Libro de Arte Coquinaria: annotation campaign and training of a model using Transkribus

Digital Humanities and Digital Knowledge, University of Bologna
January 2022.

Francesca Borriello francesca.borriello@studio.unibo.it
Constance Dami constance.dami@studio.unibo.it
Lorenza Pierucci lorenza.pierucci@studio.unibo.it

Abstract.

The present work documents a small-scale Handwritten Text Recognition (HTR) project aimed at transcribing a fifteenth-century Italian manuscript by training a model on Transkribus. The report is composed of three main sections: the first describes the annotation campaign we carried out to obtain high quality data to train the model; the second reports on the actual training on the Transkribus platform; the last section is finally dedicated to the evaluation of the obtained model.

Table of contents.

Introduction: An ancient cookbook manuscript.

We worked on an Italian fifteenth-century manuscript carrying the text of the *Libro de arte coquinaria* by Maestro Martino Da Como.

Martino da Como (1630 – late XV) — or the *prince of cooks*, as his contemporaries used to call him[1] — was a professional chef working at the papal *curia* during the second half of the fifteenth century.[2] His book of recipes, known as *Libro de Arte Coquinaria* (*The Art of Cooking*), was composed around 1465: it represents a witness of the historical passage from mediaeval to Renaissance cuisine and has been defined as «the first modern cookery book».[3]

The manuscript we chose represents one of the four versions of Martino's cookbook: it dates to the period between 1460 and 1480 and is composed of 64 well-preserved textual leaves displaying front and back an understandable, careful handwriting. The latter observation suggests that the manuscript's 'author' was a professional scribe: in this regard Paul Shaw, an expert calligrapher and designer, suggested that — following a common practice — the scribes may indeed have been two — one of them being Antonio Toffio[4], a renowned professional of his times.

The manuscript is written in *humanist minuscule*, a handwriting style created in by Italian intellectuals (the Renaissance humanists) as a reaction to the "crabbed" Gothic minuscule. The *antiqua* — as they called it — was in fact an attempt to revive 'antiquity' through a meticulous imitation of the Carolingian minuscule (XI-XII century), of which it also adopted abbreviations and graphic uses. The manuscript's handwriting, clear and stable, is in two tones of brown ink, with some *marginalia*, few layout peculiarities and few, legible corrections — mainly strikethrough text and superscript additions preceded by a caret symbol.

The manuscript is part of the Bitting collection of the Library of Congress, which makes scans available on its website in image (TIFF, JPEG, JPEG2000, GIF) and PDF format and in good resolution.[5]

1. Annotation campaign.

1.1. Task definition.

The scope of the annotation campaign was to obtain high-quality data in order to train an HTR model to automatically extract text from our manuscript.

We aimed at transcribing twenty-five selected pages from the digitized version of the manuscript: the resulting transcription was meant to represent the text as faithfully as possible, retaining its most characterizing features as carried by the document. In order to obtain a robust and effective model, we designed some guidelines and evaluated our pilot transcriptions to assess consistency between the annotators.

[1] Sacchi 1474.
[2] Jenkins 2007, pp. 97–103.
[3] Bellarini 2005.
[4] Shaw 2006.
[5] See https://memory.loc.gov/ammem/about/techStandards.pdf.

1.2. Pilot.

During the pilot phase, each of us transcribed the same few pages.[6] In particular, we identified and pointed to possible biases and difficulties that could come up during the transcription phase as a consequence of the text's characteristics. These include abbreviations and graphical uses such as:

- Superscript characters (sometimes as a form of contraction which imply preceding omitted characters);
- Brevigraphs, such as '&' for *et*;
- Tildes implying omitted letters, as in the suspended form *ĩ* for 'in';
- Punctuation marks and whitespaces used in a way that does not correspond to modern usage;
- Ligatures, equivalent *u/v*, long *s (ſ),* etc.

In order to evaluate and analyse our pilot transcriptions, we wrote and took advantage of a Python script that — along with an inter-annotator agreement measure[7]— returns all differences between input transcriptions, line by line. We did so in order to obtain a clear scenario including all transcription errors we were possibly overlooking. In this way, our attention was drawn on less evident but not secondary details, such as:

- Accents and other diacritics;
- Scribal corrections;
- Non-standardized and thus inconsistent forms;

We discussed each of these points and agreed on some temporary annotation guidelines. At first, in fact, we opted for carrying out a diplomatic yet slightly normalized transcription: our original goal was to obtain a text that could be read and searched both by experts and by the general public. We thus decided to expand abbreviations and accent words that would otherwise prove ambiguous or difficult to understand.[8] These choices were not in contrast with the Transkribus' transcription conventions: although they advised against an excessive normalization of the text, they did not entirely exclude it.[9]

We thus transcribed the same three pages following these guidelines and computed the inter-annotator agreement. In doing so, we noticed that:

- The level of agreement was not satisfying (see *Table 1*);
- The task ('making the text readable') was vague and the guidelines were consequently unclear and inconsistent;
- The normalization risked to conceal the most characterizing features of our document;

We thus decided to update our task and our annotation model and guidelines towards a stricter diplomatic transcription, in the attempt to alter the textual features as little as possible (see *Annotation model* and *Guidelines*). We did so in order to:

1. Minimise transcription errors depending on the annotators' interpretation;

[6] We selected representative pages according to the criteria exposed in § 1.5.
[7] See § 1.3. Of course, the pilot pages were not included in the Ground truth.
[8] E.g. 'e' used as a preposition (*and*) and 'e' used as the third person singular of the verb to be (*he/she is*), which in current Italian is accented ('è').
[9] «Neural networks are often able to learn to recognise and reproduce expansions. E.g. Latin prefixes and suffices such as "cum", "con" or "us" and "orum" are learned easily by the machine.» Cfr. How to enrich documents with mark-up, last updated in April 2021.

