

Quantifying the Unknown: How many manuscripts of the marquise de Sévigné still exist?

Simon Gabay, Lucie Rondeau Du Noyer, Matthias Gille Levenson, Ljudmila Petkovic, Alexandre Bartz

▶ **To cite this version:**

Simon Gabay, Lucie Rondeau Du Noyer, Matthias Gille Levenson, Ljudmila Petkovic, Alexandre Bartz. Quantifying the Unknown: How many manuscripts of the marquise de Sévigné still exist?. Digital Humanities DH2020, ADHO, Jul 2020, Ottawa, Canada. hal-02898929

HAL Id: hal-02898929
https://hal.science/hal-02898929

Submitted on 14 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Quantifying the Unknown
How many manuscripts of the marquise de Sévigné still exist?

Simon Gabay[1], Lucie Rondeau du Noyer[2], Matthias Gille Levenson[3], Ljudmila Petkovic[1], and Alexandre Bartz[4]

[1]Universités de Neuchâtel et Genève (Suisse)
[2]Independent scholar
[3]École normale supérieure de Lyon
[4]École nationale des Chartes

Abstract

Because manuscripts are lost, burned, torn apart or thrown away, it is as complex as crucial to know how many of them still exist for any philologist preparing an edition. Thanks to a (semi-)automatic and fully-open source workflow, we have extracted, structured and annotated hundreds of manuscript sale catalogues published in 19th c. Paris. The obtained level of granularity allows us not only to reconcile different sales of a single item sold multiple times, but also to identify if the manuscript is now kept in a library. Using Sévigné as a test case, we were able to calculate that c. 1% of her manuscripts still has to be found because they are circulating on the private market. All the data we produced remain available for similar research on other authors.

Manuscripts can be burned, lost, forgotten, thrown away... If scholars have already tried to measure the proportion that has survived since the apparition of moveable types with Gutenberg [Weitzman, 1987], such percentages do not help editors of texts to answer a more practical question: how many documents of a given author still exist, and among them how many are accessible to scholars?

In the present paper, we want to use Madame de Sévigné (1626-†1696) as a test case to calculate how many autograph manuscripts (AM) are still circulating on the market (M), and therefore assess precisely what is inaccessible because it is held in private collections (C) by combining three different sources of information. First, a list of French AM held in libraries – which has been created for the occasion, because there is no catalogue for French literature such as the one by P. Beal [2005]. Second, a list of Sévigné's AM held in historical private collections, drawn from Duchêne's edition [Sévigné, 1972-1978]. Third, a list of manuscripts drawn from fixed-price and auction catalogues [Bodin, 2000], which contains the description of hundreds of thousands of manuscripts sold over decades (fig. 1).

1 Quantification

Existing manuscripts (E) of an author (a) are either kept accessible in libraries and archives (L), either (more or less) hidden in private collections (C). [1]

$$E_a = L_a \cup C_a \tag{1}$$

The problem is that while we know L, for which we have catalogues, C is unknown. To know what it represents, we can divide it in two: on the one hand we have historical collections (H), usually inherited by old families, well documented and extremely static, and on the other hand there is an unknown amount of documents circulating between private collectors (P).

$$C_a = H_a \cup P_a \tag{2}$$

If we cannot know P, we can use a proxy: we can deduce what is still on the market (M) by subtracting what is owned (because it has been bought) by libraries (L) from everything that has been sold (S).

$$M_a = S_a \backslash L_a \tag{3}$$

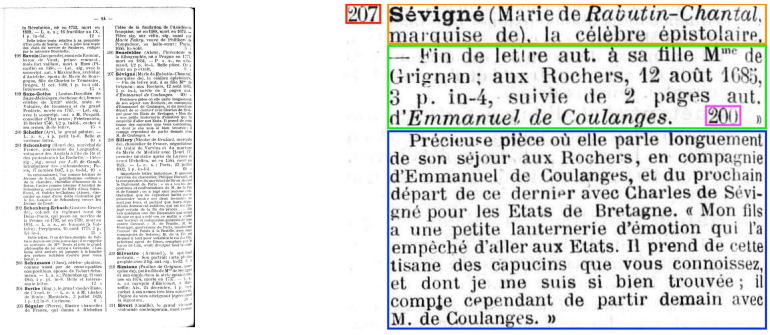
Because buyers are constantly intervening on the market (S), any value is only true at time t – *i.e.* the date of the last sale catalogue taken into account, assuming that all the previous ones have been analysed.

