

LUND
UNIVERSITY

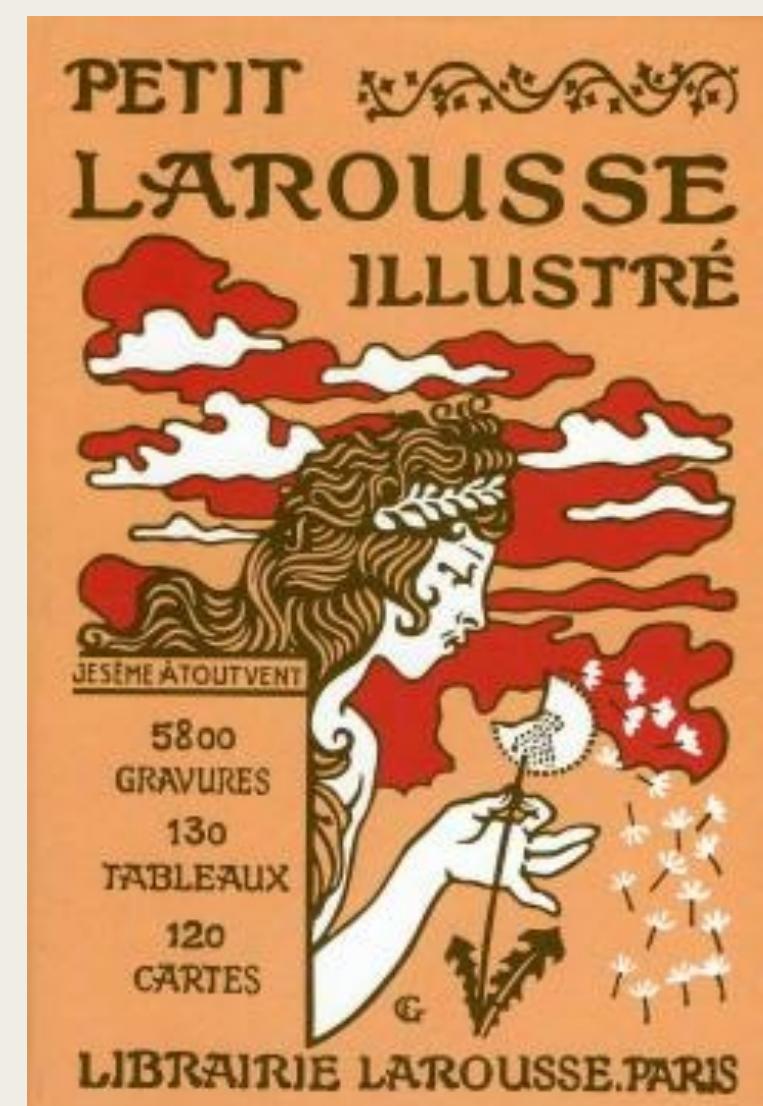
Connecting a French Dictionary from the Beginning of the 20th Century to Wikidata