2. Follow more closely the Transkribus conventions adopted by several authoritative projects (see for example the Italian Administrative Hands model).
3. Obtain a more robust model that could in principle be reusable, e.g. as a base model for other documents;

1.3. Inter-annotator agreement.

In order to evaluate our agreement on the pilot transcriptions and to assess whether our model and guidelines were good enough to support our final transcription, we created a Python script, accessible at this link. Specifically, the script takes in input two text files (our transcriptions) split into lines, checks whether they contain the same number of lines, compares the lines and returns:

a. A list of all the differences between them, line by line;

b. The percentage of agreement, obtained by dividing the number of agreed words by the number of overall words on the page. This metric corresponds to the so-called "Index of crude agreement"[10] or "Observed Agreement" (A_o), the simplest index for the I.A.A.[11]

We thus compared the files two by two for each of the three pages of the pilot and spotted the errors thanks to the output of the script. We agreed that in order to state that the inter-annotator agreement was sufficient, it would have to reach 95% of correctness. Indeed, one of Transkribus' pioneers, Dirk Alvermann stated that «A good HTR model should recognize 95% of a handwriting correctly, the CER is not more than 5%. [...] As a rule, the WER is three to four times higher than the CER and is proportional to it».[12] Since our Python function calculates the Word Error Rate (WER), we actually set higher standards than those advised by Alvermann. After understanding the type of errors each of us made and correcting the guidelines to make them clearer, we redid the transcriptions, ran the text again, and reached the following agreement:

First run	Ann1 - Ann2	Ann1 - Ann3	Ann2 - Ann3
p.3	92.7%	90.1%	92.3%
p.34	87.5%	87.9%	90.3%
p.40	90.4%	92.1%	90.4%

Table 1

[10] Goodman & Kruskal, 1954.

[11] More accurate measurements exist, of course (Cohen's κ, Fleiss' κ etc). Cfr. Gloria Gagliardi, 2018.

[12] Alvermann, 2021. https://rechtsprechung-im-ostseeraum.archiv.uni-greifswald.de/word-error-rate-character-error-rate-how-to-evaluate-a-model/

Last run	Ann1 - Ann2 (final)	Ann1 - Ann3 (final)	Ann2 - Ann3 (final)
p.3	97.5%	97.5%	99.2%
p.34	96.3%	97.0%	97.0%
p.40	94.1%	94.1%	95.6%
Average IAA	96.0%	96.2%	97.3%

Table 2

87%-89%	89%-91%	91-93%	93%-95%	95%-97%	97%-99%	99%-100%

Even though two of the results for page 40 were below our threshold of 95 %, the average of Inter-annotator agreement for the three couples of annotator was sufficient, also keeping in mind that our threshold was higher than required (as previously discussed).
Based on this result, we stated that our inter-annotator agreement was satisfying and that we could divide between the three of us the pages to transcribe during the actual campaign.

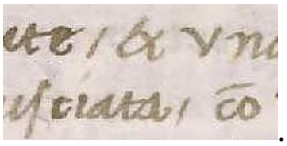
1.4. Annotation model.

As mentioned in the *Pilot* section, the definition of our annotation model was the result of an iterative process. Originally, we adopted a normalising approach, but it soon resulted to be error-prone and ineffective when it came to annotating and (presumably) to training.
We thus turned to a stricter diplomatic transcription, aimed at meticulously reproducing what is on the page, even in case it substantially differs from the current custom. We thus decided that our transcription should not only include any abbreviation and graphical convention, but also represent gaps, corrections and possible inconsistencies.
For each of the observed textual features and in compliance with this annotation model, we proceeded to define our *transcription guidelines*.

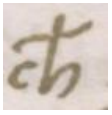
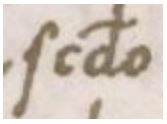
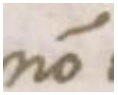
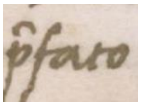
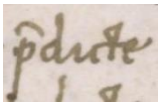
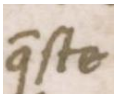
1.5. Transcription guidelines.

To transcribe our text, we chose to follow the guidelines suggested by the Transkribus team, who published these Transkribus Transcription Conventions. We also added some specifications from the OCR-d Ground Truth Guidelines of Level 1, when they applied to our text.
Firstly, after uploading the document in Transkribus, we ran the layout parser to split the documents into text regions and baselines. Before transcribing the text, we checked that the baselines were correctly detected by the software and, if needed, corrected them manually.
Then we transcribed the text line by line, following these rules:
We wanted to produce a diplomatic transcription, transcribing character by character according to what was shown on the image, even though the orthography may be different nowadays. It included combining words according to the original text, even if they would be separated in modern language. The reproduction of spaces was limited to the separation of words.