$$C_a(t) \approx H_a(t) \cup M_a(t) \tag{4}$$

With all these information, we can now deduce how many manuscripts still exist (E) if we know S.

$$E_a(t) \approx L_a(t) \sqcup H_a(t) \sqcup M_a(t) \approx L_a(t) \cup H_a(t) \cup S_a(t) \tag{5}$$

2 Extraction and structuring



207 **Sévigné** (Marie de *Rabutin-Chantal*, marquise de), la célèbre épistolaire. — Fin de lettre aut. à sa fille Mme de Grignan ; aux Rochers, 12 août 1685, 3 p. in-4, suivie de 2 pages aut. d'*Emmanuel de Coulanges*. 200 »

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 connoissez, et dont je me suis si bien trouvée ; il compte cependant de partir demain avec M. de Coulanges. »

Figure 1: *RDA*, May 1894 (N°166)

[1] Private collections kept in libraries (*e.g.* Morgan, Bodmer. . .) are counted in L.

Looking for French AM, we have concentrated our efforts on documents sold in Paris, and for financial and time reasons, we have focused on catalogues published before c. 1900. We have retro-converted:

- 250 fixed-price catalogues of the *Revue des autographes* (*RDA*, cf. fig. 1).
- 300 fixed-price catalogues of the *Lettres autographes et documents* (*LAD*).
- 100 auction catalogues.

Because of similarities between such catalogues and dictionaries, we have been able to use *GROBID dictionaries* [Khemakhem et al., 2018] to process the images and transform them into a fully TEI-conformant semantic encoding (fig. 2) thanks to a custom workflow [Gabay et al., 2020].

```
<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>Fin de lettre aut. à sa fille Mme de Grignan; aux Rochers,
       12 août 1685, 3 p. in-4, suivie de 2 pages aut. d'Emmanuel de
       Coulanges. 200 »</desc>
  <note>Précieuse pièce où elle parle longuement de son séjour aux
       Rochers, en compagnie d'Emmanuel de Coulanges, et du prochin
       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 connoissez, et dont je me suis si bien trouvée;
       il compte cependant de partir demain avec M. de Coulanges.»</note>
</item>
```

Figure 2: XML-TEI encoding of an entry

The workflow keeps undergoing constant improvements (*e.g.* Rondeau Du Noyer et al. [2019]), which have led to the creation of a dedicated tool for catalogues [Khemakhem et al., 2020]. In its last version, on top of traditional features for information extraction (special characters, position on the page. . . in red in fig. 3), we now use typographical information (bold, italics, size of the font. . . in blue in fig. 3) for more precise results.

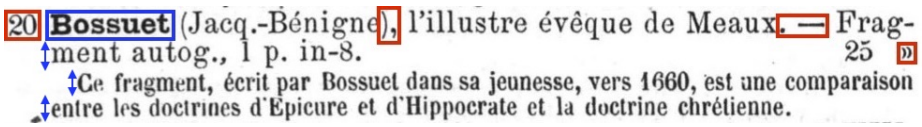
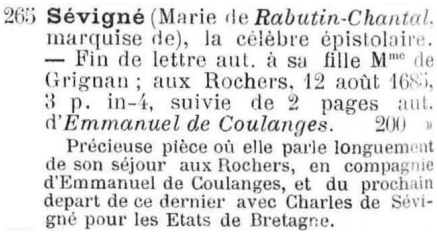


Figure 3: *RDA*, May 1873 (N°37)

3 Annotation and reconciliation

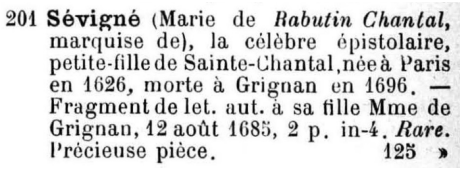
The letter previously mentioned is not the only one of Sévigné sold during the 19th c., and it has not been sold only once:



