

Taller de Inteligencia Artificial y Reconocimiento Automático de Texto Manuscrito para las Humanidades

Proyecto Flotas de la Nueva España
Diciembre 2024

Sesión de práctica con Transkribus

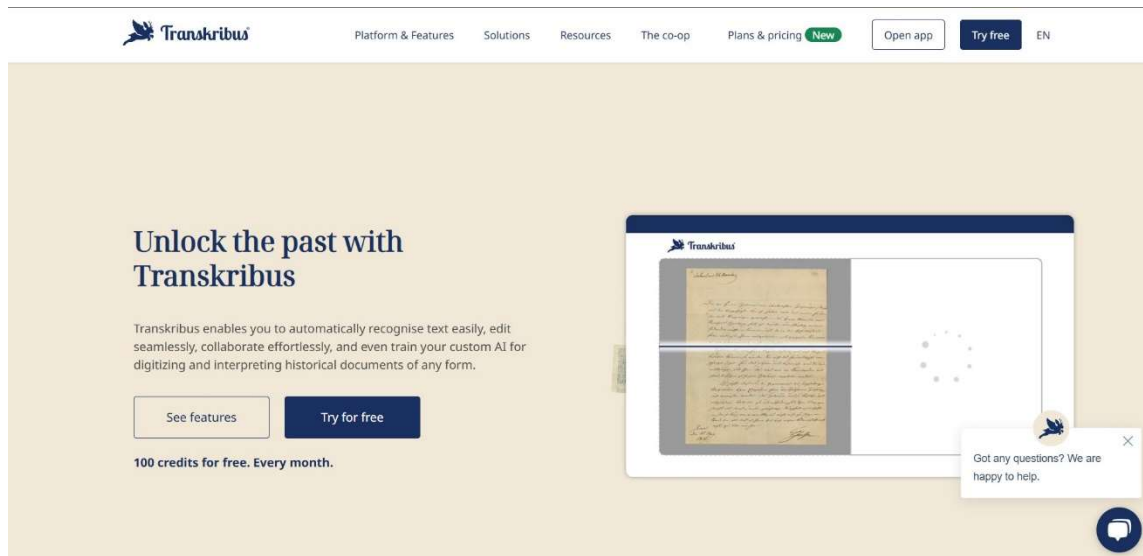
Contenido

INTRODUCCIÓN A TRANSKRIBUS	2
¿Qué es <i>Transkribus</i> ?	2
¿Cómo trabaja?	2
INICIO Y CONFIGURACIÓN	4
IMPORTACIÓN DE DOCUMENTOS	8
Editar metadatos	11
Compartir una colección	13
Búsqueda de documentos	14
INTERFAZ DE TRANSCRIPCIÓN	15
Control de versiones	17
TRANSCRIPCIÓN MANUAL	19
<i>LAYOUT ANALYSIS</i>	22
Limpieza del layout (diseño)	24
TRANSCRIPCIÓN AUTOMÁTICA	28
ANOTACIÓN	30
EXPORTACIÓN	32

INTRODUCCIÓN A TRANSKRIBUS

¿Qué es Transkribus?

Es una plataforma colaborativa que, impulsada por inteligencia artificial, está diseñada para facilitar la transcripción automática y manual de documentos históricos. Al ser colaborativa, se puede acceder a colecciones de documentos y modelos de diferentes usuarios. Además, la plataforma permite el entrenamiento de modelos personalizados para el reconocimiento de texto manuscrito e impreso. Para acceder a la plataforma es necesario ingresar a la siguiente dirección: <https://www.transkribus.org/>, dar clic en “Try free” y registrarse con un correo electrónico.



¿Cómo trabaja?

Transkribus es una herramienta que permite importar archivos históricos en formato de imagen para convertirlos en textos legibles por máquina (*machine-readable text*) a través de tres pasos:



Paso 1: Importación de documentos

Para importar documentos debemos asegurarnos de tener imágenes en alguno de estos cuatro formatos:

- JPEG/JPG de 10 MB
- PNG de 10 MB
- PDF de 200MB con un máximo de 3000 páginas
- IIIF manifest (formato de intercambio de imágenes que tienen algunas bibliotecas en línea)

Paso 2: Reconocimiento de documentos

En este paso, la computadora analiza qué parte de la imagen importada es texto. La forma de hacerlo es en dos etapas:

- 1) Análisis de la estructura (*layout analysis*), donde se divide el documento en regiones para después reconocer las líneas base o renglones y construir los polígonos de línea.
- 2) Reconocimiento de texto por medio de modelos de *Handwritten Text Recognition* (HTR), que convierten las imágenes en texto legible por máquina. Se pueden usar los modelos abiertos que tiene Transkribus en línea, o puedes generar tus propios modelos.