Uppercase letters were transcribed, when clearly identifiable as such even in the middle of sentences. In case of doubt and if the modern spelling rules would identify it as lowercase instead of uppercase, we chose the former.
Hyphenated words between two lines were transcribed following the original text, with the hyphen when it was present. We also made sure to correct the baseline if the hyphen had not been properly detected by the software.
Strikethrough, **superscript** and **underlined** passages were tagged as such using the button in the Text Editor field. When a character or a section was **unreadable**, it was tagged with the tag "gap". If some characters were legible, they were transcribed normally. Eventually if the transcriber had a proposition to make for the unclear passage, it was tagged as "supplied".
Punctuation was transcribed as close to the original text as possible and always used in conjunction with the preceding word. We chose to transcribe the *virgula suspensiva* (punctuation mark used to signal brief pauses and followed by either a lowercase or capital letter) as commas ",":



Abbreviations. We chose to keep the abbreviations as consistent as possible with the original text and not expand them. We took reference from the only model for historical Italian available on Transkribus to be as consistent as possible with existing authoritative data. To do so, we used the extra Unicode characters available on the Transkribus interface, mainly from Latin-1 Extended A and B.
Here is an illustrative list of the most frequent abbreviations found in *De arte coquinaria*:

-  cħ *(che)*
-  sc͡do *(secondo)*,  nō *(non)*
-  ꝑ *(per)*
-  p͡fato *(prefato)*,  p͡dicte *(predicte)*
-  q͡ste *(queste)*
-  ī *(in)*
-  bollīr *(bollire)*,  forī *(fore)*
-  & *(et)*

-  Menest[e] *(Menestre)*
-  đ *(di)*
-  S[re]. *(Signore)*

The letters u and v were used indifferently at the time and are now used separately; we therefore transcribed them according to modern usage. The same goes for the letter i and j.

Accentuation. Although Transkribus' guidelines advise against transcribing the accents, we transcribed the accents evidently written in the original text, such as "ó":  , that came up quite often in *De arte coquinaria*.

Apart from this specific case, all accents were ignored. Indeed, the use of accents is very sporadic and inconsistent in this manuscript, as it was usually the case until 1501.[13] Likewise, we did not add apostrophes since they are absent from the original text and were introduced in Italy only in 1501.[14]

Some letters have different forms, like normal and long "s":  *se*, or the "c"  being slightly transformed in the ligature with the "t":  *lacte*. The transcription did not reflect these differences and we transcribed with normal letters.

1.6. Corpus selection.

By carefully examining all the pages in our collection, we selected some of them that presented some peculiarities (notes in the margin, titles in capital letters, etc.) and included them in the selection of the pages to transcribe, in order to cover all of the specific particularities of the text and train the model to recognize as many characters as possible. Besides these 6 pages[15], we randomly selected 19 pages that we also randomly divided between the three of us, in order to reach a total of 25 pages, the minimum number advised by the Transkribus team to start training a model.

1.7. Transcription process on Transkribus

We each individually transcribed between 8 and 9 pages, following the previously established guidelines. We downloaded Transkribus' Expert Client and worked on our individual computers, using the feature allowing several users to work on the same document. After rechecking it several times, we also read our colleagues' pages, to make sure that we all agreed on the transcription. During the process, we continuously exchanged when encountering a peculiar character or word in the manuscript, and agreed upon the decision to take, to ensure the consistency of transcription choices across the whole annotation campaign.

[13] Giampaolo Tognetti, 1982.
[14] Idem.
[15] Pages 1, 2, 19, 22, 24, 49 of the text of *De arte coquinaria*

2. Training.

After establishing our ground truth, we moved on to the training for creating a model on Transkribus.

In doing so, we decided to both train a model from scratch and relying on an already existing base model, to detect analogies and differences in both performances.

