UNIVERSITÄT
BERN

Faculty of Business, Economics
and Social Sciences

Department of Economics

Quantifying Trade from Renaissance Merchant Letters

Fabio Gatti

24-03

March, 2024

DISCUSSION PAPERS

Schanzeneckstrasse 1
CH-3012 Bern, Switzerland
http://www.vwi.unibe.ch

Quantifying Trade from Renaissance Merchant Letters*

Fabio Gatti†

March 2024

Abstract

Medieval and Early-Modern business correspondence between European companies constitutes a rich source of economic, business, and trade information in that the writing of letters was the very instrument through which merchants ordered and organized the shipments of goods, and performed financial operations. While a comprehensive analysis of such material enables scholars to re-construct the supply chains and sales of various goods, as well as identify the trading networks in the Europe, much of the archival sources have not undergone any systematic and quantitative analysis. In this paper we develop a new holistic and quantitative approach for analysing the entire outgoing, and so far unexploited, correspondence of a major Renaissance merchant-bank - the *Saminiati & Guasconi* company of Florence - for the first years of its activity. After digitization of the letters, we employ an AI-based HTR model on the Transkribus platform and perform an automated-text analysis over the HTR-model's output. For each letter (6,376 epistles) this results in the identification of the addressee (446 merchants), their place of residence (65 towns), and the traded goods (27 main goods). The approach developed arguably provides a best-practice methodology for the quantitative treatment of medieval and early-modern merchant letters and the use of the derived historical text as data.

*Preliminary draft, please do not quote without permission of the authors.

We are thankful to Tiziana Dassi and the staff of the Bocconi University Historical Archives, for their active and indispensable support in researching and digitizing the historical materials employed in this article.

†Department of Economics, University of Bern

1 Introduction

Towards the end of the middle-ages economic growth resulted in an increased demand for goods in Europe (Malanima, 1983). To satisfy such demand, new trade roots were opened and a new class of people, the merchants, emerged to act as intermediaries between communities and cultures, allowing for the exchange of products and information. Until the opening of the Atlantic routes and the development of colonial economies, Italian merchant-bankers constituted the main actors, connecting the major European commercial hubs through transportation and financial services. Through the sending and receiving of letters, early-modern merchants organized the shipment and the transportation of goods, designing commercial links and trade networks which covered all the European Continent, comprising the supply chains and sales of a large variety of products. Arguably a systematic analysis of this mercantile correspondence could provide scholars detailed insight into the span and depth of early-modern merchant networks, as well as identify how different goods were being produced to satisfy international and interregional demand during this period. However, existing studies on mercantile correspondence have been mostly qualitative and have only focused on a limited amount of letters. More precisely, previously scholars read through a self-selected sample of the letters and tried to compile some useful information in order to construct a cohesive tale. As argued by Gaffield and Baskerville (1985), this approach of economic historical research of "find(ing) and choos(ing) historical sources for examination" can be considered "quite unsystematic".[1] To address this, the current paper develops a best-practice methodology for the quantitative treatment of medieval and early-modern merchant letters using the correspondence of a major Renaissance merchant-bank - the *Saminiati & Guasconi* company of Florence - for the first years of its activity.

Scholars interested in early-modern merchant letters have thus far focused on different aspects. For example, Bartolomei et al. (2018) analyses communication features in the let-

[1] Already in the 1980s scholars imagined a future in which the source of historical research would have been an "automated archivist"(Gaffield and Baskerville, 1985, p.167, 170-172) or an "artificially intelligent database" (Gaffield and Baskerville, 1985, p.170) to be queried and analyzed by the researcher. As stated by Fernand Braduel in a speech stressing the importance of the use of archival sources, and especially of commercial letters and registers in assessing and analysing the economic world of the *Ancient Regime*: "one day, no doubt, computer technology will overtake our old historian's profession, at least in these methods" (Spallanzani, 1969, introduction,p.6).

ters received by two prominent early-modern French corporations, while Galliano (2018), focusing as well on communication aspects, describes the "language of trust" emerging from the epistles of a Genoese merchant in mid-XVI century. The correspondence of the Salviati's, a major merchant-bank in Lyon, is the object of a paper by Matrigne (2021), which, based on the rich literature on the bills of exchange (DaSilva, 1969, DeRoover, 1944, 1953, Mandich, 1953), focuses on the use of business letters to offset credit risk in international transactions. Other works tried to reconstruct some particular tale out of a set of letters. For instance, Maréchaux analyses the organisation of Genoese naval entrepreneurs, using merchant letters to reconstruct the network of agents involved in the galleys fleet management(Maréchaux, 2023). Also Marsilio (2005) exploited the correspondence of a Genoese merchant: Paolo Gerolamo Pallavicini, classifying his correspondents by town of settlement and most traded goods, while Tucci (1957) published 300 letters by Andrea Barengo, a Venetian merchant involved in the East-Mediterranean trade in mid XVI century. Letters of English businessman were the object of the works of Price (1979), who published a set of some hundred letters sent in the 1770s by Hoshua Johnson, a major Londoner merchant, and Guttridge (1934), who was mostly interested in the political content of the letters sent by the Bristol merchant Richard Champions. A more comprehensive study has been done by Martin (1965), which published almost 500 merchant letters from the rich Historical Archive of Valladolid in Spain. Also, Martin (1965) classified by town almost 3,500 letters sent by Spanish merchants to Italian places, with a focus on those toward/from Florence, and used the epistolary information to gain insight into workings of the Mediterranean economy in late XVI century.

The contribution of this paper to the literature on the analysis of merchant letters is twofold. Firstly, we put light on as of yet essentially unexploited material by processing the complete outgoing correspondence (6,376 letters) from the unpublished *Copialettere* of the first seven years of activity of the *Saminiati and Guasconi* merchant bank, preserved in the historical archive of Bocconi University. The *Saminiati and Guasconi* merchant bank was founded in December 1626, appropriating the structures and activities of an already well established Tornaquinci's merchant-bank. Importantly, because of its takeover of the Tornaquinci's bank, the *Saminiati and Guasconi* was from the start one of the biggest companies in Florence (Cipolla, 1988) and well embedded in a network of hundreds of Italian

merchant-bankers, which, settled in the main commercial hubs of Europe, connected the economies of the Continent though financial and payment operation, and the transport of precious goods and textiles. Notably, the archival sources related to the *Saminiati*'s merchant-banking activity have been an object of only a few studies (DeMaddalena, 1959, Saba, 2019, S.Groppi, 1990), and solely for the purpose of presenting to a broader audience the Bocconi Historical archive and with no economic analysis of its content. In this regard, the only use of the letters used in this paper was for the extraction of some quotation by Groppi to present the archive in his *Inventario* (S.Groppi, 1990).

Our second contribution concerns our use of the *Saminiati and Guasconi* merchant bank letters to develop a comprehensive and systematic approach, exploiting digitization, AI for handwriting recognition, and text analysis techniques, to create a database that allows the classification and quantification of the trade in which the merchant-company was involved in. This rigorous quantitative analysis of the letters allows us to understand, for example, how much Florentine silk cloths were sought outside of Italy, how Genoa merchants were actually involved in the trade of precious-metal coins, or how the Low Countries were the European hot-place for diamonds in XVII century, thus confirming what has been stated in previous works on commodities and goods trade in early-modern Europe (Barrett, 1990, Hofmeester, 2019, L.Pezzolo and G.Tattara, 2008, Malanima, 1983, Morelli, 1976, Sabel, 2019).

The remainder of the paper is organized as follows. In Section 2 we highlight the importance of early-modern business correspondence as an instrument to carry out economic history analysis, and we describe the structures and the activities of the *Saminiati & Guasconi* merchant-bank. In Section 3 we outline the consistency of the Saminiati-Pazzi Archive of Bocconi University, especially in being a store of early-modern correspondence material. Section 4 illustrates the digitization process we carried out on the archival sources, and the steps for developing an AI-based HTR model for the correspondence transcription. Details of the data formatting and correction procedures needed to make the model output suitable for a text analysis, as well as the subsequent building of an R code for the counting of relevant keys-words, are provided in Section 5. In Section 6 we provide descriptive statistics of the constructed data set. Section 7 concludes.

