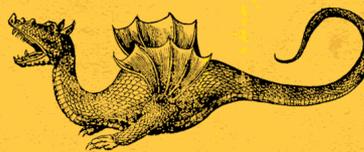




PROGETTO  
MAMBRINO

## HISTORIAS FINGIDAS



### El Proyecto Mambrino y las tecnologías OCR: estado de la cuestión

Stefano Bazzaco  
(Università di Verona)\*

#### Abstract

En el presente artículo se tratan los últimos resultados obtenidos por parte del Progetto Mambrino con las recientes tecnologías OCR. Se indican al respecto los resultados obtenidos con ABBY FineReader y Transkribus y se valoran las posibilidades de estas dos plataformas en relación con el corpus caballeresco italiano.

Palabras clave: libros de caballerías, Italia, ciclo de *Amadís de Gaula*, reconocimiento automático de los caracteres, OCR, HTR, libros antiguos impresos, Transkribus.

This article deals with the latest results obtained by the Mambrino Project with OCR technologies. The results obtained with ABBY FineReader and Transkribus platforms are discussed in relation to the Italian chivalric corpus.

Keywords: chivalric fiction, Italy, Amadis of Gaul's cycle, Optical Character Recognition, OCR, HTR, early printed materials, Transkribus.



#### Presentación: el Proyecto Mambrino

Los libros de caballerías españoles son *best-sellers* de la literatura cortesana de entretenimiento que se escribieron durante el Renacimiento y que gozaron de una enorme difusión en toda Europa. En países como Italia, Francia, Alemania, Inglaterra y Holanda se publicaron varias de estas obras, en forma de traducciones y continuaciones originales que durante los siglos XVI y XVII ocuparon las estanterías de librerías e

---

\* Este trabajo forma parte de las actividades realizadas por el Progetto Mambrino en el marco del Progetto di Eccellenza *Le digital humanities applicate alle Lingue e Letterature Straniere (2018-2022)* del Dipartimento di Lingue e Letterature Straniere de la Universidad de Verona.

impresores llegando a fascinar una enorme cantidad de lectores (Bognolo, 2012; Bognolo, Cara, Neri, 2013).

A pesar de ello, la importancia de estos textos ha sido rescatada sólo recientemente por parte de la crítica. En España, en las últimas décadas, el género ha sido revalorado a través del empeño de algunos importantes grupos de investigación, que utilizaron las nuevas tecnologías para la creación de valiosas herramientas, como bases de datos y ediciones digitales<sup>1</sup>. Italia fue el país donde los libros de caballerías de inspiración española fueron más apreciados durante (y más allá) de la época renacentista, llegando a conformar un género reconocible en su aspecto tipográfico. Sin embargo, tales novelas siguen siendo un patrimonio todavía por explorar.

Por esto, en 2003 nació el Progetto Mambrino, un grupo de investigación dirigido por Anna Bognolo y Stefano Neri, cuyo objetivo es estudiar el extenso corpus de los libros de caballerías italianos de derivación española. Entre los propósitos del proyecto se encuentra la necesidad de analizar los contenidos de las traducciones e imitaciones caballerescas publicadas en la península y de valorar el complejo panorama de producción y difusión de estos libros, censando y localizando los ejemplares que se conservan en bibliotecas y colecciones particulares.

En el año 2010, gracias a la colaboración con la Biblioteca Cívica de Verona, ha sido posible realizar un archivo digital que contiene las imágenes en alta definición de los ejemplares del entero ciclo de *Amadís de Gaula* italiano allí conservados, integrado con casi todas las traducciones y continuaciones. La distribución de estos materiales, en CD-ROM (<<http://www.quiedit.it/scheda.aspx?id=333>>) y descargables en pdf desde la web del proyecto, ha constituido el primer paso hacia la cooperación con otras bibliotecas nacionales e internacionales quienes han puesto a disposición también sus colecciones caballerescas. El trabajo de digitalización posibilitó además la transcripción de dos obras originales

---

<sup>1</sup> Se señala al respecto la base de datos bibliográfica «Clarisel» de la Universidad de Zaragoza, dirigida por Juan Manuel Cacho Bleuca y María Jesús Lacarra y consultable en la red al URL <<http://155.210.12.154/clarisel/index.htm>>.

italianas, A5: *Il secondo libro delle prodezze di Splandiano* (1564) y A10: *Aggiunta al Florisello. Le prodezze di don Florarlano* (1564), realizadas respectivamente por Paola Bellomi y Federica Colombini.