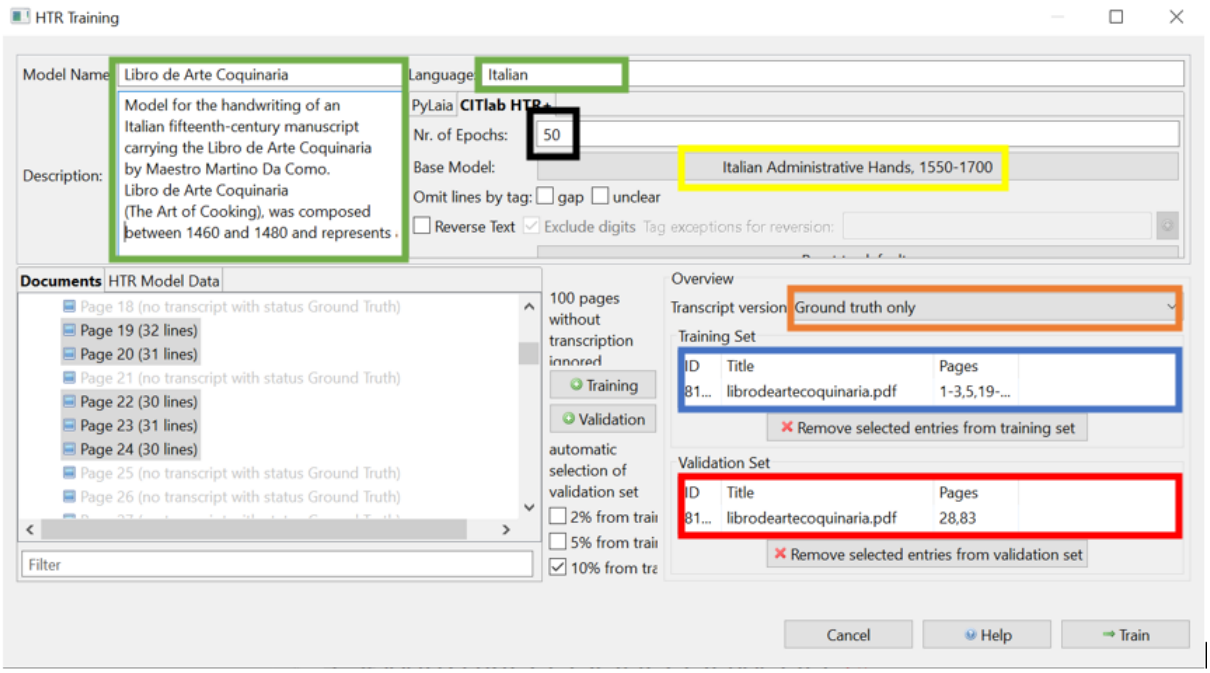


Figure 1. "HTR Training" window.

We started by inserting the model's name, its language and a brief description specifying the kind of source we have used and the period it belongs to. (*Fig.1*, green box)

We then set the *Nr. of epochs* to 50 (*Fig.1, black box*). *Nr. of epochs* «refers to the number of times that the training data is evaluated»[16]: in other words, it indicates how many times the computer goes over the training pages and checks them with the transcription. Although larger projects require a high number of epochs, we considered 50 to be enough for our small scale-one.[17] The pages we transcribed for the training were set to *Ground truth* status. We selected them by choosing the *Ground truth only* option for the *Transcript version* (*Fig.1, orange box*). We used these settings both for the model with base and without.

2.1. Training with a base model.

When training a new model, Transkribus makes it possible to select a base model from the ones publicly available on its server. Leveraging on its knowledge, such a base model is supposed to speed up and boost the performance of the to-be-trained, new model. Indeed, it seems to «make more sense to adapt existing models, benefitting from the training data that is already in the

[16] https://readcoop.eu/transkribus/howto/how-to-train-a-handwritten-text-recognition-model-in-transkribus

[17] For further details see: Huygens ING, "Transkribus webinar by Annemieke Romein", https://www.youtube.com/watch?v=5YCfaFNMol4&t=6630s .

system rather than training every new model from scratch. »[18] In addition, base models allow starting the process with a smaller amount of training data so as to reduce the workload.
For our first training, we thus decided to add a base model and check its results.
Typically, the first step in choosing the right base model is to assess whether it deals with the same language and period in time as the trained one. It is then useful to test it and see how it behaves on the new material.

As a base model, we chose the *Italian Administrative Hands 1550-1700* (*Fig.1*, *yellow box*):

> The Italian Administrative Hands model features a variety of Italian-language documents from state archives in Milan, Venice, Florence, Pisa, and Genoa. The training set represents a spectrum of humanistic, italic and cursive hands characteristic of administrative records, employed by secretaries and news writers. The model has been trained to perform well with a mix of quantitative and qualitative information as well as many common proper nouns for the period, such as locations in Europe and contemporary rulers.[19]

Although the base model belongs to a later period, we decided to rely on it because it a) was the only HTR public model specifically trained on Italian, b) was based on some humanistic hands, c) proved sufficiently good when tested on a page of the *Libro de Arte Coquinaria*.

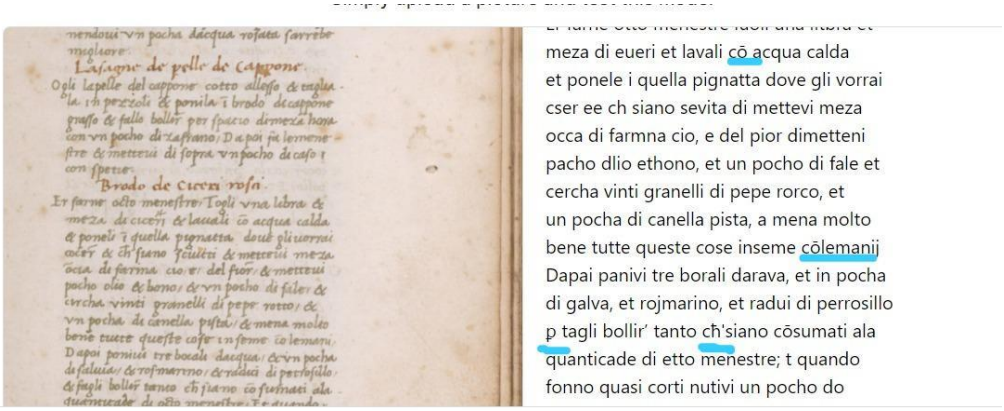


Figure 2. Output of the test performed on a page of the Libro de Arte Coquinaria, in blue the abbreviations detected.

We further checked the list of characters in the model specification and, since a good number of characters found in our manuscript were available, we concluded it was a good starting point for our training.