2 The *Saminiati & Guasconi* Merchant-Bank: Structures and Activites

Successful Renaissance Florentine merchant-companies were firmly embedded in international networks (Bratchel, 1990, Caracausi and Jeggle, 2014, Kohn, 1999, Mauro, 1990, Subrahamanyam, 1996) and composed mainly of other Italian merchants positioned in the main commercial and financial centers of Europe. The maintenance of these relationships, and the consequent exchange of goods, money and information, required the assiduous writing and receiving of letters, such correspondence constituting also a private source for "reserved" information (McCusker, 2005), and provides a competitive advantage to the best-structured mercantile organisations. *Saminiati & Guasconi*, the merchant bank of focus in this paper, was well embedded in this Italian-speaking network of merchants moving raw and laboured silks and wools from and to the Peninsula, and providing banking and payment services to each corner of the Continent (Saba, 2019)[2]

The *Saminiati & Guasconi*'s Merchant-Bank (i.e., the *Ascanio Saminiati, Giovacchino Guasconi e comp. di Banco, di Firenze*) was founded in Florence in 1626, appropriating the structures and the activities of a previous bank, the *Tornaquinci*'s, in which Ascanio Saminiati and Giovacchino Guasconi were minority partners and workers at the orders of the majority shareholders Luca and Matteo Tornaquinci and Andrea Bettini (S.Groppi, 1990). Since the beginning of its activity, in December 1626, the company was twinned by a Venetian branch (the *Casa di Venezia*) thanks to the co-participation in the corporate structure of prominent personalities from the lagoon, such as Alberto Gozzi and, since 1628, his uncle, Domenico Tironi Dalla Sedda (S.Groppi, 1990). In 1634, the *Compagnia* opened also a branch in Messina, willing to intensify the traffic of silk from the island, the transport of valuable merchandise being one of the main businesses of the company, together with operations of international banking and payments, and its participation in the International Fairs of Exchange (S.Groppi, 1990). The profitability of the bank came from the commission it charged on its economic counterparts for the provisions of various services, namely the 0.33% of commission for the *star del credere* over a bill of exchange, the 0.4% as commission for the trading of a bill of exchange and the 2% for the trade of

[2]See Saba(2019), p.129-137 for the list of the places inserted in the Saminiati's merchants network.

commodities or valuable goods (S.Groppi, 1990).

The Company was ruled according to established structures and customs well-specified in the founding contract, and some variation in the ownership composition was by no means unusual for banking companies of the time (S.Groppi, 1990). Through a well-oiled mechanism youths could start as trainees in such companies, to then become minority shareholders and at last taking the place of the elderly merchants (S.Groppi, 1990). These youngsters were usually linked to the main partners by family ties and were often employed in the most humblest and menial jobs, such as copying the letters to be sent by post (S.Groppi, 1990), thus creating exactly that resource which is used in this paper as the center of the analysis.

In Florence of the 1620s and 1630s there was certainly no shortage of competition to a merchant bankers' company, but the *Saminiati & Guasconi*'s appears to have been particularly sound and active since the very year of its foundation. As a matter of fact, using the postal charges to proxy the dimension of different merchant bankers active in the towns, the *Saminiati*'s was the fifth biggest in town for the period July 1627-June 1628, the fourth between June 1628 and June 1631 and the third from July 1631 to June 1632 (Cipolla, 1988). The relevance and socio-economic dimension of such *Compagnia*, as well as the status of its main holders, are reflected by, among other things, the election of the Gozzi as Senator for the *Serenissima* Republic of Venice in 1646, while the two main partners, Ascanio Saminiati and Nicolo Guasconi became senators in Florence in 1653 and 1657, respectively (S.Groppi, 1990).

3 Our Archivial Sources

3.1 The Saminiati-Pazzi Archive of Bocconi university

The Saminiati-Pazzi Archive of Bocconi University (https://asboc.unibocconi.it/oggetti/34-archivio-saminiati-pazzi/) is a collection of documents of various kind (registers, letters, balances, inventories, newspapers..) concerning the business and economic life of two Tuscan families of early-modern merchants. The bulk of the documentation consists of more than a thousand registers regarding the administration of different Tuscan *latifondi*, as well as the mercantile activities that the Saminiati and the Pazzi carried out in partnerships

with other Tuscan families, such as the Guasconi and the Strozzi, and a very rich set of banking and mercantile letters dated from 1624 to 1719 (DeMaddalena, 1959). The documentary holdings were acquired by Bocconi University in 1937 from the private archives of such Tuscan families and were re-ordered by Aldo De Maddalena (DeMaddalena, 1959), who in 1959 presented the Archive in an article on a french economics journal (S.Groppi, 1990). The Saminiati-Pazzi Archive was then declared of notable historical interest on June 16, 1997, and since April 2014 its charts and registers are stored in the new historical fonds conservation room at the Bocconi University Library. Still as of today the incredibly rich documentation of the archive is mostly unexploited, and other than the *L'archivio Saminiati-Pazzi* of Sergio Groppi and the introductory work of De Maddalena, few other scholars had studied the *Saminiati*'s charts and cited it in their articles (DaSilva, 1969, McCusker and Gravesteijn, 1991, Saba, 2019).

3.2 The *Copialettere* of the *Saminiati & Guasconi* Bank

The use of early-modern letters as a potential source of data has already been recognized (Camiciotti, 2014). However, differently to the analysis of static and end-period accounting documents, such as the *libro mastro*, the analysis of the economic content of the correspondence provides the scholar with almost day-by-day information about the business conduct of the mercantile organization. In this regard, a *Copialettere* is a book in which the early-modern merchant transcribed all the outgoing correspondence, in order to be able to then re-check dates, quantities and figures of the deals in which he was involved. The *Copialettere*[3] of the *Saminiati*s are books of roughly 38 x 27 cm per 300/400 pages each, bound in an elegant parchment. After the parchment cover, the books start with a page containing an invocation to God and to the Saints[4] and a drawing of the company logo. Subsequently, the pages of the books contain the transcriptions of the letters sent by post by the company.

The side of each page contains *grosso modo* the transcription of three average-sized letters, separated by a horizontal line. Some letters, especially those addressed to important partners and concerning large deals, extended over more pages, usually starting with the

[3]Saminiati-Pazzi Historical Archive

[4]From the Register n. 195: *Nel nome di Dio, della Gloriosa Vergine Maria, di San Giovanni Batista nostro Protettore, di tutti i Santi e Sante del Paradiso Ammen*

caption "*segue*", i.e., "follows". In contrast, letters which only acknowledged the receipt or shipment of a good may only consist of few lines of texts, requiring 6-7 of such letters to fill an entire page.

We here focus our attention on 6 *Copialettere*, containing all the outgoing mercantile correspondence of the *Saminiati & Guasconi's* in its first years of activity starting in December 1626 (S.Groppi, 1990). One should note, however, that the analysis could be potentially scaled up to subsequent years to produce a textual analysis of all the merchant-bank's correspondence stored in the Bocconi's archive.

The documents we used are paired temporally as follows. *Register 192* and *Register 193* cover more or less the same time span from January 1627 (plus a few letters dated December 1626) to August 1628. These *Copialettere* were held separately only for practical reasons, with *Copialettere 193* containing more transcriptions of the letters toward Venice and Rome. *Register 194* is the temporal continuation of 192, while *Register 195* is the sequel of 193 (hence containing the letters toward Venetian and Roman partners), and both roughly span from August 1628 to September 1632. The last two Registers, 196 and 197, are respectively the temporal sequel of the 194 and 195 registers and cover the time span from September 1632 to April 1634 (*Register 196)* and February 1634 (*Register 197).*

4 From the handwritten original document to .txt format

4.1 Digitization of archival materials

The first step of our analysis was to check the actual consistency of the original materials stored in the Bocconi's archives. In this regard, we were able to confirm that all pages were present and intact, the writing well distinguishable from the background, and in general no evidence of damage or stains was present on the sheets. This importantly poses a prerequisite for the *Copialettere* to successfully undergo the OCR process, since machine-made decipherability is strongly dependent on the regularity, clearness and sharpness of the text (Hamad and Kaya, 2016). We then asked the professional archivist of Bocconi University to produce for us high-resolution scans of the material.[5] The 6 *Copialettere*

[5]For a document to be worked by an OCR, images taken by a camera or by any hand-held device has the limitation of producing a "less fixed image" and potentially skewed text lines from the orientation took by the camera (Hamad and Kaya, 2016).

were scanned in high quality, producing pdf files in which each page contains scans of the two adjacent pages. A representative example of a two-page scan of Register 193 is shown in Figure 1. As can be seen, characters, words and lines of text are easily distinguishable and readable, and each letter is separated by a horizontal line from the others.

