1. The Dataset of Parisian Auction Sales (1830-1939)

- **Auction catalogues** of Parisian auction sales: transcription of the whole information in a dataset
- Matched with the **minutes** of the auction sales, through the catalogue number (hammer prices, sellers and purchasers of each artefact)

= **216,445 paintings, drawings and sculptures** from 1831 through 1939, with their prices.
2. Pricing the Words, in Auction Catalogues

This dataset was first analyzed through hedonic regressions, in order to « explain » the hammer prices of artworks, with a series of variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Made By&quot;</td>
<td>-10.13971</td>
<td>19.66024</td>
<td>0.606</td>
</tr>
<tr>
<td>&quot;Attributed to&quot;</td>
<td>-28.86939</td>
<td>33.13995</td>
<td>0.384</td>
</tr>
<tr>
<td>Engraved</td>
<td>490.0131</td>
<td>141.6278</td>
<td>0.001</td>
</tr>
<tr>
<td>Nb of words</td>
<td>2.651543</td>
<td>0.170860</td>
<td>0.000</td>
</tr>
<tr>
<td>Touch</td>
<td>-37.72552</td>
<td>26.72803</td>
<td>0.158</td>
</tr>
<tr>
<td>Color</td>
<td>-2.691323</td>
<td>26.95307</td>
<td>0.920</td>
</tr>
<tr>
<td>Scarce</td>
<td>-31.94795</td>
<td>87.71657</td>
<td>0.707</td>
</tr>
<tr>
<td>cons</td>
<td>21.12359</td>
<td>16.99545</td>
<td>0.212</td>
</tr>
</tbody>
</table>

« Traditional » variables in art market studies

Variables regarding the description of the artworks

A variable measuring the length of the description of the works, in auction catalogues

Series of dummies, indicating whether or not the description of the works mentions the touch of the artist, deals with the color of the work or, for instance, indicates whether or not the work is scarce.
2. Pricing the Words, in Auction Catalogues

The Need to Analyse Critical Comments:

The hammer price is very correlated with the length of the description in the catalogue but it is independent from some qualitative aspects!

... as if the size of the description was much more important than the concrete meaning of the words!

The quantitative analysis fails to measure the qualitative opinion on the work, that can be found in exhibition or auction catalogues.

As a consequence, no evidence is made between the price of an artwork and the critical comments on the latter, even if it seems obvious that such a link exists.

- Sentiment analysis is now a well-established sub-domain of NLP (natural language processing)
- It could help analyse subjective comments on the different works
- However
  - Few available systems and resources if the target language is not English
  - Sentiment analysis is language-dependent
  - Available systems are domain-dependent and not robust

Our strategy:

- Use existing resources as far as possible
  - Emolex, Emobase (for French)
- (Re)train on a sample of the target corpus
- Check if the result is consistent

- This is still work ongoing
- With this approach, we hope to be able to answer the question: how can one measure opinions on artworks and, more broadly, tastes?