[18] Muehlberger et al. 2019.
[19] Jake Dyble, Antonio Iodice, Sara Mansutti and Rachel Midura (Primary Investigators), "Italian Administrative Hands, 1550-1700 Handwritten Text Recognition (HTR) Model for Transkribus," Early Modern Digital Itineraries, July 2020, https://emdigit.org/tool/2020/07/21/italian-administrative-hands.html.

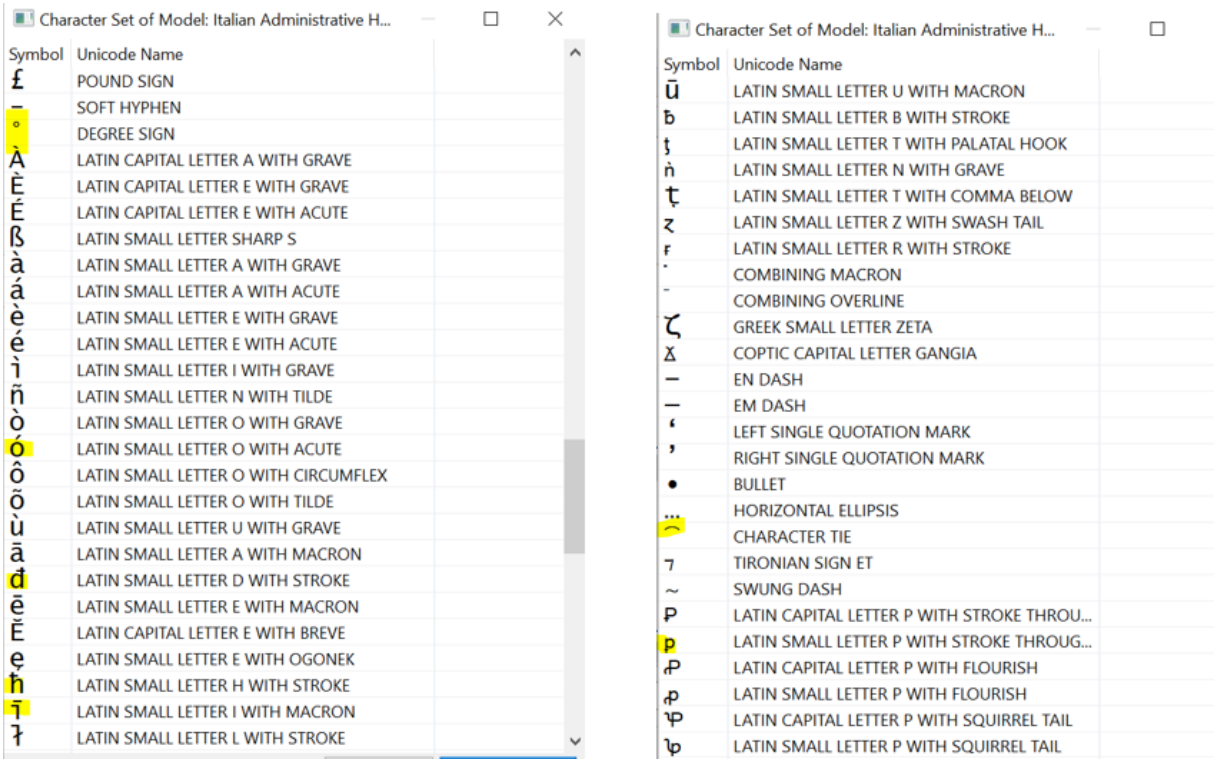


Figure 3. Base model's character set, in yellow the characters also found in the manuscript.

Next, we selected the pages to be included in our set of training and validation data for both our models (with and without base).

2.2. Training set and Validation set.

We selected all the ground truth pages, sending 90% of them in the *training set* (*Fig.1, blue box)* and 10% in the *validation* one (*Fig.1,* red box*)*. For the latter, we chose the option that automatically picks 10% of the training set pages. Having transcribed 25 pages out of 127, 23 were picked for the training set and 2 were used as validation.

2.3. Training process.

The training process for the model with a base took 1h 30m 19s 11ms; the one we created from scratch, instead, was ready in 1h 28m 58sec 747 ms. The fact that model-with-base was not faster (as we expected) may be due to the small quantity of training material: the difference in speed probably emerges when a greater number of pages is involved.

3. Evaluation.

After training our models we analysed their performances and measured their accuracy in different ways. More specifically, we a) examined their learning curves, b) interpreted the Character Error Rate (CER), c) compared the two models by testing them on the same page, d) analysed the models' errors with respect to our Ground Truth. The outcome of our analysis follows.

3.1 Learning curve

The graphs below show the learning curve for both models: in them, the y-axis represents the CER, while the x-axis represents the epochs. The CER is the average percentage of characters transcribed incorrectly by the machine and is typically used to measure models' accuracy.

The two lines in the graphs thus represent the models' accuracy throughout the training: the red line represents accuracy on the validation set, while the blue line represents accuracy on the training set. The validation line is the most significant since it tells how well the model performed on 'unseen' text, *i.e.* the pages it has not been trained on. Ideally, the two lines should be close to each other[20]: this is indeed the case for the model-with-base's learning curve, while the other's lines are slightly more detached (*Figures 4, 5*).