En 2013 el Progetto Mambrino publicó el *Repertorio delle continuazioni italiane ai romanzi cavallereschi spagnoli. Ciclo di Amadis di Gaula* (Bognolo-Cara-Neri, 2013), un volumen que contiene los resúmenes y los índices de personajes de los libros italianos de *Amadis*, a los que se antepone un estudio sobre diversos aspectos que atañen la recepción italiana y europea del género caballeresco.

A partir de 2015, gracias a esta labor previa y a la luz de las posibilidades que ofrecen las nuevas tecnologías, ha tomado cuerpo la idea de realizar una edición científica digital, o *Digital Scholarly Edition*, de las obras italianas que componen el ciclo de *Amadis de Gaula* (más de veinte textos: véase la base de datos «Spagnole Romanzerie – Ciclo di Amadis di Gaula» < <http://www.mambrino.it/spagnole/amadis.php> >).

Muchas son las ventajas relacionadas con la edición en formato digital de los textos caballerescos. Por un lado, la realización de una edición científica digital permite la libre consultación de las imágenes en alta calidad, fácilmente navegables en la red por parte de estudiosos y lectores; mientras que por otro permite la visualización de unas transcripciones de las obras, cuyos contenidos podrán ser investigados con modernos instrumentos de minería de textos y puestos en relación con otros materiales textuales y multimediales. En otras palabras, el intento es facilitar unos objetos digitales de libre consultación de los cuales se pueda extraer una gran cantidad de informaciones, que difícilmente se podrían sacar de manera inmediata y sincrética a partir de una obra en papel. Este último aspecto es propiamente el que califica las ediciones científicas digitales frente a la mera publicación de las imágenes escaneadas en formato digital, puesto que, como afirma Sahle, «a digitised edition is not a digital edition», porque «a digital edition cannot be given in print without significant loss of content and functionality» (2016, 27).

El camino hacia la creación de unas ediciones científicas digitales es un proceso largo, compuesto por varias fases. La primera de ellas, representada por el inicial trabajo de digitalización que ya comentamos, constituyó un primer avance con respecto al proceso para mejorar la

accesibilidad de las obras caballerescas italianas. Se están llevando a cabo acuerdos con varias bibliotecas del territorio (la Biblioteca Civica de Verona, la Biblioteca Bertoliana y la Biblioteca La Vigna de Vicenza, la Biblioteca Marciana y la Biblioteca del Museo Correr de Venecia, junto con bibliotecas extranjeras como la Biblioteca Nacional de España) para impulsar un programa de digitalización en alta resolución de los materiales caballerescos conservados, que esperamos poner al alcance de los estudiosos.

En 2010 se emprendió el proceso de transcripción manual de las obras que componen el ciclo *amadisiamo*, para realizar una edición moderna que facilitara la indexación de los contenidos de estos libros. Sin embargo, los textos son muy extensos, ya que en general superan los 800 folios, por consiguiente la transcripción manual resultaba un trabajo muy costoso en términos económicos y de tiempo: se pasó pues a considerar las posibles opciones para transcribir de modo automático o semiautomático tales obras, mirando a las posibilidades que ofrecían las nuevas técnicas de Optical Character Recognition (OCR), o reconocimiento óptico de los caracteres.

## **El problema de la transcripción automática de materiales históricos**

Los softwares de OCR son herramientas específicamente diseñadas para la obtención de unas transcripciones a partir de las imágenes escaneadas de las fuentes en papel. Se trata pues de instrumentos capaces de convertir los signos gráficos contenidos en una imagen digital en un texto procesable por parte de la máquina, lo cual en algunos casos puede aligerar o sustituir los métodos de transcripción manual.

Sin embargo, dependiendo de la época de publicación de las fuentes, las transcripciones generadas con herramientas de OCR están lejos de ser exentas de errores.

Con respecto a los textos contemporáneos, los problemas parecen casi completamente solucionados, ya que hay varias tecnologías (comerciales y de libre acceso) capaces de ofrecer muy buenos resultados de transcripción automática. Normalmente, tales instrumentos de OCR di-

viden la página escaneada en zonas, cada una correspondiente a una letra, y buscan una similitud con unos inventarios de distintos *fonts* prototípicos ya aprendidos por parte de la máquina. Los procesos de reconocimiento automático funcionan de manera apropiada con estas obras por varios motivos, como la separación uniforme entre las letras, el fondo blanco de la página, la concreta semejanza entre los *fonts* y los caracteres impresos, un léxico más estandarizado, etc. (Sprigmann y Lüdeling, 2017).

Por otra parte, con los materiales históricos impresos, que se publicaron desde el siglo XV hasta principios del siglo XX, los procedimientos de OCR no parecen adecuados, puesto que varios problemas pueden complicar el proceso de reconocimiento automático. Las razones de la insatisfactoria aplicación de los métodos de OCR a los documentos antiguos dependen de muchos factores: los problemas más evidentes están relacionados con el estado de las fuentes y la calidad de las fotoreproducciones; sin embargo, la cuestión es más bien de carácter tipográfico, ya que en tales obras se emplean unas letras muy heterogéneas, en muchas ocasiones difíciles de descifrar por parte del ordenador, que alteran toda la operación de transcripción automática<sup>2</sup>.