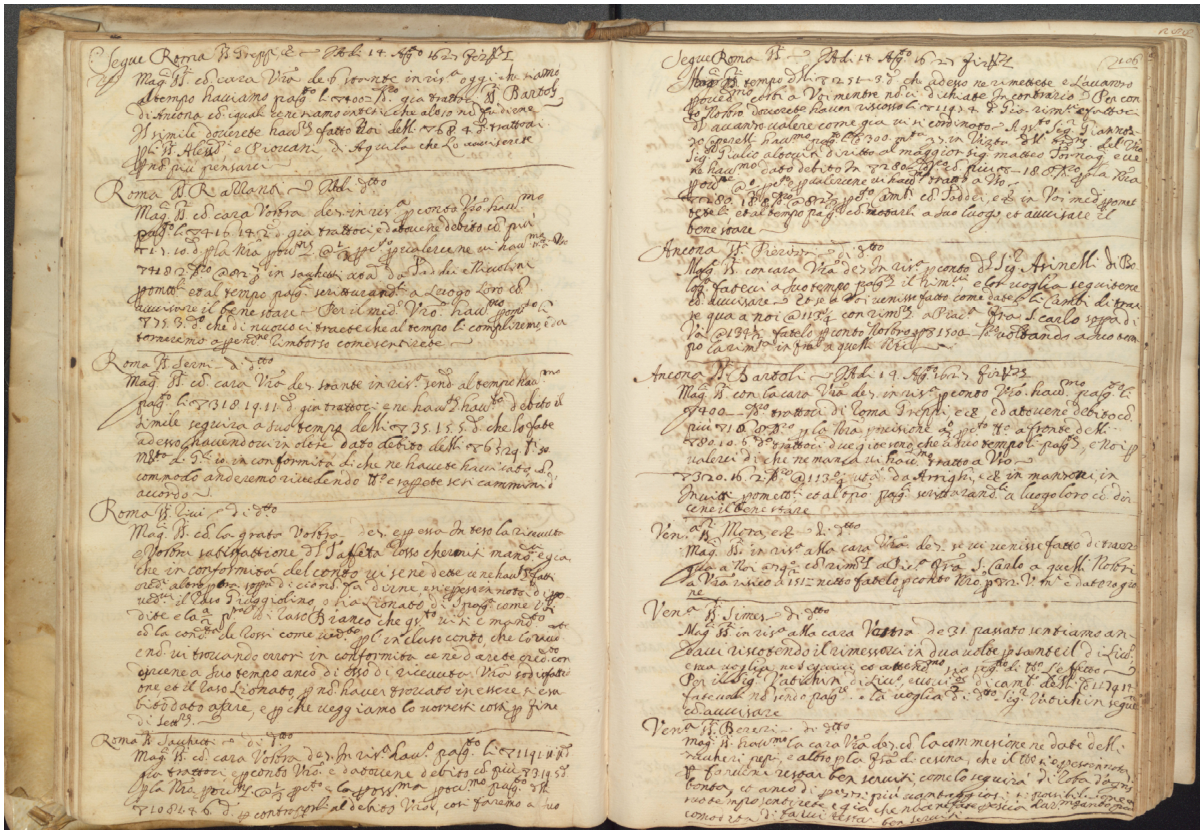

Figure 1: Example of scan from Register 193

4.2 An AI model for the *Saminiati's* ortoghtaphy: the Transkribus Platform

After the digitization of the material, the next step consisted of the application of an Handwritten Text Recognition (HTR) technology on the documents to produce a searchable text for computer-made text analysis. To do so we uploaded the scans on Transkribus, the most popular user-facing platform for producing transcripts of historical texts (Nockels et al., 2022). More specifically, Transkribus is a platform run by a Cooperative (READ-COOP) comprising hundreds of institutions, which employs machine learning technology in order to decipher the writing of a given scripter. One should note that such platforms

have already been used to transcribe early-modern manuscripts in different languages and scripts by researchers in the area of digital humanities (Burlacu and Rabus, 2021, Massot et al., 2019, Schlagdenhauffen, 2020).

The learning component of the textual analysis is based on a training dataset and results in the production of a neural network (an HTR model) targeted at the ortography of the writer (Schlagdenhauffen, 2020). For the specific project object of this research, the Cooperative generously granted us 5,000 credits, allowing us to apply the AI tool we developed on up to 5,000 pages of documents. After the uploading of the documents, the first step consisted of conducting a layout analysis over the pages using the layout analysis tool (Gila, 2023). This tool works well in recognizing text regions and lines so that, for our case, the output only needed minor corrections. One should note that the layout analysis step is crucial since the employed technology follows a "line-oriented approach where the image of a baseline (a horizontal line running underneath a line of text in a digitised image) and the corresponding correctly transcribed text represent the input for the learning algorithms of neural networks" (Muehlberger et al., 2018, p. 958).

The next step of the training of the AI-HTR model entailed feeding the platform with hand-made transcriptions of the original documents, in order to allow the Neural Network to learn to which character each sign in the *Saminiati's* orthography corresponded. Since Transkribus analyses the lines of text character by character, in order to ensure an optimal learning process it is reccommended to "scrupulously respect the letter of the manuscript" (Massot, Sforzini, and Ventresque, 2019, p.10). In our case we reported the transcriptions of early-modern business-Italian abbreviations as such. For example, we replaced *q.sto* for *questo* (this) and *vro* for *vostro* (your).

It is generally accepted that a training dataset of 15,000 transcribed words (around 75 pages) is "sufficient for training an HTR engine to recognize text written in one hand" (Muehlberger et al., 2018, p.959). For our documents it was possible to distinguish three different main spelling styles, although it is hard to say whether they belonged to different authors or just variations of the same hand. However, it seems likely that at least two of these orthographies belonged to different people because of their significant differences. We trained different models, increasing each time the number of hand-made transcriptions until we could not detect any significant difference in the precision of the trained

HTR-model in the transcription. In this regard, we progressively trained 12 models, differentiated by the quantity of hand-written transcription that we fed into them, and the first few being highly imprecise in the transcriptions and having a high Character Error Rate (CER) both for the training and the validation sets. For example, the first model we trained with just one page of transcription had a CER of 11.50% on the training and 19.40% on the validation set. After providing 17,599 words the model obtained a CER of 1.71% on the training set and of 5.20%. However, after this point the number of words we fed into the model did not seem to have a significant impact on the precision of transcriptions. Indeed, the "final_model", the one we used to produce the actual transcriptions, was fed by 63,014 words, i.e., 119 (mostly) double pages, and had only a CER of 1.00% for the training and 4.90% for the validation set. One should note that the precision of the model is measured on the number of characters transcribed correctly. In our case, we trained the model to recognize 110 different characters, namely the 23 characters of the latin alphabet (both in lower and in upper case), the 9 numerical digits, as well as some "special characters" (resembling some from the coptic and greek alphabets) other than commas, punctuation and algebraic operators, used by the *Saminiati & Guasconi* to denominate prices, quantities and currencies. In Figure 2 we provide a screenshot of the HTR model from Transkribus used. One should note that the error rate can also be assessed relative to the public models available on the Transkribus platform. More precisely, an HTR which can produce automated transcripts of handwritten material with a CER of below 5 per cent (meaning that 95 per cent of the characters are correct) is considered good (Burlacu and Rabus, 2021, Muehlberger et al., 2018), and our model clearly passes this hurdle.

5 Classifying the letters: from .txt files to a dataset

5.1 Data Formatting

After having applied the AI-HTR model we trained on Transkribus on the entire corpus of documents, and then exported the output in terms of 6 .txt files (6 different files for the 6 different *Copialettere*). Each .txt files consists of texts in which each line represents the transcription of a line in the original document. An empty line separated the lines