The final CER for the model trained with a base is of 1.48%; for the one without base, it is slightly higher (1.59%).[21]

> In the best cases, HTR 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). If an HTR model is less accurate, with a CER of more than 10 per cent, experiments suggest that automated transcriptions become less useful as a research resource in themselves because correcting myriad errors is more time consuming than manual transcription. [22]

Significantly, both our models were below the 5% critical threshold and thus can be categorized as very good models for HTR on this collection.

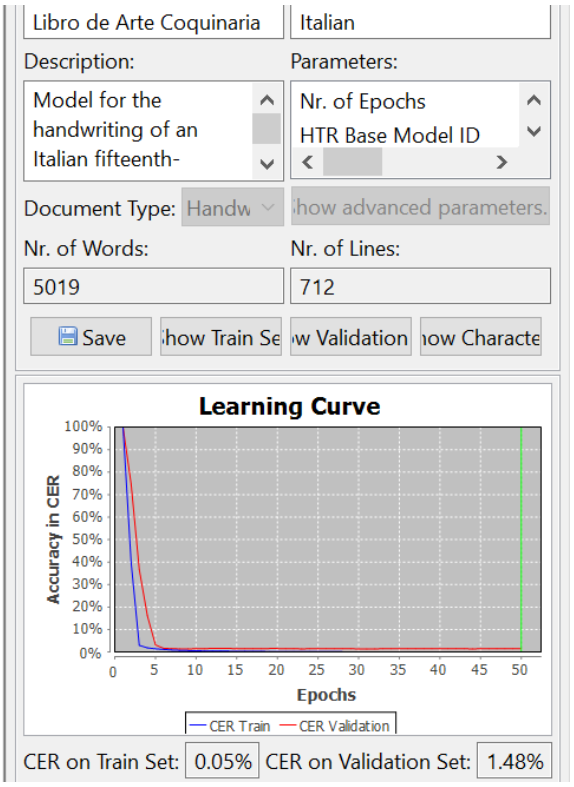


Figure 4 learning curve of the model created starting from the base model.

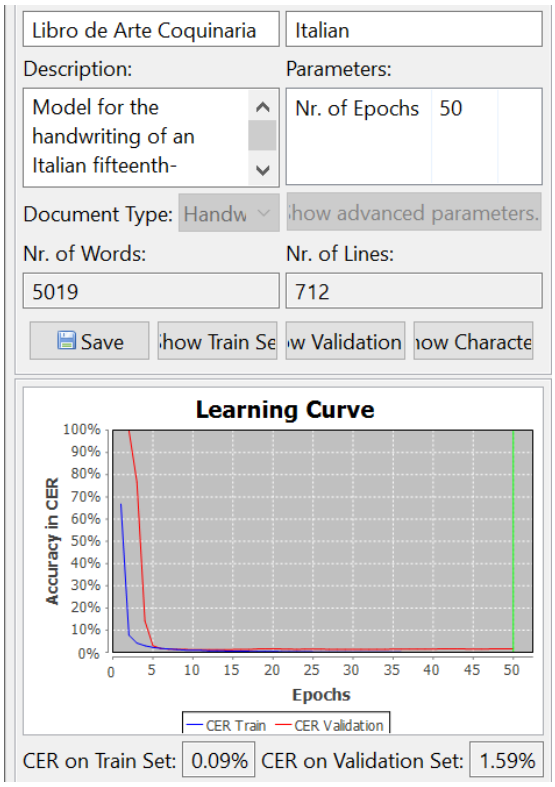


Figure 5 learning curve of the model created from scratch.

[20] If not, some additional training material is probably needed.
[21] See *Figures 4, 5.*
[22] Muehlberger et al. 2019.

3.2. Comparing models' accuracy.

For further analysing the accuracy of both models, we decided to test them on a specific page and compare their performances. We thus transcribed a new page, set it to *Ground Truth* and used the *Compare* function provided by Transkribus to compare the manually transcribed page (*Reference*, correct text) with the same page transcribed by both models (*Hypothesis*).

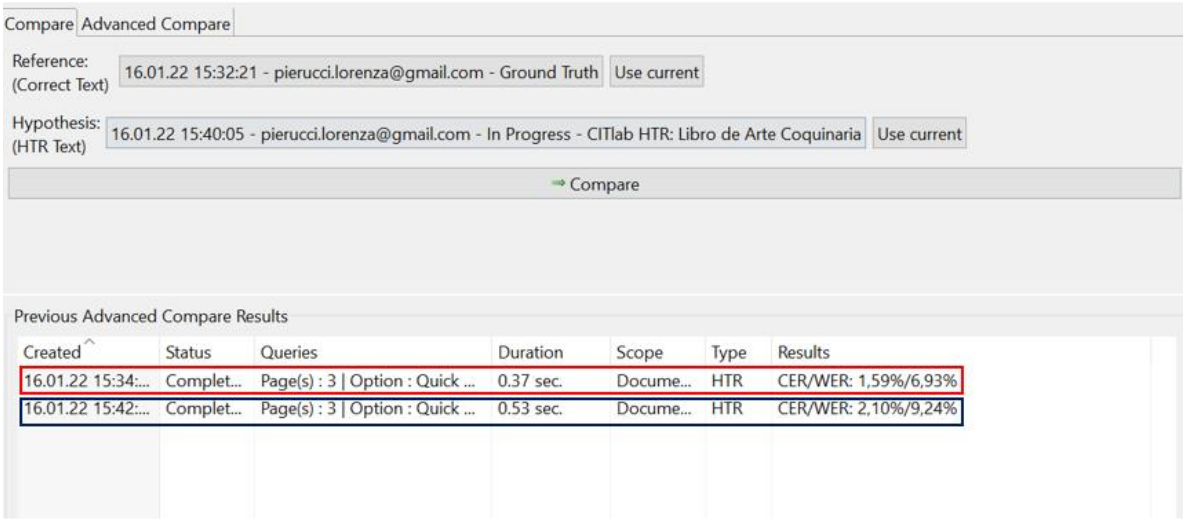


Figure 6. Comparison of the accuracy of both models on the same page.