PIERRE NUGUES

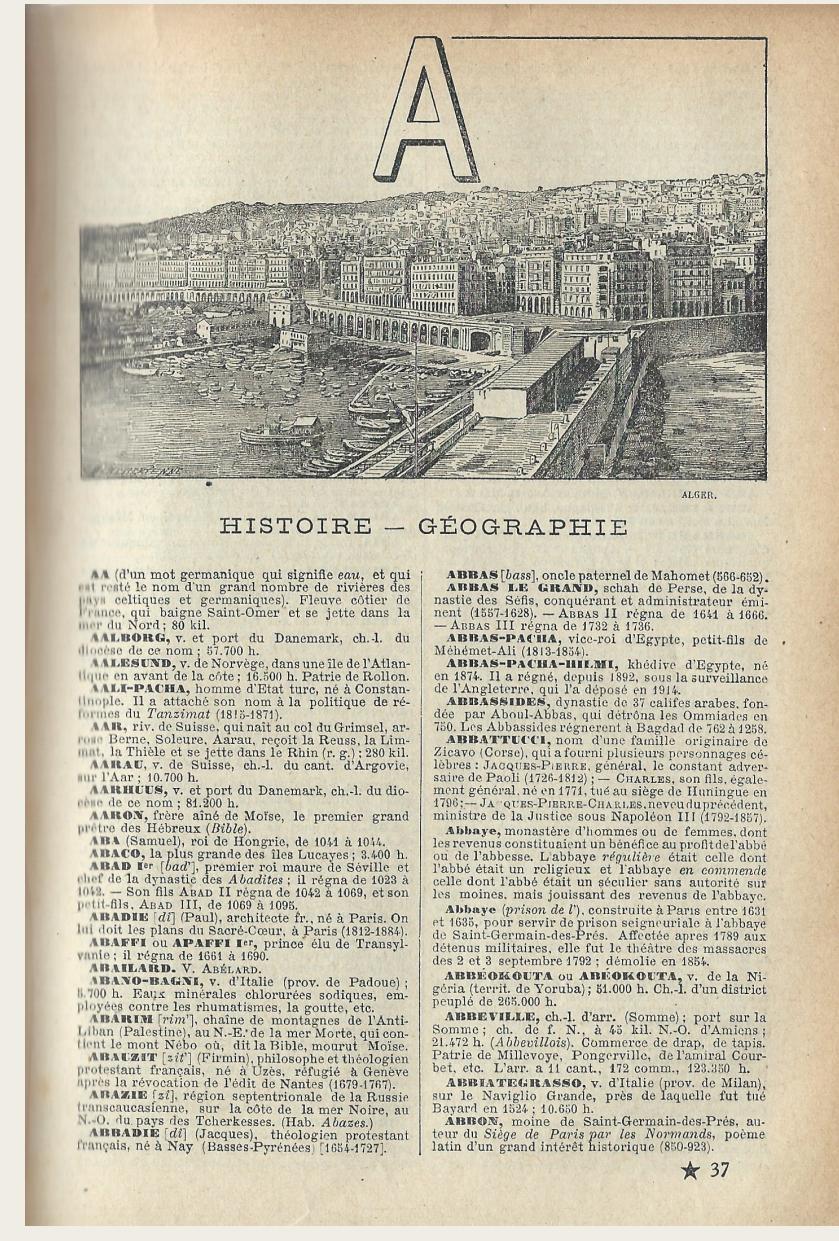
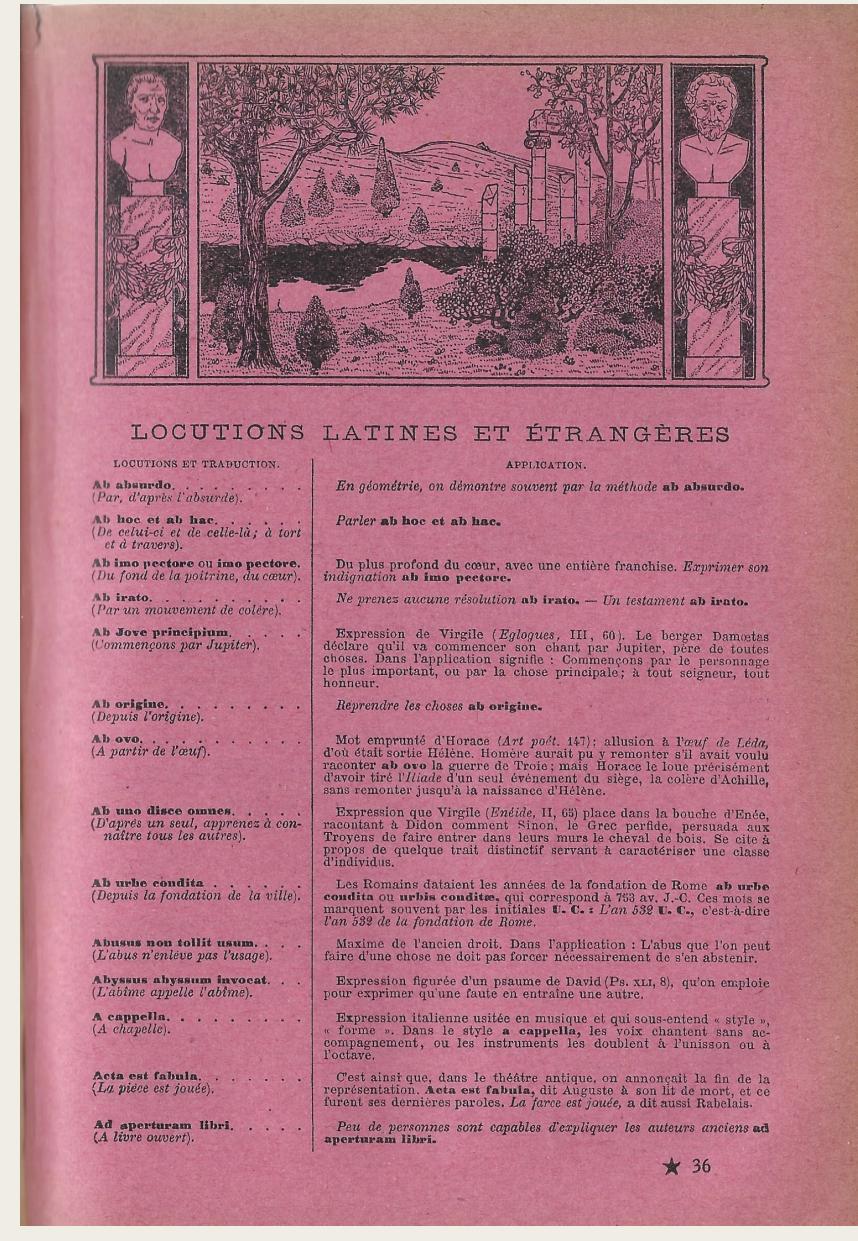
COMPUTER SCIENCE, LUND UNIVERSITY, PIERRE.NUGUES@CS.LTH.SE