265 **Sévigné** (Marie de *Rabutin-Chantal*, marquise de), la célèbre épistolaire. — Fin de lettre aut. à sa fille Mme de Grignan ; aux Rochers, 12 août 1685, 3 p. in-4, suivie de 2 pages aut. d'*Emmanuel de Coulanges*. 200 »

Précieuse pièce où elle parle longuement de son séjour aux Rochers, en compagnie d'Emmanuel de Coulanges, et du prochain depart de ce dernier avec Charles de Sévigné pour les Etats de Bretagne.

Figure 4: RDA, July 1897 (N°200)



201 **Sévigné** (Marie de *Rabutin Chantal*, marquise de), la célèbre épistolaire, petite-fille de Sainte-Chantal, née à Paris en 1626, morte à Grignan en 1696. — Fragment de let. aut. à sa fille Mme de Grignan, 12 août 1685, 2 p. in-4. *Rare.* Précieuse pièce. 125 »

Figure 5: RDA, April 1902 (N°257)

Because the same item can be sold multiple times, it is crucial to transform the list obtained with the digitisation of sale catalogues into a set of unique types (or classes, cf. blue and red boxes in fig. 6), prior to comparing these types with existing documents held in libraries. Doing so, we can identify AM that have never appeared on the market (in pink and in black), document the history of those that are now in library collections (in blue and in orange) or identify "ghost" manuscripts that are still circulating on the private market (in green and red).

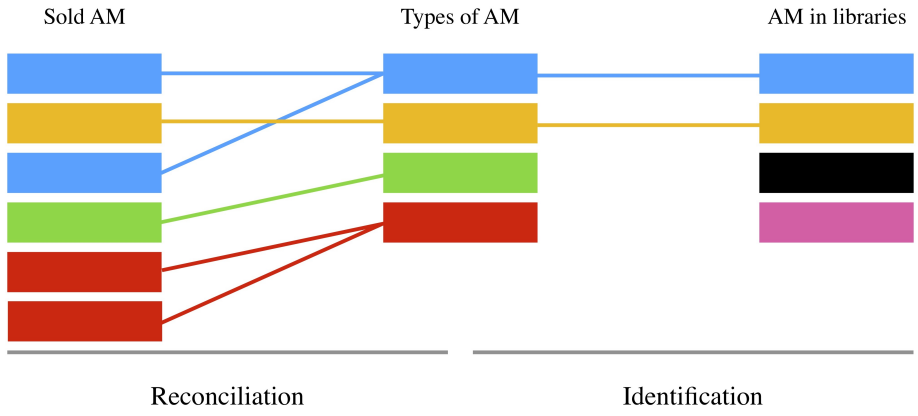


Figure 6: Reconciliation-identification process

To carry out this task, more information is required than those provided by *GROBID dictionaries*, we have therefore added an extra layer of information, including the type of document (*L.a.s.* for autograph letter signed, *D.s.* signed document. . .), its length (number of pages or folios. . .), its format (in-octavo, in-quarto. . .), its date or its price. Since these information follow either an extremely strict (*1 p.*, *L.a.s.*. . .), either a fairly common pattern (*12 janvier 1798*, *19 sep. 1820*. . .), they are tagged with regexes and dedicated python libraries in order to obtain a more fine-grained encoding:

```
<desc><term>Fin de lettre aut.</term> à sa fille Mme de Grignan; au
    Rochers, <date>12 août 1685</date>, <measure type="length">3 p.
    </measure><measure type="format">in-4</measure>, suivie de 2 pages
    aut. d'Emmanuel de Coulanges. <measure type="price">200</measure>
»</desc>
```

Figure 7: Annotated <desc>

Combined with the name of the author, such information provide a unique combination of features that can be used to compare sold documents over time, and identify not only same AM sold twice, but also different fragments of a single manuscript (tab. 1), which share part of the information only (same date, same format but different length).