The comparison confirmed and reflected the trend shown by the learning curve: the model trained with a base (*Fig.6, red box*) performed slightly better — CER and WER[23]of 1,59% and 6,93% respectively — than the model without base (*Fig.6, blue box*, CER/WER: 2,10%/9,24%).

The model-with-base's total character accuracy was thus of 97,77% against the 97,15% model-without-base's one. The total word accuracies were of 89,27% and 86,2 %, respectively. These results confirmed the ones found on the validation set and the CER is still much lower than the critical threshold of 5%, characterizing these HTR models as good.

[23]Word Error Rate (WER) is the number of errors divided by the total words. Word Error Rate = (Substitutions + Insertions + Deletions) / Number of Words. See more at https://en.wikipedia.org/wiki/Word_error_rate

3.3. Errors.

The quantitative metrics we used until now provide a measure of how the models perform but do not account for the *nature* of errors. For this reason, we compared the text versions generated by both models: as shown by the images below, the mistakes for the model with base mainly concerned spacing and punctuation, with just a few spelling mistakes and one non-detected abbreviation (*Figure 7,8: line 1-17, fano* instead of *faño*). The model without training made more misspelling errors *(Figure 7, lines: 1-2, 1-3, 1-7, 1-24, 1-25*). Generally speaking, though, errors involved minor changes not compromising the legibility of the page.

1-1 # ponilo in un pocha dacqua calda, & ī ~~breve~~ breve
1-2 # spatio serra deliguato, et converso ~~iesalimo~~ in salimo-
1-3 # ra, laquale ~~comesia rasertata~~ come sia rasectata se potra ~~ponerr~~ poner
1-4 # nel caldaro pianamente, azio che non ~~vō~~ ve
1-5 # andassi la terra, che serra sul fondo: et
1-6 # se la carne fusse vecchia et dura, spetial-
1-7 # mente Cappone et ~~Gallma,~~ Gallina, cavala parechie
1-8 # volte dal aqua bollente, et rinfredala nel
1-9 # aqua fresca, et in q⁀sto modo serra piu bella
1-10 # et piu presto cotta:
1-11 # per fare ogni bello arrosto:
1-12 # ~~P er~~ Per fare bello arrosto de pollastri, de ~~capponi~~ capponi,
1-13 # de Capretti, ó de qualunche altra carne ch̄
1-14 # meriti essere arrosta: prima se fosse carne
1-15 # grossa, fagli trare un boglio, excepto se fosse
1-16 # de vitello giovine, et poi lardala, come se
1-17 # ~~fano~~ faño li arrosti: ~~Se~~ se fosse Cappone, Fasano, pol-
1-18 # lastro, capretto, ó qualunchaltra carne, ch̄
1-19 # meriti arrosto, fa ch̄ sia ben netta, et polita,
1-20 # poi mettila in aqua bollente, et subito cavala
1-21 # fore, et ponila in aqua freda, et questo se
1-22 # fa, azio ch̄ sia piu bella, & meglio se possa
1-23 # ~~conciare~~ conciare: poi lardala, zio, e, con lardo bactuto,
1-24 # & altre chose ~~convemenli~~ convenienti odorifere onta bn̄,
1-25 # secondo el gusto del tuo ~~gignore.:~~ signore: et drento se
1-26 # te piace gli poni de bone herbe con prune ~~secche~~ secche,
1-27 # marasche, et viscioli, ó in tempo, ~~delagresto:~~ delagresto,
1-28 # et altre chose simile: poi mittila ~~ordinata:~~ ordinata-
1-29 # mente nel speto, et ponila al foco, et daglilo
1-30 # nel principio ad ascio ad ascio, perche sia bello
1-31 # & bono arrosto se deve cocere pian ~~piano,~~ piano: et
2-1 # ~~ca. 2z°.~~ .a.2

Figure 7 in green the GT, in red the errors of the transcription of the model created from scratch.