Paso 3: Exportación de documentos

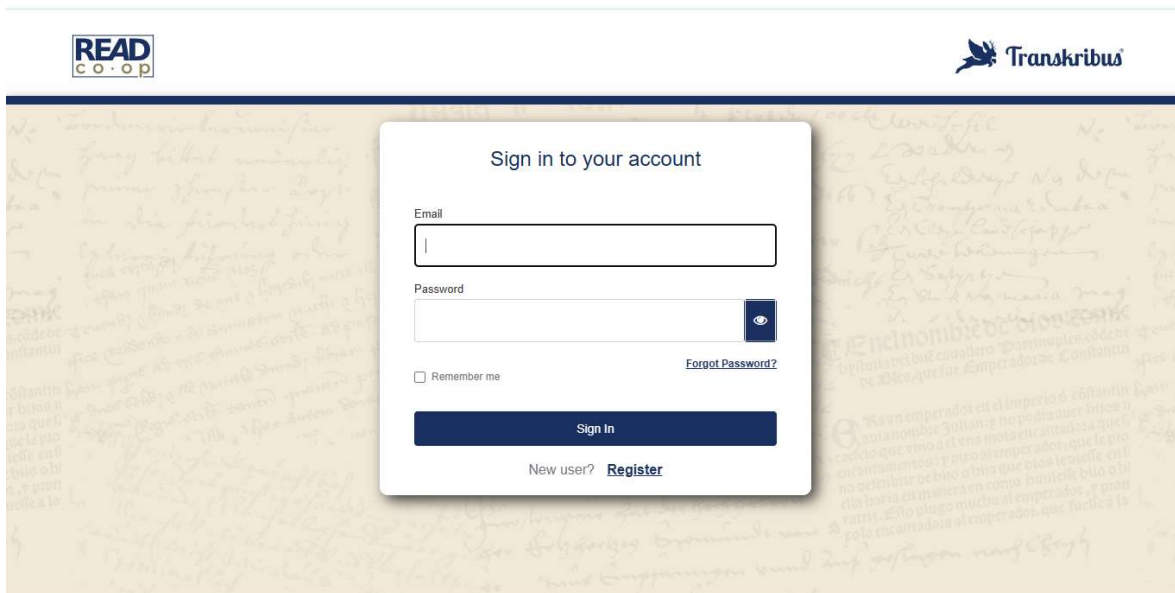
Transkribus te permite exportar cada página transcrita en diferentes formatos:

- Imagen: con diferentes tamaños, es decir, diferentes resoluciones.
- XML: junto con el texto, se incluyen las coordenadas de cada una de las regiones y los polígonos de los renglones, así como las anotaciones que hayamos incluido.
- Texto simple: en formato TXT o DOCX.

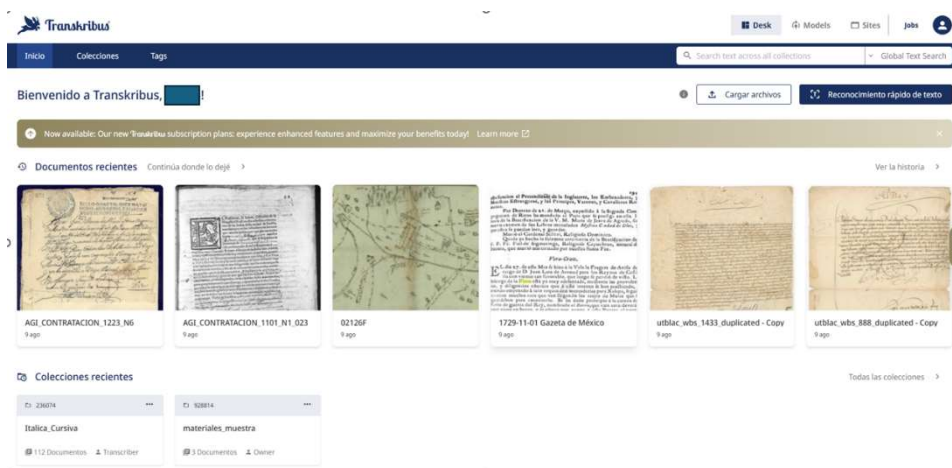
INICIO Y CONFIGURACIÓN

Para trabajar con Transkribus, siempre es preferible usar una computadora de escritorio o laptop, y no una tableta o teléfono, pues tanto la capacidad del equipo como el tipo de navegador pueden afectar el funcionamiento de la plataforma.