Las fuentes italianas del ciclo de *Amadís de Gaula* producen estos mismos inconvenientes. En primer lugar, por ser ediciones “de bolsillo” en formato octavo, presentan varios grados de deterioro, derivados del paso del tiempo, del constante manejo y de las condiciones de preservación (Figura 1). Pero más en general, los textos en cursiva exhiben varios caracteres especiales, como abreviaturas, ligaduras, signos tironianos, que ponen en dificultad los comunes instrumentos de reconocimiento automático (Figura 2).

Si los problemas relativos al estado de las fuentes se pueden parcialmente solucionar en fase de reproducción, a través de unos programas que ayudan a regularizar las transparencias y los efectos de iluminación impropia de las imágenes, la interpretación errónea de los

---

<sup>2</sup> Un listado de estos factores se encuentra en Barbuti y Caldarola (2013, 262); Sprigmann y Lüdeling (2017). Al respecto véase también Berg-Kirkpatrick y Klein (2014, 118).

elementos textuales especiales por parte del ordenador representa un concreto obstáculo para cualquier tecnología de OCR.

Una vez obtenidas unas imágenes fácilmente procesables por parte de la máquina, se empezó entonces una fase de experimentación consistente en valorar la eficacia de las herramientas de OCR actualmente disponibles en relación con las obras de nuestro corpus.

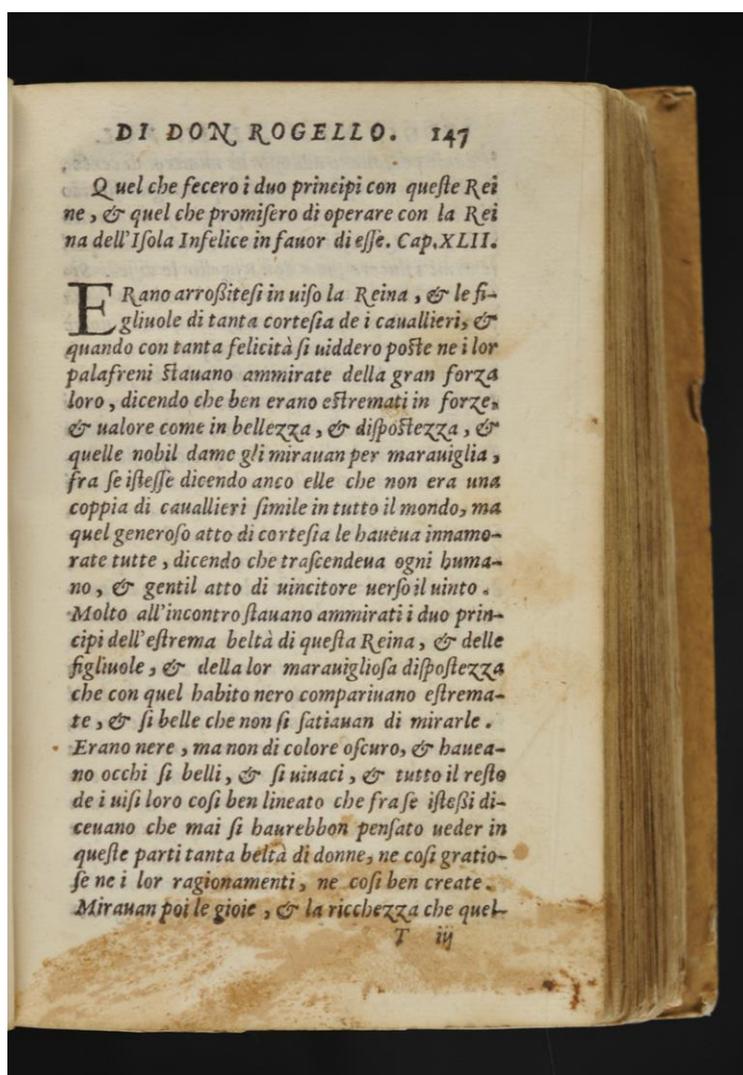


Figura 1. Manchas y desgastes en una página de la *Aggiunta a Rogello di Grecia*, 1564 (Verona, Biblioteca Civica, Cinq. 350-15).