1-1 # ponilo in un pocha dacqua calda, & ī breve
1-2 # spatio serra deliguato, et converso ~~inesalimo~~ in salimo-
1-3 # ra, laquale ~~comesia~~ come sia rasectata se potra ~~ponerr~~ poner
1-4 # nel caldaro pianamente, azio che non ve
1-5 # andassi la terra, ~~chesserra~~ che serra sul fondo: et
1-6 # se la carne ~~susse~~ fusse vecchia et dura, spetial-
1-7 # mente Cappone et ~~ballma,~~ Gallina, cavala parechie
1-8 # volte dal aqua bollente, et rinfredala nel
1-9 # aqua fresca, et in q⁀sto modo serra piu bella
1-10 # et piu presto cotta:
1-11 # ~~Per~~ per fare ogni bello arrosto:
1-12 # Per fare bello arrosto de pollastri, de ~~capponi~~ capponi,
1-13 # de Capretti, ó de qualunche altra carne ch̄
1-14 # meriti essere arrosta: prima se fosse carne
1-15 # grossa, fagli trare un boglio, excepto se fosse
1-16 # de vitello giovine, et poi lardala, come se
1-17 # ~~fano~~ faño li arrosti: ~~Se~~ se fosse Cappone, Fasano, pol-
1-18 # lastro, capretto, ó qualunchaltra carne, ch̄
1-19 # meriti arrosto, fa ch̄ sia ben netta, et polita,
1-20 # poi mettila in aqua bollente, et subito cavala
1-21 # fore, et ponila in aqua freda, et questo se
1-22 # fa, azio ch̄ sia piu bella, & meglio se possa
1-23 # conciare: poi lardala, zio, e, con lardo bactuto,
1-24 # & altre chose convenienti odorifere onta bn̄,
1-25 # secondo el gusto del tuo signore: et drento se
1-26 # te piace ~~guponi~~ gli poni de bone herbe con prune ~~secche~~ secche,
1-27 # marasche, et viscioli, ó in tempo, ~~del agresto,~~ delagresto,
1-28 # et altre chose simile: poi mittila ~~ordinata.~~ ordinata-
1-29 # mente nel speto, et ponila al foco, et daglilo
1-30 # nel principio ad ascio ad ascio, perche sia bello
1-31 # & bono arrosto se deve cocere pian piano: et
2-1 # ~~a. R.~~ .a.2

Figure 8 in green the GT in red the errors of the transcription of the model with the base

4. Export.

To export a transcription, Transkribus offers the possibility to do it in several standard formats: as a Transkribus document, to allow a user to directly load the document in its own collection; as a PDF file, where the text can be extracted, copied and pasted from the image; as a TEI document in format XML; as a DOCX or TXT file, and as three formats for exporting only the tags. Since we used very few tags, we only exported our work in PDF, TEI, DOCX, TXT as well as in a Transkribus repository.

Unfortunately, the model in itself is only shareable to Transkribus users that we manually add to the collection. It is not available otherwise, since our small-scale annotation campaign — leading to our very specific model for this particular manuscript — is not destined to be one of the official Transkribus' publicly available models.

Conclusions.

We carried out a small-scale annotation campaign, trained two models using Transkribus, and evaluated their efficiency, comparing the performances of a model trained from scratch and of one relying on a base model.

The results — as discussed in §3.1 **Learning curve**— proved satisfactory, and the difference between the two models' performance did not result to be highly significant, probably due to the small size of the training set. Nonetheless, as we expected, the model trained with a base was generally more accurate.

Of course, our collection was small and homogeneous, which made it easier for the model to learn efficiently in comparison with collections composed of a variety documents and handwritings. We are indeed aware that by transcribing a larger number of pages for our ground truth — rather than the minimum required — we could have obtained even better results.

This experiment could be seen as a starting point for a general model for handwritten Italian of the end of 15th century and a tool to help for research on this topic. It would of course need a bigger set of documents from different topics and handwritings, but the workflow would essentially be the same.

Bibliography.

Bellarini, Luigi, "Introduction." *The Art of Cooking: The First Modern Cookery Book*, University of California Press, Berkeley, 2005.

Gagliardi, Gloria, "Inter-Annotator Agreement in linguistica: una rassegna critica". *Proceedings of the fifth italian conference on computational linguistics*, Torino: Accademia University Press, 2018.
10.4000/books.accademia.2802

Goodman, Leo A., and Kruskal, William H. "Measures of association for cross classifications". *Journal of the American Statistical Association*, vol. 49, 1954, pp.732–769.
https://doi.org/10.2307/2281536

Jenkins, Nancy Harmon. "Two Ways of Looking at Maestro Martino." *Gastronomica*, vol. 7, no. 2, Berkeley: University of California Press, 2007, pp. 97–103.
https://doi.org/10.1525/gfc.2007.7.2.97

Muehlberger, Guenter, et al. "Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study" *Journal of Documentation*, vol. 75, no. 5, 2019, pp. 954-976. https://doi.org/10.1108/JD-07-2018-0114.

Sacchi, Bartolomeo (Platina). *De honesta voluptate et valetudine*. Rome: Udalricus Gallus, 1474.

Shaw, Paul, "A note on calligraphy. *Libro de Arte Coquinaria by Maestro Martino. CD-ROM with notes and translation by Gillian Riley*, by Albala, K. Oakland: Octavo Editions, 2005.

Tognetti, Giampaolo, *Criteri per la trascrizione di testi medievali latini e italiani*, Roma: Quaderni della Rassegna degli Archivi di Stato, 1982.

Sitography.

https://rechtsprechung-im-ostseeraum.archiv.uni-greifswald.de/author/alverman/
https://emdigit.org/tool/2020/07/21/italian-administrative-hands.html
https://readcoop.eu/transkribus/howto/how-to-train-a-handwritten-text-recognition-model-in-transkribus/
https://readcoop.eu/transkribus/howto/transkribus-transcription-conventions/
https://en.wikipedia.org/wiki/Word_error_rate
https://www.youtube.com/watch?v=5YCfaFNMol4&t=6630s
https://ocr-d.de/en/gt-guidelines/trans/index.html