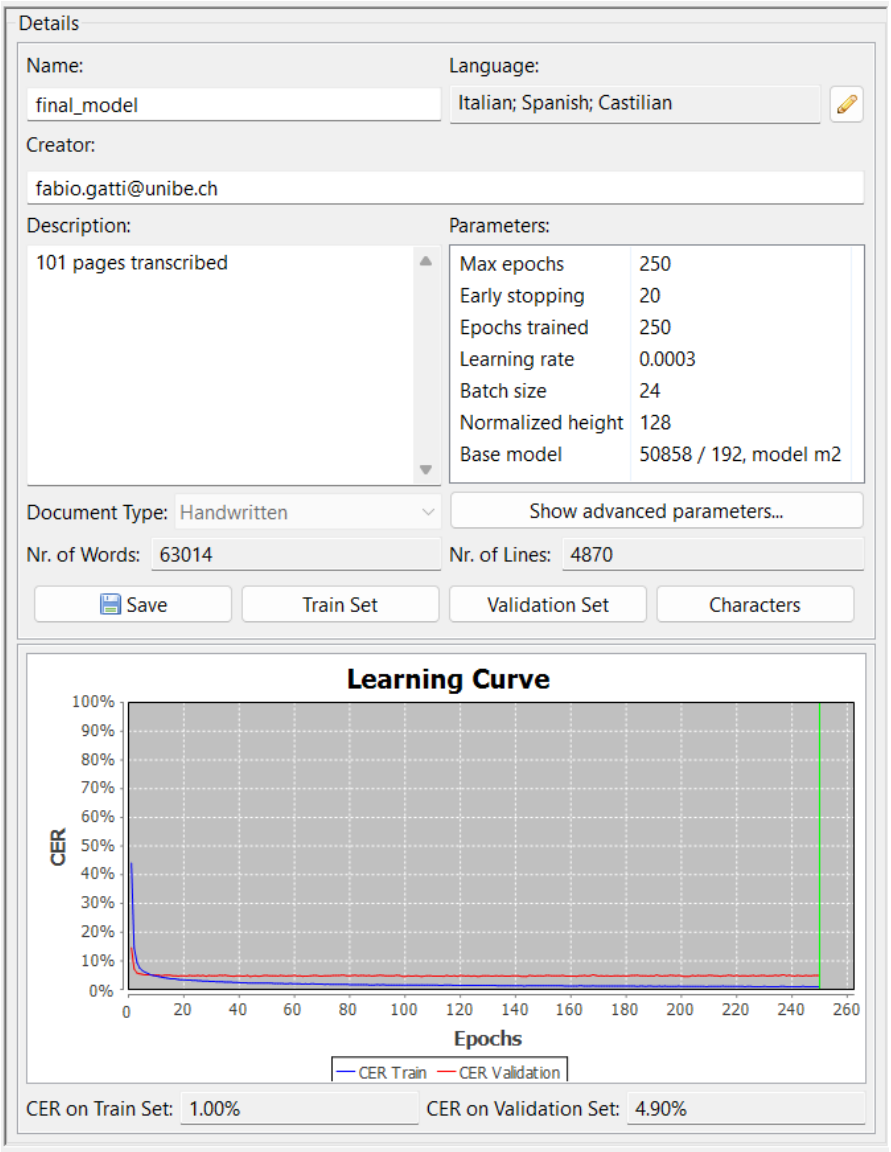

Figure 2: Description of the HTR model emplyed in this paper

whenever a different text region was recognized on the platform (roughly a single-page was considered by the platform as a text region). These 6 .txt files were combined into 2 files since the 6 *Copialettere*, as noted earlier, were temporally paired as 192-193, 194-195, 196-197, so that uniting the texts of 192-194-196 and 193-195-197 was the first step for the subsequent merge of all the files into a comprehensive document. For each of the two files each letter transcribed in the *Copialettere* was separated by an empty line, with occasional reference checks to the digitized material. Since from the letters the precise date was not always clear, but it was quite easy to identify the month, we decided to use

the month as the temporal unit to merge and then analyse the text. Checking back to the original documents, we added a *_month_* etiquette in the text files before the first letter sent for each month. We give each *_month_* a progressive number, and then merged the two texts on the base of this template. At that point we had a unique .txt file containing the transcriptions of all the letters sent by the *Saminiati & Guasconi* from late December 1626 to February 1634, plus the letters sent in March and April 1634, for all the none Roman and Venetian addressees. After trimming the initial and final letters to have a consistent time frame[6] with all outgoing correspondence, our final file consisted of all letters sent from January 1627 to February 1634.

5.2 Choice of Keywords and search for variations/errors

Usually a good starting point in performing an automated text analysis consists of stripping out elements of the wording considered not relevant, such as punctuation, very common words (usually called "stop words") or very rare ones, and then to proceed to a comprehensive analysis of the remaining bulk of the text (Gentzkow et al., 2017). In our case, we instead decided to ignore a large part of the wording (written in a cryptic renaissance business jargon) and focused our attention only on a few hundreds relevant key-words, ruling out any "holistic" method trying to interpret and classify the letters on the base of their overall content. Indeed, in face of the complexity of the language used (and the ambiguity resulting from the AI decoding errors), such an approach would "necessarily fail" (Grimmer and Stewart, 2013, p.270).

In order to perform a textual analysis of the content of the letters for the purpose of quantifying trade we focused on extracting information on the addressee, place of destination, goods traded and means of transportation. In this regard we were able to determine 446 different merchant companies, 65 towns and a set of 27 goods and 7 different means of naval transport. Regarding trade in goods, we identified several typical products of the Florentine industries, such as different types of silk fabrics[7] (Goldthwaite, 2009, Malanima, 1983, Morelli, 1976) and woolen clothes[8] (Ammannati, 2020, Edler, 1934, Malanima,

[6] Only a handful of letters were sent on December 1626, the company being founded the 26th, while for March and April 1634 we only had the letters sent to just a part of the correspondents

[7] *drappi, rasi, ermisini, chermisini, tele, taffette, velluti, dommaschi, broccati*

[8] *rascie, saie, pannine, baiette, perpignani, ciambellotti*

1983, Munro, 2005, Spallantani and Bruscoli, 2003), among the most frequently cited items. Also, we searched for the mention of raw silk, wool and cotton, as well as for the mentioning of edible goods and spices such as pepper (Berglez, 2021), saffron, sugar and oil. Among the precious goods, the most cited were diamonds (G.Lenzen, 1970, Hofmeester, 2019, Sabel, 2019), necklaces and silver, while the spanish *reali* and *doppie* were the most frequently cited precious metal coins (Barrett, 1990, Chen, Palma, and Ward, 2021, L.Pezzolo and G.Tattara, 2008). The seven different types of sea crafts we identified were *galere, feluche, vascelli, navi, fregate, barche* and *galeoni*.

Conscious of the fact that the 27 main goods we identified were not the only merchandise ever traded by our company (sometimes we found mentions of the trade of other, occasional and possibly just one time-traded goods, like a portrait or elephant tusk), we decided to search also for different types of packaging to help us to identify every letter as being a "merchandise transaction" or not. In this way were able to isolate 10 possible ways through the *Saminiati* traded goods, namely, *casse, balle, fagotti, sacchetti, barili, scatole, salme, fiaschi, vasetti e botti*.

After the identification of the merchants, their towns, the commodities, and packaging systems we needed to address the imprecision of the AI-HTR tool we developed in order to conduct a meaningful text analysis. More precisely, although our model arguably had a very satisfactory rate of error in recognizing single characters, the number of mispelled words was much higher, since each of the key-word we were interested in was composed by multiple characters. We thus inspected the .txt file produced by the model, and compared it with the digitised *Copialettettere*, taking note of all the errors/variations in which the AI generated HTR model transcribed the input pdf into .txt. As an example, the model transcription for the words *Seta - Sete* (Silk - Silks) took 48 different forms (like *Seta, Site, Setas, Gete..*), while the town of Venice as a place for letters' destinations was decoded in 33 different ways (like *Venezia*, *Ven.a*, *Vinizia*, *Vonezia..*). One should note that some of these varieties were actually due to the abbreviations written in the original documents themselves, like *cas.ta* or *cass.ta* for *cassetta* (i.e., crate).

After identifying the variations we encountered in the spelling of the 446 merchants and 65 towns in the first lines of each letter, and of 27 goods, 7 sea crafts and 10 packaging systems in the body of the letter, we performed a final check before conducting the final

text analysis. To this end we checked that each of those variations was referred to the intended merchant/town/good/sea craft/packaging system in the original text, and we corrected the text file where this was not true. Examples include the case of *La* being not intended to be a mentioning of *Lana* (wool), but actually meaning *la*, namely "the" and that of the merchant Della Sedda, whose name was sometimes transcribed in a way (*Seta*) which we would have originally identified as a "Silk" related word. Other cases in which we had to edit the .txt file were those in which the model decoded (more or less) correctly a word, but included a space between the syllables. For instance, *Napoli* was sometimes transcribed as *Na poli* and *Sete* as *Se te*.

5.3 A code for text-analyis

After the cleaning process just described, we wrote an R code that identifies each letter and stores the contents of each of the 6,376 epistles in separate lists. Then we manipulated the texts by lowering all the the capital letters and segmenting (tokenizing) (Jokers, 2018, Silge and D.Robinson, 2017) the content of each letter into words (Gentzkow et al., 2017)[9], while conserving the words containing punctuation[10]: we now had a set of lists each one containing the "loose words" i.e. the *unigrams*(Grimmer and Stewart, 2013) of each letter. While what we have described so far serves as the starting point of the textual analysis over the *corpus* of the letters, a partly different process was implemented for the identification of the addressee merchant and the related town of destination, such information being written in the first text line of each letter. In this case we wrote an ad-hoc code which took care for the segregation of the original text into letter-blocks and then the tokenization in words, but only considered the first line of each epistle. This differentiation is important since in the *corpus* of the letters other merchants and towns with respect to those to which the letter was addressed are usually mentioned, potentially causing confusion in the identification of the true recipient. After having produced such letter-based blocks of texts, filled by the tokenized words contained in each letter, the proper counting started.[11]

[9]Inside each letter, the original order of the words is ignored, according to a so called *bag-of-words* procedure (Gentzkow et al., 2017)

[10]Because of the abbreviations used in medieval-business italian, with words shortened through dots or commas in the middle: like *b.a* for *balla* (bale)

[11]It is worth noting that the searching (and counting) of the key-words accounting for their encountered variations is a process in a certain way analogous (even if, in facts, opposite) to the *stemming* practice often used in automated text-analysis (Grimmer and Stewart, 2013). Indeed, while the stemming is a

More specifically, the code searched and counted in the text-blocks containing the first line of each epsitle the name of each of the 65 identified towns and 446 merchants, and for each one we provided our search engine with a list of the variations/errors we previously found confronting the HTR output file and the digitized documents.[12] In contrast to the search for words related to the traded products, the counting was done on the text-blocks comprehensive of the whole content of each letter. As before, we accounted for all the variations we found double-checking and comparing the original documents with the Transkirbus's output.[13] The same process used for the products was also implemented for the search of the mention of means of naval transport[14] and that for the counting of the packaging mentions[15].