Petit Larousse illustré (1905)

- One-volume dictionary of French from 1905
- Definitions of words, things, and people
- Numerous illustrations
- Very popular and far-reaching cultural influence.
- Updated yearly
- Digitized by the Nénufar project

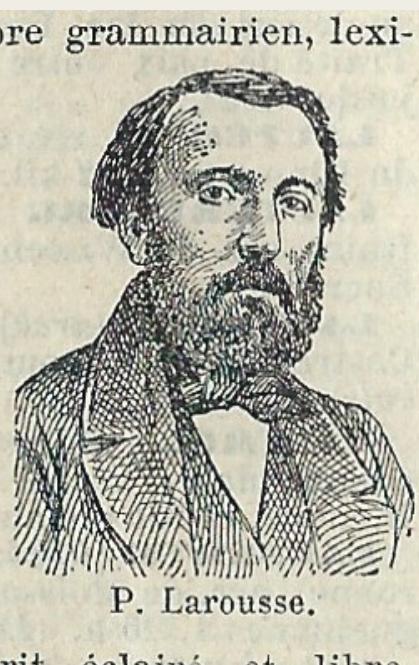


The Three Parts of the *Petit Larousse*



Wikidata: A Part of Wikipedia

Petit Larousse illustré



Wikidata item

[Item](#) [Discussion](#) [Read](#) [View history](#)

Pierre Larousse (Q313709)

French grammarian, lexicographer and encyclopædist

Pierre Athanase Larousse

In more languages

| Language | Label | Description |
|----------|-----------------|------------------------------------------------------|
| English | Pierre Larousse | French grammarian, lexicographer and encyclopædist |
| French | Pierre Larousse | grammairien, lexicographe et encyclopédiste français |
| Italian | Pierre Larousse | linguista, pedagogista e editore francese |
| Swedish | Pierre Larousse | No description defined |

All entered languages

Statements

| instance of | human | + 2 references |
|-------------|-------|----------------|
| image | | |