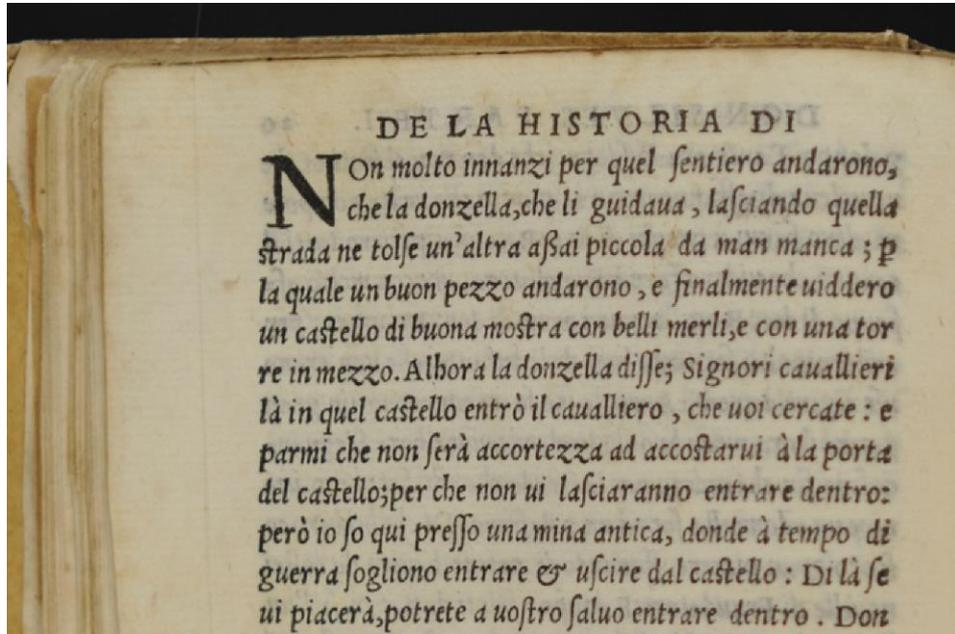


Figura 2. Ligaduras, abreviaturas y signos tironianos en el *Don Silves della Selva*, 1551 (Verona, Biblioteca Civica, Cinq. 350-16).

## La transcripción semi-automática de los libros de caballerías italianos

El trabajo de transcripción automática de las obras del corpus es fruto de una experimentación iniciada en el año 2016.

En un primer momento, algunos intentos de transcripción automática han sido llevados a cabo con ABBYY FineReader, un software propietario muy conocido, principalmente empleado para el reconocimiento de textos impresos. El programa, obviamente, resultó inadecuado al momento de interpretar signos tironianos y ligaduras, por lo cual, se decidió entrenar la plataforma a través de la identificación de unos patrones, para asignarle unos precisos valores a las distintas realizaciones de cada letra o grupo de letras. Los resultados del entrenamiento permitieron generar unas transcripciones más fiables, pero todavía inaceptables para nuestros propósitos, porque requerían un ingente trabajo de correc-

ción en fase de postproducción<sup>3</sup>.

Pasamos entonces, en tiempos más recientes, a la adopción de otro software apto al reconocimiento y la transcripción de textos llamado Transkribus, un programa *open access* creado y distribuido por el proyecto europeo READ (*Retrieval and Enrichment of Archival Documents*) y específicamente desarrollado para la transcripción de textos manuscritos (<<https://transkribus.eu/Transkribus/>>). Desde el primer momento, los resultados obtenidos con esta herramienta han sido muy prometedores también con respecto a nuestras obras impresas, porque el programa gracias a la aplicación de un sistema de HTR (*Handwritten Text Recognition*)<sup>4</sup> y unos *recurrent neural networks*<sup>5</sup> ha permitido solucionar los problemas derivados de las abreviaturas y ligaduras de la cursiva.

Transkribus emplea unos algoritmos que se sirven de materiales textuales ya transcritos (llamados técnicamente *groundtruth*) para el propio entrenamiento del software; con la consecuencia de que las transcripciones resultantes pueden ser gradualmente implementadas por medio de la identificación de unos patrones recurrentes que consienten la obtención de unos textos cada vez más fiables.

En concreto, el *training* está compuesto por distintas fases, que comentamos a continuación.

- 1) La primera fase corresponde a la preparación de los materiales digitalizados. Al respecto, se han subido a la plataforma las fotoreproducciones de distintas obras caballerescas italianas: primero las páginas digitalizadas en alta resolución de algunas *princeps* conservadas en la Biblioteca Cívica de Verona, y sucesivamente otras imágenes sacadas de la web, que comprendían materiales escaneados por las bibliotecas (italianas y extranjeras) y otros sacados directamente de Googlebooks, dado que para esta primera etapa puramente técnica de experimenta-

---

<sup>3</sup> A Tiziana Mancinelli debemos una descripción de los momentos significativos de este proceso (2016, 256-258).