The output of each research process is a list in which each letter is associated with the counting of the mentioning of each key-word in its *corpus* (summing of all the detected variations). These lists were then transformed into R data frames and merged into a single dataset. Figure 3 exemplifies the working of the text analysis code on the transcriptions, where in the first lines it identifies the town of destination (Krakow) and the addressee merchant (mister Orsetti), and in the corpus of the letter the mentioning of 6 different textiles and of a packaging system.

5.4 Exporting the merged data.frame, and cleaning of the final dataset

The merged data.frame is exported in a data sheet where each letter_ID (from 1 to 6,376) is associated with the name of the addressee merchant, the town of destination, and the count of the mentioning, in the *corpus* of each letter, of each of the 27 goods, 7 sea crafts and 10 packaging systems of interest. At that point we checked whether our dataset was coherent and complete and found that for a few letters no town of destination nor addressee

procedure aimed at removing the ends of the words, to reduce the dimensionality of text but preserve the basic concepts (assumed to be incorporated in the root of each word), in our case we could not rely on the root of words, the errors in character transcription potentially occurring in any part of them, as for the case of *Lete* indicating *Sete* (Silks) or *Pucchiri* for indicating *Zuccheri* (sugars)

[12] As an example for Naples we searched for *"napoli","napli", "nap.li"*. Similarly it was done for the identification of the addressee merchants, as an example, for the Ferrari's' merchant company we searched for the words *"ferrari","ferrard","ferraro","perrari"*

[13] As an example for the mentioning of olive oil we searched for *"olé","oli","ole","olii","dioli","olio"*.

[14] As example, for the word *vassello*, we searched for *"vasselli","vassello","varelli","vasello","vassallo","aussello","vassellio","uasselli","raselli"*

[15] As an example, for the word *scatola* (box), the code searched for *"scatola","scat.ta","siatola", "vcatola","scabolini","scatolino","scatolini"*

_letter_4611
Cracovia Sig.re orsetti Y Adi 23 - ottobre 1632 firenze D
Mag.co Si.re con cara vŕa de 23 - sc.so et qnto a' drappi di Napoli comm.te
per ord.e vŕo speriamo frà pochi gni Sentirli speditti per qui cosi
la Carsa velluti di Genzova che qndo Saranno in poternrone
faremo subita la spedisione per costi conf.te al ordine perche
li dommaschi di Gendua gia che non ci vénné la mostra del
opera et delli coloni attendendo con quanto vi si scrisse con lualtina
ci diciate meglio il pensier vŕo per potere in contrare di ordina.
e farvi restar Servito pontualm.te per che le coltre di Napoli che
qui erano in vendita presero luogo che per.o non haviamo modo
prontam.te di comprlire la Cassa delle telette e ermisini Sensa
commetterle a Napoli che lordine vŕo non Si estende a poterlo
fare tanto meno che non Sappiamo le commettendi a Napoli
fussero di maggior costo di quéllo accennate di β6 la Coltre
e cosi staremo attendendo se volete Sene dia lordine libero nel
qual Caso lo faremo Subbito et Speriamo Sara procacciato
mercanzia buona da darvi giusto ci piace avessi riceuto
il conto del costo e spese della Cassa velluti et Rasi
gia Speditavi et dovete averlo trovato giusto et datocene Crd
et quando ricever.te la mercanzia piaccavi avisarlo con il
ben Servito Vvoi vi dolete de Nŕi di Ven.a Sopra le robe mand.o
e che allo ro avete Scritto il vŕo Senso giovand.vi credere vi
haranno risp.to conplita.te e che non vi mosterrete tanto difcile
in approvare il fatto da vŕi amici tanto con fid. come tuttele
case Nŕe et rimcov.ro da voi quel pensiero del nd volere accett.re
per vŕo conto le mercan.te di vŕo ordine si provegg.no et Si mandono confo
sopra questo quanto vi Sene scrisse con lultima come in Copia preganda
facibitare il negozio accio li possa continuare e Pervirur di buona
voglia come molto lo desideri certificand. non lo possete sperare d qualsvag.
con piu app.to damore et Sincerita di quello noi profess.m

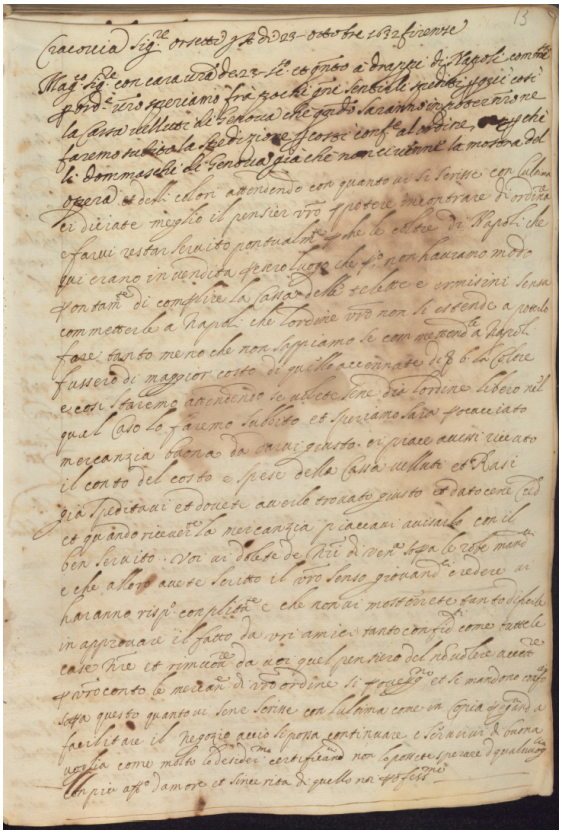

Figure 3: At left: AI-based HTR transcription worked by the text analysis code, with relveant words, catvhed by the program, highlighted. At right the original letter

merchant was detected. The first two issues were solved by referring back to the digitized material and the .txt file. In some case we missed to list some of the variations/errors in the model transcriptions, and we added the new detected variations in the search engine. In the final dataset only 13 letters out of 6,376 were considered to be addressed to "other merchants" and only because the recipient company was actually not written or there was too much mispelling in order to be identified. The same approach was taken for towns of destination, and only 2 letters in our dataset were finally categorized as to "other" towns.

There were also cases where the letters were addressed to more than one merchant. While some merchant companies were run by a partnership between two or more merchants, in some of these cases the *Saminiati* for time saving reasons only wrote the name of the first, or the second merchant giving the name of the company. As an example, to the Roman merchant company *Acciaioloi & Martelli*, sometimes letters were sent addressed to *Acciaioli*, and other times to *Martelli*. In total, checking back in the original documents, we identified 35 of such companies, and we decided to correct, in the final dataset, the

"merchant" variable by substituting the mention of one, or more of the merchants that made up the partnership, with the extended name of Company.[16]

As a final validation step[17] we by hand compared the keyword count for more than 60[18] letters, comparing what was reported in the data sheet, with the .txt file and the scanned documents and found a satisfactory match.

6 Sample Descriptive statistics

To demonstrate how our constructed dataset can provide insight into the *Saminiati & Guasconi* trading operations, we present here some descriptive statistics. It is worth noting that among the 6,376 letters considered, 2,415 are classified as "goods transactions" because they are letters that mentioned at least one of the 27 products or 7 packaging systems noted above, while the other letters are solely financial transactions disconnected from any commodity exchange, such as exchange rate speculations, loans or international money transfers. The *Saminiati & Guasconi* trade was conducted with correspondents located in 65 different places spatially spread from Portugal to Constantinople, and from Sicily to the North Sea, as shown in Figure 4. Letters and goods moved among the network nodes mostly through a system of land couriers (the *condotte*), while sea transport was used for trade with the maritime hubs of Constantinople, Genoa, Leghorn, Messina, and Palermo, as is evident from the mention of naval systems in the letters, depicted in Figure 5.

Not all the towns in the network had the same importance as markets for the *Saminiati*'s traded products. For instance, from Figure 6 one can see how Venice (seat of the second "House" of the merchant-bank), Naples and Rome were the main places of destination of the *Saminiati*'s mail. Goods were also not traded at the same frequency, where it is apparent from Figure 7 that Silk was the most traded good during our time period, followed by other Florentine textiles and by Spanish silver and golden coins (*Reali* and *Doppie*). Figure 8 further aggregates the 27 considered goods into 5 main categories, confirming the

[16] As an example, the code reported *Acciaioli*, or *Acciaioli, Martelli* or *Martelli* in the "merchant" variable of the letters sent to such Roman correspondents, and we substituted all such mentioning with *Acciaioli & Martelli*

[17] Validation is an essential step for any automated text analysis (Grimmer and Stewart, 2013).