Larousse (Pierre) → wd:Q313709

Connects to all the
Wikipedia versions

And contains structured
information

Wikipedia (35 entries)

| ar | بَيْر لَارُوس |
|-----|--------------------------|
| arz | بَيْر لَارُوس |
| az | Pýer Laruss |
| bg | Пиер Ларус |
| br | Pierre Larusse |
| ca | Pierre Larousse |
| cs | Pierre Athanase Larousse |
| da | Pierre Larousse |
| de | Pierre Larousse |
| el | Πίερ Λαρούς |
| en | Pierre Larousse |
| eo | Pierre Larousse |
| es | Pierre Larousse |
| eu | Pierre Larousse |
| fi | Pierre Larousse |
| fr | Pierre Larousse |
| hr | Pierre Larousse |

| | | |
|-----------------|---------------------------|-----------------|
| date of birth | 23 October 1817 Gregorian | + 12 references |
| place of birth | Toucy | + 2 references |
| date of death | 3 January 1875 Gregorian | + 11 references |
| place of death | Paris | + 2 references |
| cause of death | cerebrovascular disease | + 1 reference |
| place of burial | Montparnasse Cemetery | + 0 references |

Entity Linking

- We connected manually all the 20,245 entries from the history and geography part to wikidata items (and hence to wikipedia information).
- This enables us to:
 - Verify information;
 - Extract more data on the entities;
 - Apply some automatic analyses;
 - Compare historically-situated representations with current descriptions

Resource

The annotated dataset is available in JSON
https://github.com/pnugues/petit_larousse_1905/.

Link Example

AALI-PACHA, homme d'Etat turc, né à Constantinople. Il a attaché son nom à la politique de réformes du Tanzimat (1815-1871).

'AALI-PASHA, Turkish statesman, born in Constantinople. He attached his name to the policy of reforms of the Tanzimat (1815-1871).'

JSON

```
{"texte": "AALI-PACHA, homme d'Etat turc, né à Constantinople. Il a attaché son nom à la politique de réformes du Tanzimat (1815-1871).", "qid": ["Q439237"]}
```

| | | |
|----------------|----------------------------|----------------|
| given name | Mehmed | + 0 references |
| Emin | + 0 references | |
| All | + 0 references | |
| noble title | pasha | + 0 references |
| date of birth | 5 March 1815 Gregorian | + 4 references |
| place of birth | Istanbul | + 1 reference |
| date of death | 7 September 1871 Gregorian | + 3 references |
| place of death | Istanbul | |