<sup>4</sup> Para una evaluación de las características de los sistemas de HTR véase Juan et al. (2010).

<sup>5</sup> Los avances técnicos en el reconocimiento automático (*recurrent neural networks*, *LSTM technologies*, etc.) se discuten en Breuel et al. (2013, 683-685).

ción era indiferente la elección de los textos-base para la edición. Todos estos materiales tenían una calidad muy variable: algunas fotoreproducciones en colores tenían una buena resolución; otras eran en blanco y negro y muy defectuosas (Tabla 1). Sin embargo, todas juntas constituían un extenso campo de aplicación útil para valorar las posibilidades que ofrecía la nueva herramienta.

	<b>Obra</b>	<b>Fuente</b>	<b>Imágenes</b>
<b>1-4</b>	AMADIS DI GAULA, 1546	Berkeley, University of California, Bancroft Library, PQ 6275.I2 1547	b/n
<b>A.10</b>	AGGIUNTA AL FLORISELLO (LE PRODEZZE DI DON FLORARLANO), 1564	Verona, Biblioteca Civica, Cinq. E 350-12	color
<b>A.11</b>	AGGIUNTA A ROGELLO DI GRECIA, 1564	Verona, Biblioteca Civica, Cinq. 350-15	color
<b>12</b>	SILVES DE LA SELVA, 1551	Verona, Biblioteca Civica, Cinq. 350-16	color
<b>13.1</b>	SFERAMUNDI. PRIMA PARTE, 1558	Madrid, Biblioteca Nacional de España, 5-4978	b/n
<b>13.2</b>	SFERAMUNDI. SECONDA PARTE, 1560	Wien, Österreichische Nationalbibliothek, 40.J.16 (Vol. 19) - googlebooks	b/n
<b>13.3</b>	SFERAMUNDI. TERZA PARTE, 1563	Wien, Österreichische Nationalbibliothek, 40.J.16 (Vol. 20) - googlebooks	b/n
<b>13.4</b>	SFERAMUNDI. QUARTA PARTE, 1563	München, Bayerische Staatsbibliothek, P.o.hisp. 105 k-4.	b/n
<b>13.5</b>	SFERAMUNDI. QUINTA PARTE, 1565	Wien, Österreichische Nationalbibliothek, 40.J.16 (Vol. 22) - googlebooks	b/n

Tabla 1. Listado de las obras elegidas para el entrenamiento de Transkribus.

- 2) Una vez subidas las imágenes al *cloud* de Transkribus, ha empezado la propia fase de entrenamiento del software, a través de la aplicación simultánea de sistemas de HTR y de OCR dentro de la misma plataforma de Transkribus. Esta fase se articula a su vez en tres momentos distintos.

Primero, en la operación de *Layout Analysis* (LA), el programa fracciona automáticamente la página digitalizada en distintas zonas. La más extensa corresponde a la caja de texto, mientras que las otras zonas corresponden a las líneas del texto contenido en la imagen (Figura 3). Entonces, Transkribus necesita conocer la correcta transcripción de una parte del texto. Por esta razón, es necesario transcribir una veintena de páginas manualmente (*groundtruth*), de manera que a cada línea de la imagen corresponda una sola línea del texto transcrito. Aprovechando tal correspondencia, la plataforma entrena al algoritmo para que busque una correlación iterativa entre los signos presentes en la imagen y sus respectivas transcripciones, hasta llegar a una identificación entre líneas cada vez más precisa (Jander, 2016).

