Quantifying the Unknown

How many manuscripts of the marquise de Sévigné still exist?

S. Gabay¹,² L. Rondeau du Noyer M. Gille Levenson³
L. Petkovic² A. Bartz⁴

¹University of Geneva (CH)
²University of Neuchâtel (CH)
³École normale supérieure de Lyon (FR)
⁴École nationale des Chartes (FR)
1 - Madame de Sévigné

Marie de Rabutin-Chantal, marquise de Sévigné 1626, Paris - †1696, Grignan

The most famous letter-writer in French literature.

Her letters are published anonymously after her death.

In the middle of the 18th c. her grand-daughter burns most of the remaining material, but some manuscripts still survive.

Figure – Portrait of Mme de Sévigné, Claude Lefèbvre, Musée Carnavalet (Sévigné). Source : Wikipedia
2 - Public vs private collections

Some manuscripts are kept in private collections

Some manuscripts are kept in public libraries

Figure – Collection of E. Lorient. Source: Librairie Traces écrites, Paris

Figure – Archives Dép. de la Côte-d’Or, titres de familles. Source: dossier Berbisey, E 81
3 - Quantifying the unknown

How many documents of a given author still exist, and among them how many are accessible to scholars?

We use Mme de Sévigné as a test case to calculate how many autograph manuscripts (AM) are still circulating on the market, and therefore assess precisely what is inaccessible because it is held in private collections by combining three different sources of information.

1. A list of French AM held in libraries (created for the occasion)

2. A list of Sévigné’s autograph manuscripts held in historical private collections, drawn from Duchêne’s edition.

3. A list of manuscripts drawn from fixed-price and auction catalogues.
Sale catalogues are published by rare books and manuscripts dealers. It is a primary, yet understudied, source for manuscript history.
5 - Workflow

We have improved our workflow for (semi-)automatic extraction [Gabay et al. 2020]

- transforming the image into structured information
- using international standards
- relying as much as possible on open-source solutions

![Workflow Diagram](image)

**Figure – Workflow**

All entries end up annotated in TEI to allow online publication, data mining and sharing.

<item n="207" xml:id="CAT_000256_e207">
  <num>207</num>
  <name>Sévigné (Marie de Rabutin-Chantal, marquise de)</name>
  <trait>
    <p>la célèbre épistolaire.</p>
  </trait>
  <desc><term>Fin de lettre aut.</term> à sa fille Mme de Grignan; aux Rochers, <date>12 août 1685</date>, <measure type="length" unit="p" n="3">3 p.</measure> <measure type="format" unit="f" n="4">in-4</measure>, suivie de 2 pages aut. d'Emmanuel de Coulanges.</desc>
  <note>Précieuse pièce où elle parle longuement de son séjour aux Rochers, en compagnie d'Emmanuel de Coulanges, et du prochain départ de ce dernier avec Charles de Sévigné pour les Etats de Bretagne. «Mon fils a une petite lanternerie d'émotion qui l'a empêché d'aller aux Etats. Il prend de cette tisane des capucins que vous connaissez, et dont je me suis si bien trouvée; il compte cependant de partir demain avec M. de Coulanges.»</note>
</item>
7 - A manuscript fingerprint

A manuscript (or several fragments of a single manuscript) keeps circulating on the manuscripts market.

<table>
<thead>
<tr>
<th>RDA, May 1894 (N°166)</th>
<th>RDA, July 1897 (N°200)</th>
<th>RDA, April 1902 (N°257)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sévigné</td>
<td>Sévigné</td>
<td>Sévigné</td>
</tr>
<tr>
<td>Fin de lettre aut.</td>
<td>Fin de lettre aut.</td>
<td>Fragment de let. aut.</td>
</tr>
<tr>
<td>12 août 1685</td>
<td>12 août 1685</td>
<td>12 août 1685</td>
</tr>
<tr>
<td>3 p.</td>
<td>3 p.</td>
<td>2 p.</td>
</tr>
<tr>
<td>in-4</td>
<td>in-4</td>
<td>in-4</td>
</tr>
<tr>
<td>200 francs</td>
<td>200 francs</td>
<td>125 francs</td>
</tr>
</tbody>
</table>

**Table – Key information of three sold items**

To reconcile entries or reconstitute manuscripts, we need to compare the items sold. To do so, we propose to use data contained in the `<desc>` of each `<item>`. Similarly to a social number, using our date or our place of birth, we can create a unique ID for each manuscript with the author’s name, the type of document, the writing date, the length, the format and the price.
8 - Reconciliation process

After filtering on a single author, all items found are compared one to another, and a score is offered for each pair.

<table>
<thead>
<tr>
<th>ID</th>
<th>Bonus</th>
<th>Malus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of document</td>
<td>+ 0.2</td>
<td>- 0.1</td>
</tr>
<tr>
<td>Writing date</td>
<td>+ 0.5</td>
<td>- 0.5</td>
</tr>
<tr>
<td>Length</td>
<td>+ 0.1</td>
<td>- 0.1</td>
</tr>
<tr>
<td>Format</td>
<td>+ 0.1</td>
<td>- 0.3</td>
</tr>
<tr>
<td>Price</td>
<td>+ 0.1</td>
<td>- 0.1</td>
</tr>
</tbody>
</table>

Table - Using a malus or a bonus to compare manuscripts

Pairs obtaining a score higher than 0.5 are then reconciliated with a graph: each cluster represents (potentially) a manuscript sold multiple times.
9 - Identifying manuscripts

<table>
<thead>
<tr>
<th>Princeton, Rare Books and Special Collections, C0710, vol. 4, f°57/N°4</th>
<th>Paris, BNF, NAF 717</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sévigné</td>
<td>Sévigné</td>
</tr>
<tr>
<td>Fin de lettre aut.</td>
<td>Fragment de let. aut.</td>
</tr>
<tr>
<td>12 août 1685</td>
<td>12 août 1685</td>
</tr>
<tr>
<td>3 p.</td>
<td>2 p.</td>
</tr>
<tr>
<td>in-4</td>
<td>in-4</td>
</tr>
</tbody>
</table>

Table – Key information of two manuscripts

Not only we can reconcile sold entries between themselves, but also these entries with library collections.
Results

We can now offer some results:

▶ 63 sales have been identified until 1903
▶ 46 AM being sold at least one time, 14 at least two times
▶ 13 letters out of the 46 sold are not in public libraries or archives today

Following these numbers, we can say that:

▶ c. 1% of the 1,350 letters identified by R. Duchêne are still circulating on the market.
▶ c. 5% of the total has survived but is inaccessible to scholars, if we add the 62 letters still held in the private collections of the Guitaut family in Burgundy.
One more thing

Due to the COVID pandemic, we are still waiting for some PDFs ordered at the library (in October 2019...). Therefore, the results obtained computationally had to be supplemented by traditional research in library. As soon as all the documents will be received, we will complete the study. In the meantime, GROBID models, XSLT stylesheets for the corpus creation and python scripts for extraction and reconciliation can be found online:

https://github.com/katabase/

Merci beaucoup à
- B. Joyeux-Prunel (UniGE)
- Th. Clérice, J.-B. Camps and C. Corbières (ENC)
- M. Khemakhem and L. Romary (INRIA)
- A. et E. Decaster
- S. Fernandez, R. Grünblatt et P. Ianetta
- A. Carnoy (bon anniversaire!)