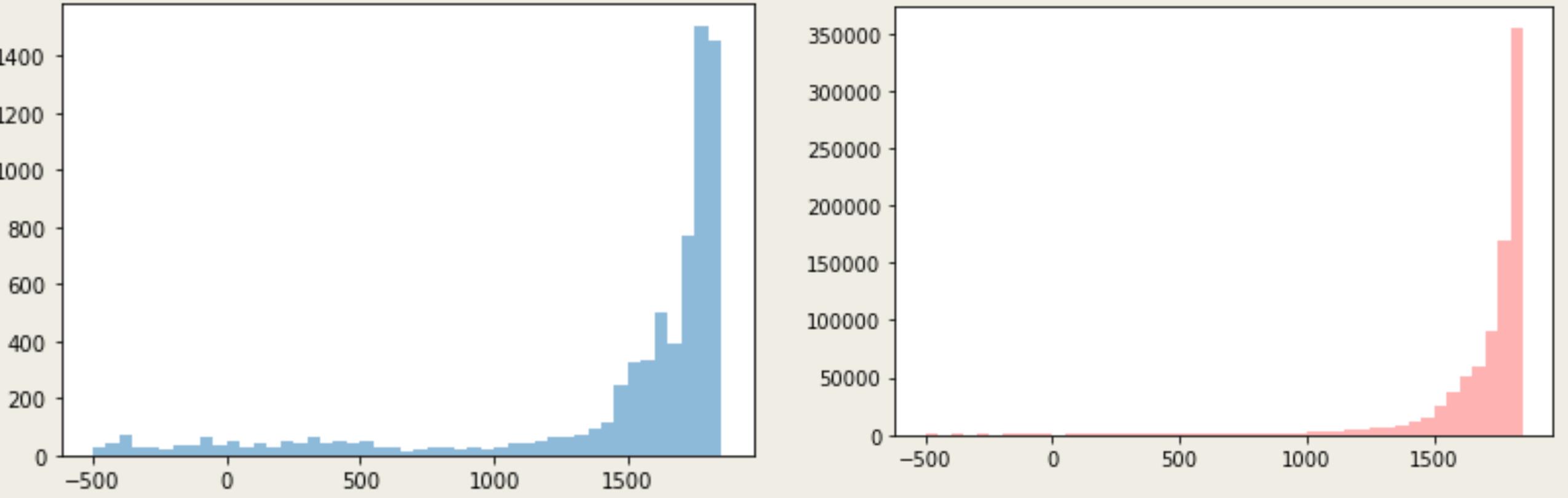
Semantic Categories

SPARQL query: wd:Q439237 wdt:P31 ?type .
?type → Q5 (human)

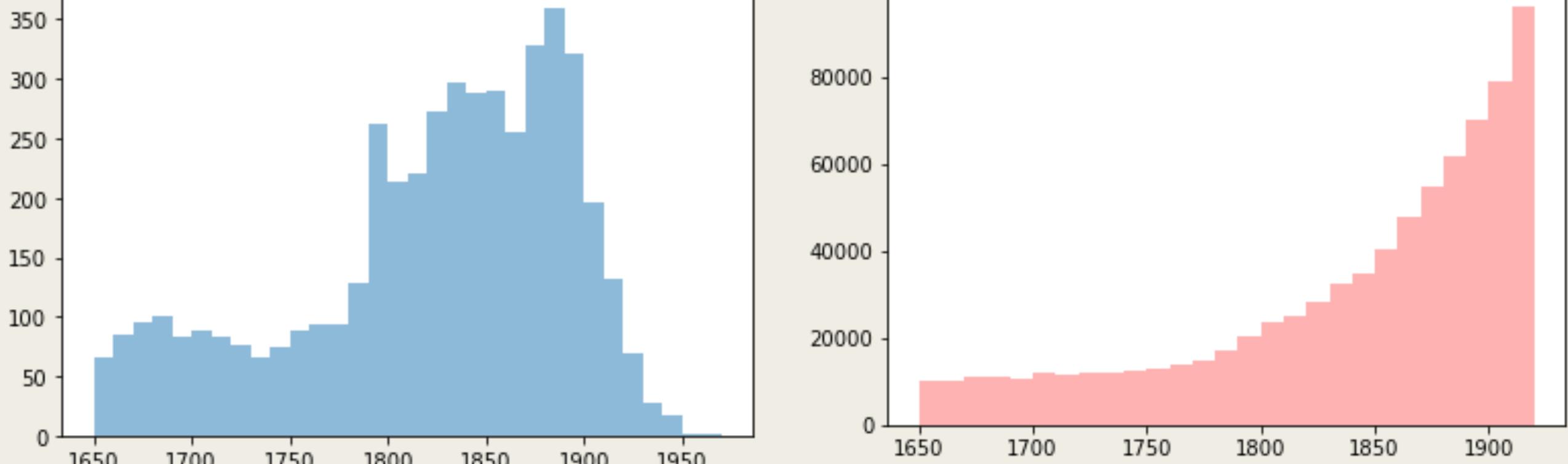
| Q-Number | Frequency | Description | Q-Number | Frequency | Description |
|----------|-----------|-------------------|----------|-----------|-----------------------------------------|
| Q5 | 7653 | human | Q4022 | 447 | river |
| Q484170 | 3022 | commune of France | Q3305213 | 308 | painting |
| Q1549591 | 849 | big city | Q747074 | 255 | comune of Italy |
| Q515 | 789 | city | Q1637706 | 231 | city with a pop. of more than 1,000,000 |
| Q7725634 | 572 | literary work | Q23442 | 221 | island |

Birth and Death Dates

wd:Q439237 wdt:P569 ?db .



wd:Q439237 wdt:P570 ?dd .



Petit Larousse illustré

Wikidata

Geographical Coordinates

?e wdt:P625 ?geo .