[18] the 1% of the letters in the sample

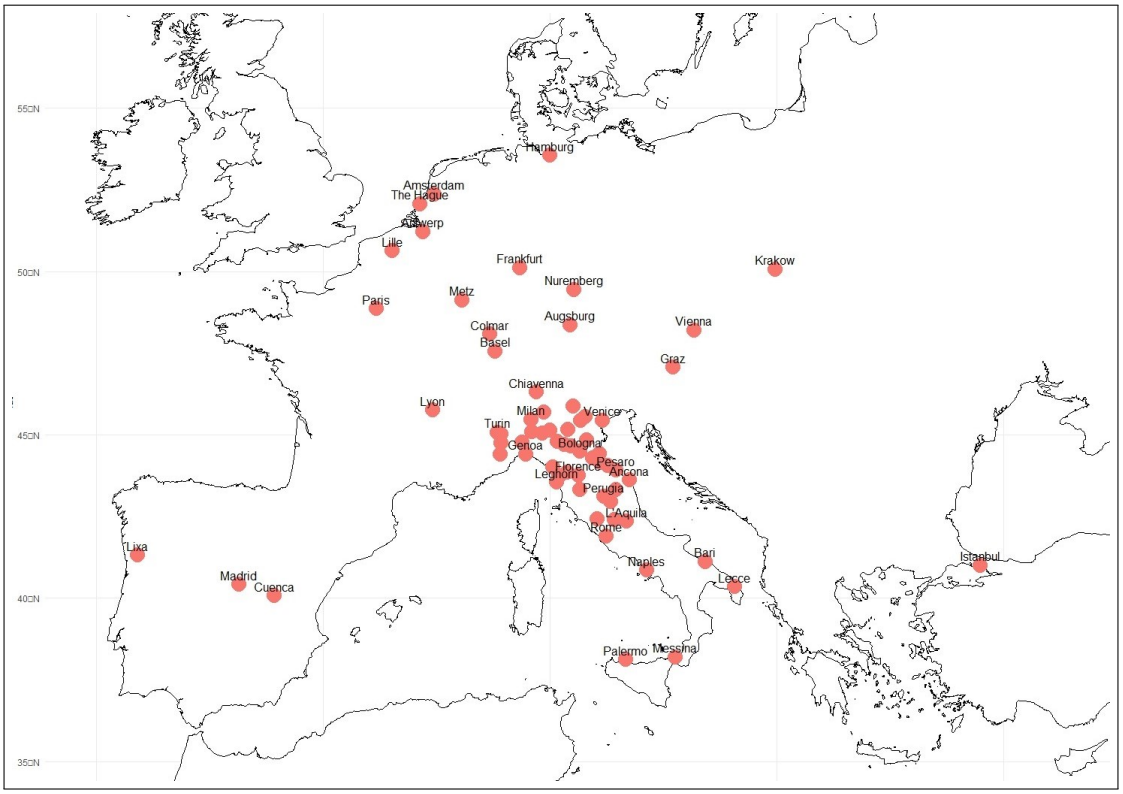

Figure 4: The geographical extent of the *Saminiati*'s network

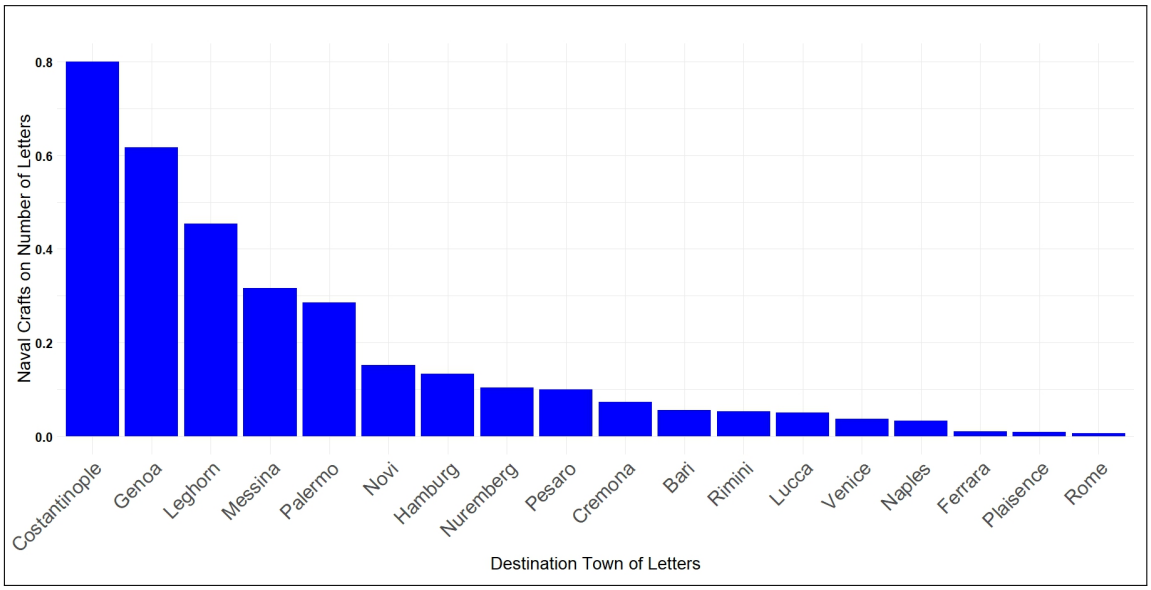

Figure 5: Mentioning of naval transport in total number of letters per town

importance of the silk sector in the Florentine Renaissance industry (Goldthwaite, 2009, Malanima, 1983).

Different goods also tended to differ in their destination markets. For instance, Figure

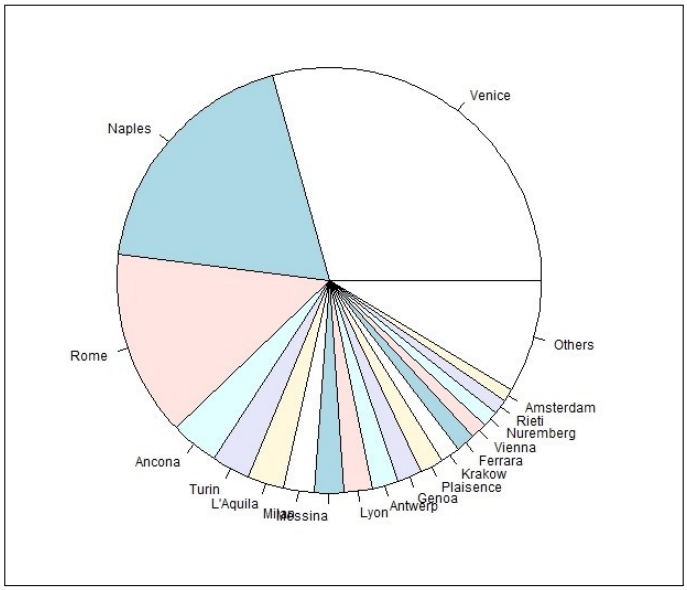

Figure 6: Letters sent by the *Saminiati* per town of destination

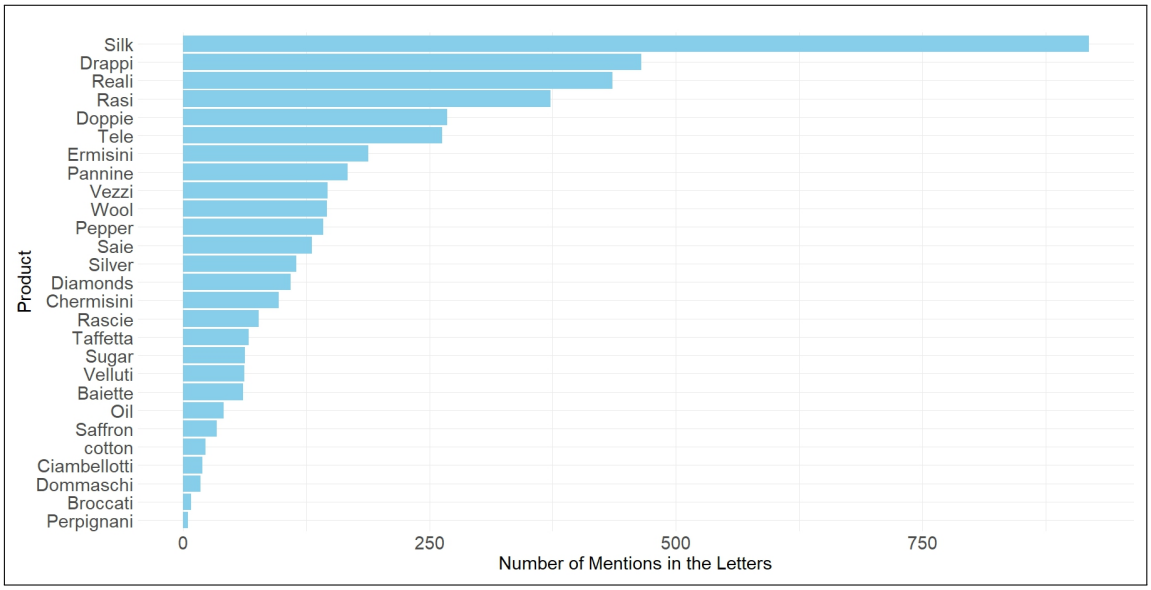

Figure 7: Number of times the considered goods were mentioned in the letters

9, showing where the letters mentioning different products were sent, demonstrates that while Venice, the second "home" of the *Saminiati*, was the most frequent market for silk, diamonds and precious coins, Genoa was important in terms of precious coins transaction because of the financial fairs run by the Genoese and its relationship with the Spanish Crown. One may also note from the figure that the Netherlands was the seventeenth-century hot spot for diamonds, and how Florentine silks enjoyed a diverse and international