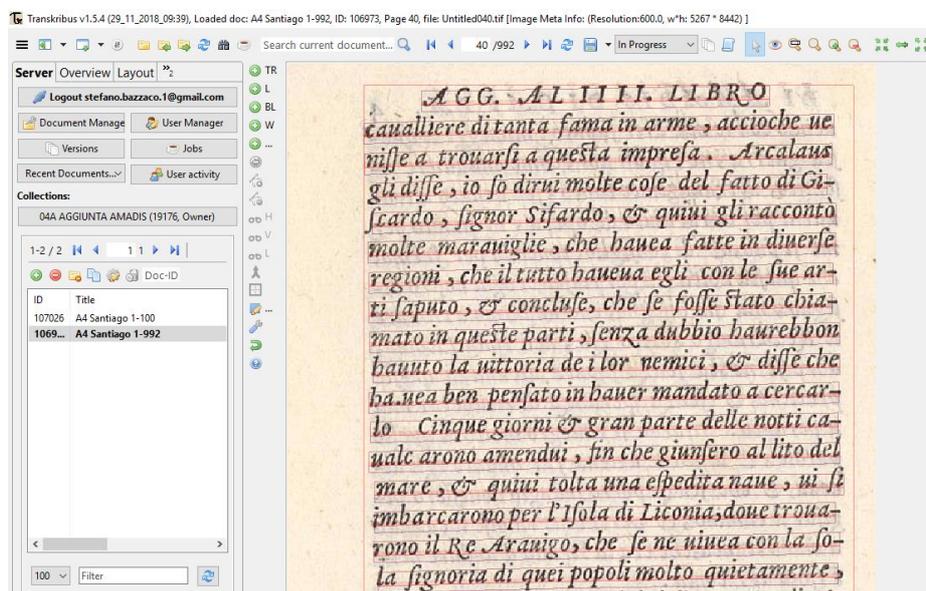


Figura 3. Un ejemplo de la pantalla de Transkribus después del proceso de Layout Analysis sobre la *Aggiunta al quarto libro di Amadis di Gaula*, 1563 (Santiago de Compostela, Biblioteca Universitaria, 13996).

3) A lo largo del proceso, la mayoría de las páginas transcritas manualmente constituye el Train Set, es decir el campo de aplicación del HTR, mientras que otras pocas páginas, agrupadas en el Test Set, son incluidas para valorar los resultados del entrenamiento.

En fin, el programa genera un modelo, cuya precisión se indica a través de unos índices llamados CER (*Character Error Rate*), que expresan en porcentaje la *edit distance* entre la transcripción correcta y la transcripción generada por el HTR. Por ejemplo, el 6 % de CER indica que mediamente, con respecto a las páginas transcritas, dentro de cien caracteres (espacios incluidos), 6 de ellos son erróneos.

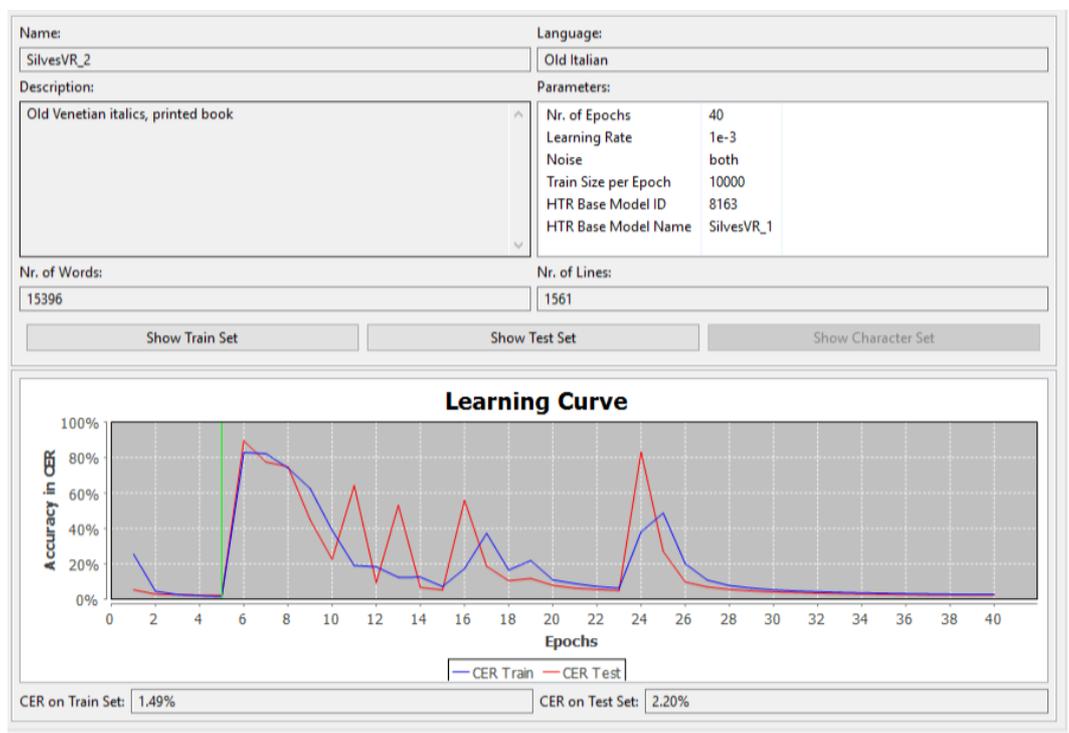


Figura 4. Modelo elaborado con Transkribus: en la parte inferior aparecen los índices del entrenamiento sobre el *Don Silves della Selva*, dozeno libro amadisiamo (Verona, Biblioteca Civica, Cinq. 350-16).

