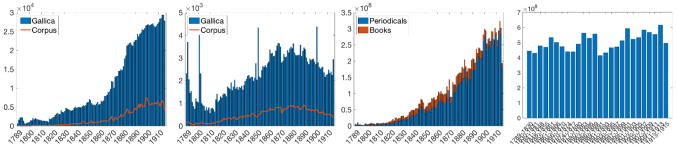
Generalisation of the bias computation

### The corpus

54.403 books and 245.188 periodicals issues published between 1789-1914 downloaded from https://gallica.bnf.fr




26 time bins 450 millions tokens each.




(a) Periodicals distribution (b) Books distribution (c) Tokens distribution (d) Num. of tokens in each bin

### The streams

1. Religious, antisemitism based on religious prejudices and accusations. Seed words: believer- unbeliever;
2. Economic: based on stereotypes concerning Jews' economic behaviours. Seed words generosity- greed;
3. Socio-political: based anti-national political behaviours. Seed words: honor - shame;
4. Racial: antisemitism based on the definition of Jews as a race, considered inferior. Seed words: pure - impure;
5. Conspiratorial: based on conspiracy theories. Seed words: loyal - disloyal;
6. Ethic: based on Jewish supposed unethical or perverse morals or behaviours. Seed words: moral - immoral.

### Word embedding

Each bin we trained with a *word2vec skip-gram* model[2] using a window *size of 5 words*, a *300 dimensions vector* and removing the words that occur less than *25* times.

### Local similarity measure[3]

Changes in the semantic space of the words:
*juif* (noun/adjective, masculine, singular)
*juifs* (noun/adjective, masculine, plural)
*juive* (noun/adjective, feminine, singular)
*juives* (noun/adjective, feminine, plural).

Computed as:

$$d(\mathbf{s}_i^{t1}, \mathbf{s}_i^{t2}) = 1 - \text{cos-sim}(\mathbf{s}_i^{t1}, \mathbf{s}_i^{t2})$$

$$s_i^t = \text{cos-sim}(\mathbf{w}_i^{(t)}, \mathbf{w}_j^{(t)}) \forall w_j \in N_k(w_i^{(t)}) \cup N_k(w_i^{(t+1)})$$

### Embedding Projections[4]

To quantify biases in word embeddings semantic spaces a word vector is projected on a semantic axis. The semantic axis can be computed as $\mathbf{g} = \mathbf{w}_i - \mathbf{w}_j$ and its projection as the dot product $\hat{b} = \mathbf{w} \cdot \mathbf{g}$. The higher the values of the projection, the more biased the word is toward that direction.

The projections are computed on six semantic axes, that correspond to six antisemitic streams.

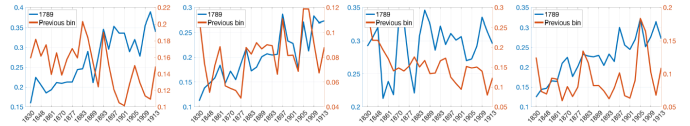
We identified for each stream a set of n antonyms pairs, to construct the bias subspace in the embedding. To quantify the biases for all the time we computed the mean bias, $b$, for each stream as the arithmetic mean of the individual biases, $\hat{b}$ on each axis, as:

$$b(w_i, s) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{w}_i \cdot (\mathbf{w}_{a_j^{neg}} - \mathbf{w}_{a_j^{pos}})$$

### References

1. Pierre Birnbaum. 2011. The anti-semitic moment: A tour of France in 1898. Chicago University Press.
2. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. NIPS 2013.
3. William L Hamilton, Jure Leskovec, and Dan Jurafsky. Cultural shift or linguistic drift? Comparing two computational measures of semantic change. EMNLP 2016.
4 Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. INIPS 2016.

### The rise of antisemitic language



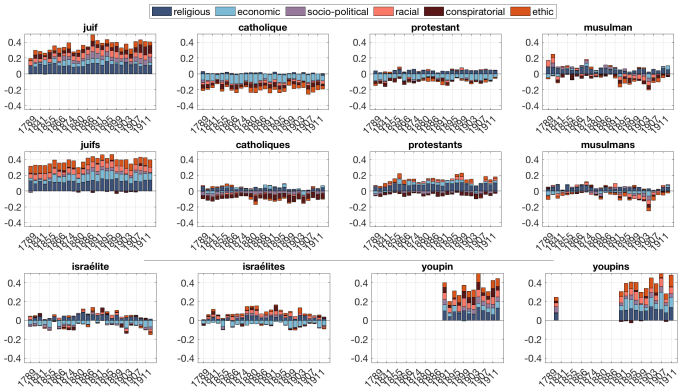

(a) juif (b) juifs (c) juive (d) juives

Figure 2: Local neighborhood measure. The $y$ axes indicates the cosine distance of the second-order vector constructed for each time period compared to the 1789 (blu line) and the preceding time period (red line).

| juif | | juifs | | juive | | juives | |
|---|---|---|---|---|---|---|---|
| ⤵ 1841 | ⤴ | ⤵ 1861 | ⤴ | ⤵ 1874 | ⤴ | ⤵ 1870 | ⤴ |
| laquedem | juive | crucifient | juif | huguenots | judaïque | syriennes | négociantes |
| mécréant | judaïque | schismatiques | israëlites | favorite | musulmane | iraniennes | samaritaines |
| rogatons | rabin | judaïsants | juive | opera | syrienne | musulmanes | réfugiées |
| blasphémateur | bouddhiste | fétichistes | rabbins | rigoletto | héroine | israëlites | ascètes |
| ⤵ 1886 | ⤴ | ⤵ 1870 | ⤴ | ⤵ 1886 | ⤴ | ⤵ 1880 | ⤴ |
| ghetto | judaïque | judaïsants | juif | drumont | iranienne | israëlites | épousées |
| déicides | rabin | hérétiques | synagogues | antisémitisme | apostasié | musulmanes | luthériennes |
| francmaçon | wanderghen | cabalistes | talmud | circoncis | lithuanienne | femmes | turques |
| aryen | anabaptiste | luciférien | sanhédrin | théâtrale | puritaine | célébrations | dissolues |
| ⤵ 1893 | ⤴ | ⤵ 1897 | ⤴ | ⤵ 1893 | ⤴ | ⤵ 1897 | ⤴ |
| déicide | talmud | antisémites | samaritains | juiverie | synagogue | juif | dissolues |
| youtre | bouddhiste | youtres | talmud | satanisme | héroine | youtres | baptisées |
| francmaçon | sodomite | youpins | idolâtres | monogamique | lapidée | antijuives | prostituaient |
| youpins | anabaptiste | enjuivés | pharisiens | opprimée | persécutrice | antisémitiques | ascètes |
| ⤵ 1897 | ⤴ | ⤵ 1905 | ⤴ | ⤵ 1901 | ⤴ | ⤵ 1905 | ⤴ |
| youtre | rabin | judaïsants | synagogues | stigmatisant | dragonnade | massacrées | courtisannes |
| sémite | usurier | hellénisants | talmud | antijuive | torturée | terrorisées | païennes |
| judaïsant | shylock | diaspora | pharisiens | antinationale | puritaine | diaspora | prostituaient |
| antisémite | anabaptiste | massacrant | ismaélites | dreyfusiste | anabaptiste | déportées | émigrées |

Table 1: Words that have been introduced (left column ⤵) or eliminated (right column ⤴) for our 4 target words in time periods with a high local neighborhood distance, compared to 1789.

### Projections




### Links

Article: https://www.aclweb.org/anthology/W19-4715
Code and data: https://github.com/roccotrip/antisem