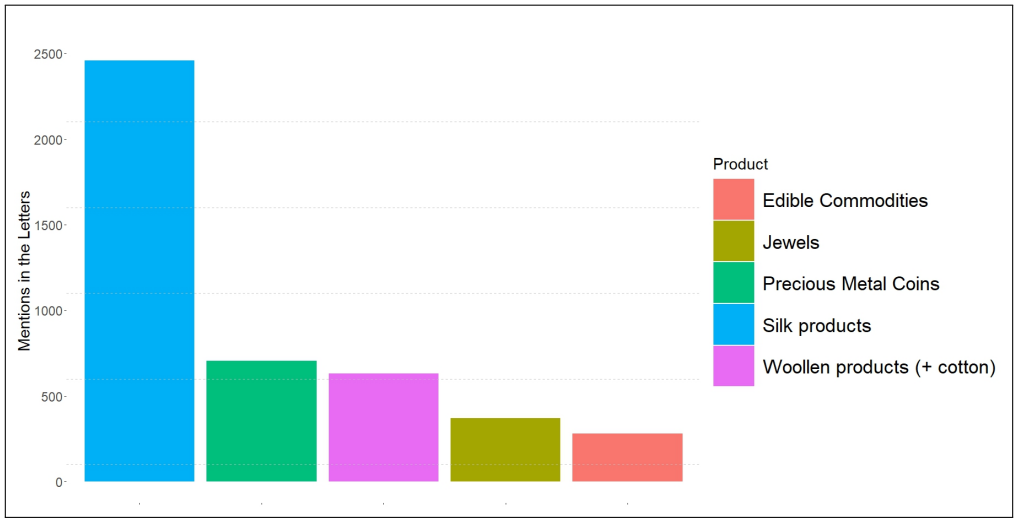

Figure 8: Mentioning of goods, aggregated by kind

demand, confirming views already found in the literature on early modern trade (Barrett, 1990, G.Lenzen, 1970, Goldthwaite, 2009, Hofmeester, 2019, L.Pezzolo and G.Tattara, 2008, Morelli, 1976).

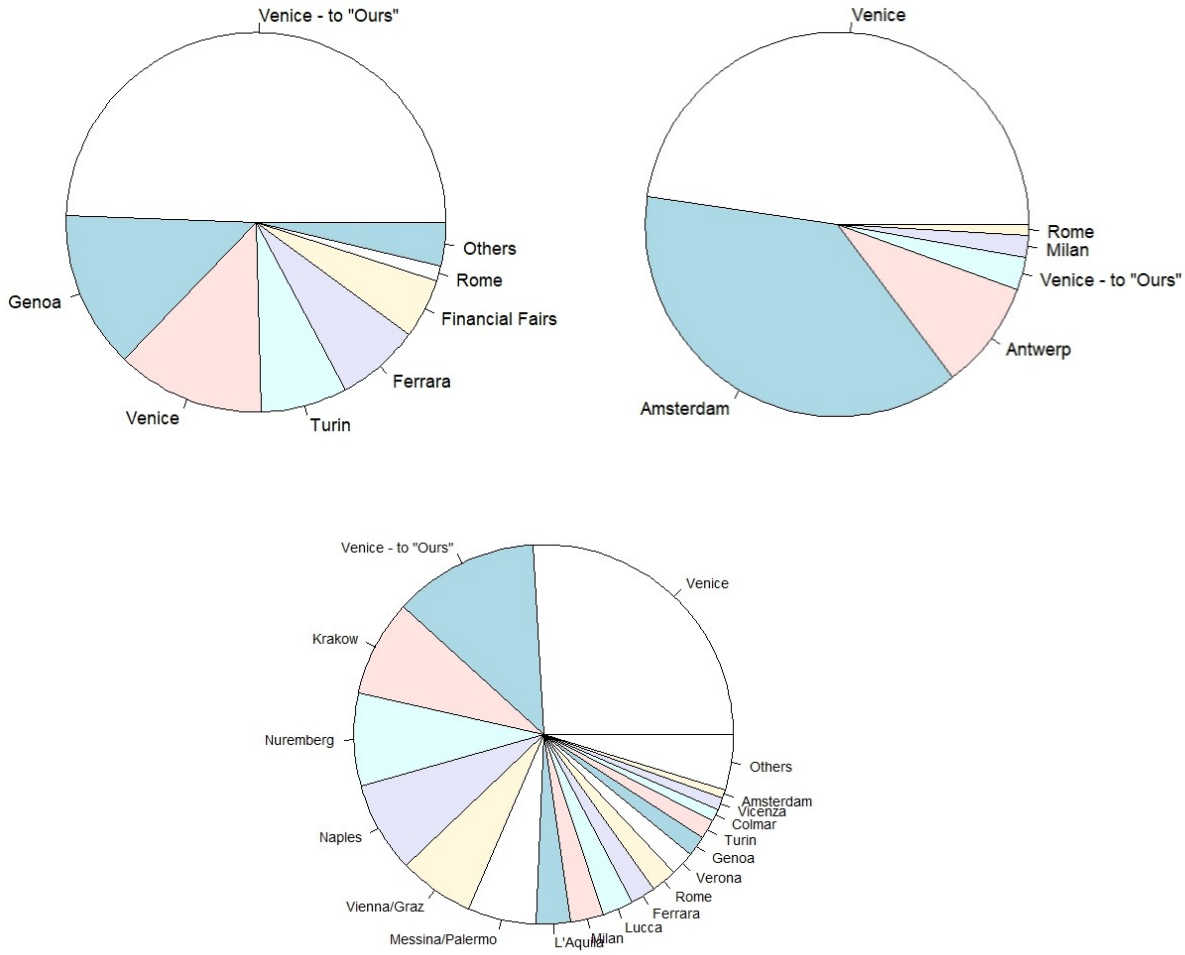

Figure 9: Top left chart: destination of letters containing the mention of precious coins. Top right chart: destination of letters mentioning diamonds. Bottom chart: destination of letters mentioning silk products.

Moving to the analysis at the merchant level, Table 1 shows the 10 companies the *Saminiati & Guasconi* sent the most letters to. In first place is the Venetian branch of the *Saminiati* itself. It is also in the first position in Table 2, representing the merchants with whom *Saminiati & Guasconi* organised most of its goods transactions. It is interesting to note that a high number of letters does not *tout court* coincide with a high number of merchandise transactions. More specifically, many partners were mostly involved in financial operations separate from any goods trade.

Merchant	Number of Letters
Saminiati's Branch in Venice	446
Andreini	221
Betti	161
Malegonnelle	145
Aprea	134
Batta	119
Massella	117
Mora	115
Aquino	108
Sacchetti	106

Table 1: Merchant companies that received the most letters

Merchant	Number of Goods Transactions
Saminiati's Branch in Venice	291
Batta	93
Aprea	89
Canova	77
Andreini	76
Gazzeri	61
Sedda	60
Lonbria	56
Soldani & Pestalozzi	54
Gozzi	52

Table 2: Companies involved in the largest number of goods transactions

7 Conclusions

Mercantile correspondence can potentially provide a unique source of information about business conduct and the economy in preindustrial times. In this paper we outlined a computer-based approach to employ a text-analysis of the out-going correspondence of a major Renaissance Florentine merchant-bank, with the aim of building a dataset suitable for quantitative analysis. The three mile-stones of the process are digitization, development of an AI-based HTR-model, and automated text-analysis. While there was some necessary editing of the textual output of the model, the approach of constructing a dataset from a large amount of historical documents (6'376 letters) represented a significant saving of time compared to any manual approach. Not only can the method be used to analyze the still unexplored letters stored in the Historical Archive of Bocconi University, but could also be employed with other merchant letter collections. Such analyses could provide novel insights into merchant trade.

References

F. Ammannati. *Per filo e per segno L'Arte della Lana a Firenze nel Cinquecento*. Firenze University Press, Florence, 2020.