La primera obra elegida para el entrenamiento ha sido la *Segunda parte de Sferamundi*, decimotercero libro del ciclo de Amadís, publicado por Michele Tramezzino en Venecia en 1560. Desde el principio los resultados de reconocimiento automático obtenidos fueron muy buenos y nos empujaron a seguir con otros textos: en la siguiente tabla se señalan las obras del corpus que han sido seleccionadas para el entrenamiento, indicando por cada una la respectiva calidad de las imágenes y los índices de precisión (CER Train Set y CER Test Set) de los modelos elaborados:

	Obra	Calidad de las Imágenes	CER (Character Error Rate)
1-4	AMADIS DI GAULA, 1546	b/n media resolución	-
A.10	AGGIUNTA AL FLORISELLO (LE PRODEZZE DI DON FLORARLANO), 1564	color alta resolución	-
A.11	AGGIUNTA A ROGELLO DI GRECIA, 1564	color alta resolución	CER Train set: 0,66% CER Test set:
12	SILVES DE LA SELVA, 1551	color alta resolución	CER Train set: 0,49% CER Test set: 0,66%
13.1	SFERAMUNDI. PRIMA PARTE. 1558	b/n baja resolución	-
13.2	SFERAMUNDI. SECONDA PARTE. 1560	b/n baja resolución	CER Train set: 1,13% CER Test set: 1,21%
13.3	SFERAMUNDI. TERZA PARTE. 1563	b/n baja resolución	CER Train set: 1,61% CER Test set: 1,80%
13.4	SFERAMUNDI. QUARTA PARTE. 1563	b/n media resolución	CER Train set: 0,84% CER Test set: 1,11%
13.5	SFERAMUNDI. QUINTA PARTE. 1565	b/n baja resolución	CER Train set: 1,51% CER Test set: 1,59%

Tabla 2. Primeros resultados obtenidos con Transkribus.

## Discusión de los resultados y perspectivas futuras

Transkribus, plataforma destinada a la transcripción de textos manuscritos de difícil interpretación (por sus aspectos formales, como el *misalignment*, y textuales: inserciones, tachaduras, abreviaturas, etc.), permite también solucionar muchos de los problemas derivados de la composición tipográfica de la letra impresa. Los signos especiales y las ligaduras no representan un obstáculo para el sistema de HTR; las abreviaturas aparecen expandidas en el documento final; y hasta se ha podido entrenar el algoritmo para que aprenda básicas instrucciones filológicas y modernizaciones léxicas<sup>6</sup>.

Las transcripciones obtenidas en esta primera fase de experimentación nunca llegan a ser totalmente correctas; sin embargo, es tan bajo el número de errores por página que podemos confirmar que Transkribus es una valiosa herramienta para la transcripción de textos impresos en cursiva<sup>7</sup>. Puesto que con los textos manuscritos los índices de precisión del HTR llegan a un valor de CER comprendido entre el 30% y el 8%, se puede con razón afirmar que los resultados obtenidos con nuestras obras son mucho más alentadores, con valores de CER inferiores al 1%. Si consideramos además que las transcripciones dependen en buena medida de la calidad de las imágenes, la plataforma promete interesantes posibilidades, puesto que al momento no se han encontrado evidentes problemas de reconocimiento hasta con las imágenes más desgastadas.

---

<sup>6</sup> Es preciso señalar que tales transformaciones han sido llevadas a cabo de manera un poco forzada, puesto que Transkribus debería operar dentro de una relación entre signo gráfico y letra transcrita de 1:1. Las obras impresas con que tratamos, sin embargo, por ser bastante uniformes en sus aspectos tipográficos y textuales, han permitido extender en algunos casos los criterios de transcripción automática para reducir el sucesivo trabajo de postcorrección.

<sup>7</sup> Con respecto a los avances en las técnicas de postcorrección, se consideren las siguientes afirmaciones: «OCR engines of the latest generation employing recurrent neural networks with LSTM architecture lead to impressive results on OCR'ed historical documents over the complete history of modern printing, often reaching recognition accuracy on the character level around 95% and more. Yet, for many applications OCR'ed texts need to be close to perfect, which means that postcorrection of OCR results remains inevitable. Fully automated postcorrection only helps to improve low quality texts. When already starting from a baseline accuracy of 95%, only interactive postcorrection leads to substantial improvements of OCR'ed texts» (Fink, Schultz y Springmann, 2017).