RDA, May 1894 (N°166)	RDA, July 1897 (N°200)	RDA, April 1902 (N°257)
Sévigné	Sévigné	Sévigné
Fin de lettre aut.	Fin de lettre aut.	Fragment de let. aut.
12 août 1685	12 août 1685	12 août 1685
3 p.	3 p.	2 p.
in-4	in-4	in-4
200 francs	200 francs	125 francs

Table 1: Key information of three sold items from catalogues

Because we have catalogued all the known manuscripts of the marquise de Sévigné after extensive research in European and American libraries, it is possible to reconstitute part of their history thanks to the sale catalogues.

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

Table 2: Key information of two manuscripts

4 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

Following these numbers, we can say that:

- c. 1% of the 1,350 letters identified by Duchêne [Sévigné, 1972-1978] 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.

Such numbers, obviously, need to be taken with caution for two main reasons. On the one hand, the oldest catalogues are not precise enough to identify exactly which AM is sold. On the other hand, the market in the 19th c. is already international, and manuscripts sold outside of France are not taken into account by our study. This second

problem should receive all our attention in a near future to contribute to the history of objects [Courtin, 2020], and especially the migration of manuscripts [Burrows et al., 2019].

Acknowledgements

Many thanks to Agathe Decaster for her (crucial) help with the mathematical formulas and her brother Erwan.

References

P. Beal. Catalogue of English Literary Manuscripts 1450–1700, 2005. URL https://celm-ms.org.uk/. https://celm-ms.org.uk.

T. Bodin. Les grandes collections de manuscrits littéraires. In A. Charon and E. Parinet, editors, *Les Ventes de livres et leurs catalogues: XVIIe-XXe siècle*, number 5 in Études et rencontres, pages 169–190. École des chartes, Paris, 2000.

T. Burrows, E. Hyvönen, L. Ransom, and H. Wijsman. Mapping manuscript migrations: Digging into data for the history and provenance of medieval and renaissance manuscripts. *Manuscript Studies*, 3(1), May 2019. URL https://repository.upenn.edu/mss_sims/vol3/iss1/13.

A. Courtin. Retour sur la datavisualisation « Sur la pistes des ventes d'antiques ». *Numérique et recherche en histoire de l'art*, 2020. URL https://numrha.hypotheses.org/743.

S. Gabay, L. Rondeau Du Noyer, and M. Khemakhem. Selling autograph manuscripts in 19th c. Paris: digitising the Revue des Autographes. In *Atti del IX Convegno Annuale AIUCD. La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica*, Quaderni di Umanistica Digitale, pages 113–118, Milan, Italy, 2020. Associazione per l'Informatica Umanistica e la Cultura Digitale. https://hal.archives-ouvertes.fr/hal-02388407.

M. Khemakhem, A. Herold, and L. Romary. Enhancing Usability for Automatically Structuring Digitised Dictionaries. In *GLOBALEX workshop at LREC 2018*, Miyazaki, Japan, May 2018. URL https://hal.archives-ouvertes.fr/hal-01708137.

M. Khemakhem, S. Gabay, B. Joyeux-Prunel, L. Romary, L. Saint-Raymond, and L. Rondeau Du Noyer. Information Extraction Workflow for Digitised Entry-based Documents. In *DARIAH Annual event 2020*, Zagreb, Croatia, 2020. Digital Research Infrastructure for the Arts and Humanities. URL https://hal.archives-ouvertes.fr/hal-02508549.

L. Rondeau Du Noyer, S. Gabay, M. Khemakhem, and L. Romary. Scaling up Automatic Structuring of Manuscript Sales Catalogues. In *What is text, really? TEI and beyond*, Graz, Austria, Sept. 2019. URL https://hal.inria.fr/hal-02272962. https://hal.inria.fr/hal-02272962.

Sévigné. *Correspondance*. Gallimard, Paris, 1972-1978. ed. by R. Duchêne.

M. P. Weitzman. The Evolution of Manuscript Traditions. *Journal of the Royal Statistical Society. Series A (General)*, 150(4):287–308, 1987. ISSN 0035-9238. doi: 10.2307/2982040. https://www.jstor.org/stable/2982040.