W. Barrett. World bullion flows, 1450–1800. *The Rise of Merchant Empires Long Distance Trade in the Early Modern World 1350–1750*, pages 224–254, 1990.

A. Bartolomei, C. Lemercier, V. Rebolledo-Duhin, and N. Sougy. Becoming a correspondent: The foundations of new merchant relationships in early modern french trade (1730–1820). *Enterprise Society*, 20(3):533–574, 2018.

J. Berglez. *European Monopoly Strategies in the Early Modern Indian Ocean Spice Trade*. Master Thesis, 2021.

F. Bratchel. Italian merchant organization and business relationships in early tudor london. *Merchant Networks in the Early Modern World, 1450–1800*, Routledge:1–28, 1990.

C. Burlacu and A. Rabus. Digitising (romanian) cyrillic using transkribus: new perspectives. *Diacronia*, 2021.

G. Camiciotti. Letters and letter writing in early modern culture: An introduction. *Journal of Early Modern Studies*, (3):17–35, 2014.

A. Caracausi and C. Jeggle. *Commercial Networks and European Cities, 1400–1800*. Pickering Chatto, London, 2014.

Y. Chen, N. Palma, and F. Ward. Reconstruction of the spanish money supply, 1492–1810. *Explorations in Economic History*, 81:1–26, 2021.

C.M Cipolla. Crisi a firenze nel 1629-1630. pages 245–253, 1988.

J.G. DaSilva. *Banque et credit en Italie au XVIIe siecle*. Klincksieck, Paris, 1969.

A. DeMaddalena. Les archives saminiati : de l' economie à l'histoire de l'art. *Annales*, 14 (4):738–744, 1959.

R. DeRoover. What is dry exchange? a contribution to the study of english mercantilism. *Journal of Political Economy*, 52(250), 1944.

R. DeRoover. *L'´evolution de la lettre de change XIVe-XVIIIe si'ecles*. A. Colin, Paris, 1953.

F. Edler. *Glossary of medieval terms of business : Italian series, 1200-1600*. Waverly Press, Baltimore, 1934.

C. Gaffield and P. Baskerville. The automated archivist: Interdisciplinarity and the process of historical research. *Social Science History*, 9(2):167–184, 1985.

R. C. Galliano. Honore et utile: The approaches and practice of sixteenthcentury genoese merchant custom. *Understanding the Sources of Early Modern and Modern Commercial Law*, pages 55–86, 2018.

M. Gentzkow, B.T. Kelly, and M. Taddy. Text as data. *NATIONAL BUREAU OF ECONOMIC RESEARCH Working Paper series*, 23276, 2017.

P. Gila. From digitization and images to text and content: Transkribus as a case study. *86th Annual Meeting of the Association for Information Science Technology*, 2023.

G.Lenzen. *The History of Diamond Production and the Diamond Trade.* Praeger, New York, 1970.

R. Goldthwaite. *The Economy of Renaissance Florence.* The Johns Hopkins University Press, Baltimore, 2009.

J. Grimmer and B. M. Stewart. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, (21):267–297, 2013.

G.H. Guttridge. *The American Correspondence of a Bristol Merchant 1766-1776.* University of California Press, Berkeley, California, 1934.

K. Hamad and M. Kaya. A detailed analysis of optical character recognition technology. *International Journal of Applied Mathematics, Electronics and Computers*, 4:244–249, 2016.

L. Hofmeester. Knowledge, technique, and taste in transit: Diamond polishing in europe, 1500–1800. *Gems in the early moder world, Materials, Knowledge and Global Trade, 1450-1800*, pages 281–308, 2019.

L. Jokers. *Text Analysis with R: For Students of Literature.* Springer, 2018.

M. Kohn. Merchant banking in the medieval and early modern economy. *Dartmouth College Department of Economics Working Paper*, pages 1–18, 1999.

L.Pezzolo and G.Tattara. Una fiera senza luogo": Was bisenzone an international capital market in sixteenth-century italy? *The Journal of Economic History*, 68(4):1098–1122, 2008.

P. Malanima. *La Decadenza di un'Economia cittadina L'industria a Firenze nei secoli XVI-XVIII.* Il Mulino, Bologna, 1983.

G. Mandich. *Mandich, Giulio. Le pacte de ricorsa et le marche italien des changes au 17e siecle.* Affaires et gens d´affaires, Paris, 1953.

C. Marsilio. Nel xvii secolo dei genovesi. la corrispondenza commerciale di paolo gerolamo pallavicini nel triennio 1636-1638. *Edizioni Scientifiche Italiane*, (1):101–119, 2005.

F. Ruiz Martin. *Letteres merchandes échangées entre Florence et Medina del Campo.* S.E.V.P.E.N., Rue du Four, 13, Paris, 1965.

B. Maréchaux. Business organisation in the mediterranean sea: Genoese galley entrepreneurs in the service of the spanish empire (late sixteenth and early seventeenth centuries). *Business History*, 65(1):56–87, 2023.

M. Massot, A. Sforzini, and V. Ventresque. Transcribing foucault's handwriting with transkribus. *Journal of Data Mining Digital Humanities*, 2019.

F. Mauro. Merchant communities, 1350-1750. *The Rise of Merchant Empires Long Distance Trade in the Early Modern World 1350–1750*, Cambridge University Press:264, 1990.

J. McCusker. The demise of distance: The business press and the origins of the information revolution in the early modern atlantic world. *The American Historical Review*, 110(2): 295–321, 2005.

J. McCusker and C. Gravesteijn. The beginnings of commercial and financial journalism: The commodity price currents, exchange rate currents, and money currents of early modern europe. *The American Historical Review*, 99(2), 1991.

R. Morelli. *La seta fiorentina del cinquecento.* Giuffrè Editore, Milano, 1976.

G. Muehlberger, L. Seaward, M. Terras, S. Ares Oliveira, V. Bosch, M. Bryan, S. Colutto, H. Déjean, M. Diem, S. Fiel, B. Gatos, A. Greinoecker, T. Grüning, G. Hackl, V. Haukkovaara, G. Heyer, L. Hirvonen, T. Hodel, M. Jokinen, P. Kahle, M. Kallio, F. Kaplan, F. Kleber, R. Labahn, E.M. Lang, S. Laube, G. Leifert, G. Louloudis, R. McNicholl, J.-L. Meunier, J. Michael, E. Mühlbauer, N. Philipp, I. Pratikakis, J. Puigcerver Pérez, H. Putz, G. Retsinas, V. Romero, R. Sablatnig, J.A. Sánchez, P. Schofield, G. Sfikas, C. Sieber, N. Stamatopoulos, T. Strauß, T. Terbul, A.H. Toselli, B. Ulreich, M. Villegas, E. Vidal, J. Walcher, M. Weidemann, H. Wurster, and K. Zagoris. Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study. *Journal of Documentation*, 75(5):954–976, 2018.

J. Munro. *The woollen cloth industry in Italy: The rise, expansion, and decline of the Italian cloth industries, 1100 - 1730*. Munich Personal RePEc Archive, 2005.

J. Nockels, P. Gooding, S. Ames, and M. Terras. Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of transkribus in published research. *International Journal of Applied Mathematics, Electronics and Computers*, 22(3):367–392, 2022.

J.M. Price. *Joshua Johnson's Letterbook, 1771-1774 Letters from a Merchant in London to His Partners in Maryland*. London Record Society, 1979.

F. Saba. Commercio e banca nell'europa del xvii secolo. la corrispondenza delle compagnie di ascanio saminiati conservate nell'archivio saminiati pazzi depositato presso l'università bocconi. *Storia Economica*, 22(1):93–137, 2019.

C. Sabel. The impact of european trade with southeast asia on the mineralogical studies of robert boyle. *Gems in the early moder world, Materials, Knowledge and Global Trade, 1450-1800*, pages 87–116, 2019.

Saminiati-Pazzi Historical Archive. Section 1, registers 192-193-194-195-196-197, 1626-1634. Bocconi University, Milan.

R. Schlagdenhauffen. Optical recognition assisted transcription with transkribus: The experiment concerning eugène wilhelm's personal diary (1885-1951). *Journal of Data Mining Digital Humanities*, 2020.

S.Groppi. *L'Archivio Saminiati-Pazzi*. EGEA, Milan, 1990.

J. Silge and D.Robinson. *Text Mining with R: a tidy approach*. O'Reilly, 2017.

M. Spallantani and F. Guidi Bruscoli. *Tessuti di seta tra Firenze e il Levante (ca. 1350-1550) : le fonti*. Firenze University Press, Florence, 2003.

M. Spallanzani. *La Lana come Materia Prima*. Leo S.Olschki,, Florence, 1969.

S. Subrahamanyam. *Merchant Networks in the Early Modern World, 1450–1800*. Routledge,, New York, 1996.

U. Tucci. *Lettres d'un marchand vénitien Andrea Berengo (1553-1556)*. SEVPEN, Paris, 1957.