En otras palabras, en el futuro será posible transcribir con buena aproximación, por medio de unos iniciales *trainings*, todas las obras del corpus, porque la variedad de las imágenes cargadas a la plataforma ha permitido ampliar el espectro de aplicación del *software*, que está acostumbrado ahora a realizaciones gráficas y tipográficas muy distintas. Los modelos generados durante este proceso son modelos individuales, es decir concebidos en relación con unos específicos ejemplares digitalizados y basados sobre unas transcripciones preconfeccionadas (*groundtruth*); no obstante, la variedad tipográfica que caracteriza los textos escaneados del corpus puede llevar también a la creación de unos modelos mixtos muy eficaces. Springmann y Anke, subrayando la necesidad de unos modelos de OCR mixtos más extensivos, tratan el problema de la siguiente manera:

The problem of applying OCR methods to historical printings is thus twofold. First one needs to train an individual model for a specific book with its specific typography. This can be achieved by transcribing some portion of the printed text, which usually requires linguistic knowledge. Second, even if this model works well for the book it has been trained on, it does not normally produce good OCR results for other books, even if their fonts look similar to the human eye. We need to overcome this typography barrier in order to use OCR methods effectively in the building of a historical electronic text corpus (Springmann y Anke, 2017).

Una vez obtenidas las transcripciones, el Progetto Mambrino se ocupará de la realización de una edición científica digital de cada libro del ciclo amadisiano. En una primera etapa, los textos serán indexados por medio de un ligero etiquetado TEI y puestos en relación con las imágenes de las fuentes originales y otros materiales digitales. Gracias al soporte de algunos estudiantes de la Universidad de Verona (Dipartimento di Lingue e Letterature Straniere) ha sido posible convertir al formato XML el índice completo de los nombres de los personajes del ciclo italiano de *Amadís* presente en el repertorio (Bognolo, Cara, Neri, 2013, 509-663); mientras que en futuro se prevé también la inclusión de unos resúmenes de los libros por capítulos y de un árbol genealógico de los personajes principales. Idealmente, un artefacto digital así concebido, cuyos conte-

nidos son completamente interoperables, demuestra cómo las herramientas digitales pueden afectar de manera cualitativa los estudios literarios, proporcionando un cambio significativo en la misma relación que subsiste entre los especialistas y sus objetos de estudio.



### Bibliografía citada

- Berg-Kirkpatrick, Taylor y Dan Klein, «Improved Typesetting Models for Historical OCR», *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Volume 2: Short Papers), Baltimore, Maryland: Association for Computational Linguistics, 2014, pp. 118-123.
- Bognolo, Anna, Giovanni Cara y Stefano Neri, *Repertorio delle continuazioni italiane ai romanzi cavallereschi spagnoli. Ciclo di Amadis di Gaula*, Roma, Bulzoni Editore, 2013.
- Bognolo, Anna, «Il romanzo cavalleresco spagnolo in Italia e la collezione di *Amadis* della Biblioteca Civica di Verona», in *Progetto Mambrino, Ciclo italiano di Amadis di Gaula. Collezione della Biblioteca Civica di Verona. Studi preliminari*, Verona, Quiedit, 2012, pp. 7-30.
- Breuel, Thomas, Adnan Ul-Hasan, Mayce Ali Al-Azawi y Faisal Shafait, «High-Performance OCR for Printed English and Fraktur Using LSTM Networks», *Second International Conference on Document Analysis and Recognition (ICDAR)*, 2013, pp. 683-687.
- Fink, Florian, Klaus U. Schulz, y Uwe Springmann, «Profiling of OCR'ed Historical Texts Revisited», *Proceedings of DATECH2017*, Göttingen, 2017, sin paginación.
- Jander, Melina, «Handwritten Text Recognition – Transkribus: A User Report», *The electronic Text Reuse Acquisition Project (eTRAP)*, Institute of Computer Science, University of Göttingen, Germany, 2016.

- URL < <http://www.etrapp.eu/transkribus-a-user-report/> > (último acceso: 05/12/2018).
- Juan, Alfons, Veronica Romero, Joan Andreu Sánchez, Nicolás Serrano, Alejandro Héctor Toselli, Enrique Vidal, «Handwritten Text Recognition for Ancient Documents», *Proceedings of the First Workshop on Applications of Pattern Analysis*, eds. Tom Diethe, Nello Cristianini, John Shawe-Taylor, 2010, pp. 58-65.
- Mancinelli, Tiziana, «Early printed edition and OCR techniques: what is the state-of-art? Strategies to be developed from the working-progress Mambrino project work», *Historias Fingidas*, 4 (2016), pp. 255-260.
- Sahle, Patrick, «What is a scholarly digital edition (SDE)?», *Digital Scholarly Editing. Theory, Practice and Future Perspectives*, eds. Matthew Driscoll y Elena Pierazzo. Cambridge: Open Book Publishers, 2016, pp. 19-39.
- Springmann, Uwe y Anke Lüdeling, «OCR of historical printings with an application to building diachronic corpora: A case study using the RIDGES herbal corpus», *DH Quarterly*, 11-2, 2017, sin paginación <<http://www.digitalhumanities.org/dhq/vol/11/2/000288/000288.html>> (último acceso: 08/12/2018)