Ingresa en <https://app.transkribus.eu> con los datos que usaste para registrarte.



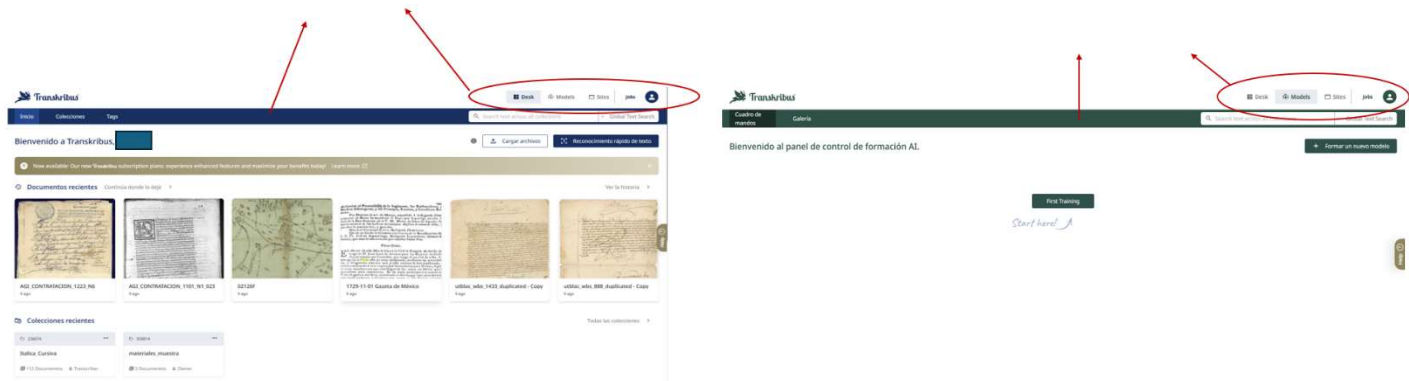
Una vez ingresados tus datos, verás una pantalla como la siguiente:



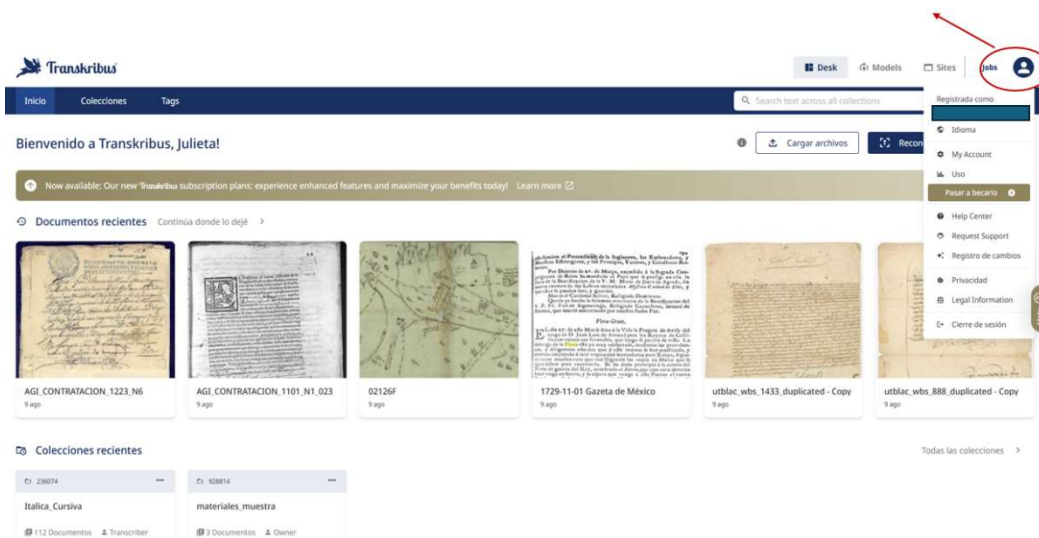
Aquí encontrarás accesos rápidos a los últimos documentos y colecciones con los que hayas estado trabajando.

En la parte superior derecha, se encuentran los accesos a las tres secciones principales de la plataforma: **Desk**, donde están tus documentos; **Models**, donde están los modelos de reconocimiento de texto y estructura de página; **Sites**, donde puedes crear tu propio sitio y publicar documentos.

Puedes guiarte y distinguir las tres secciones por los colores de sus encabezados, botones y letras. La sección Desk, donde están tus colecciones y documentos, es de color AZUL, mientras que la sección de Modelos, es en color VERDE.



En la esquina superior derecha, podrás ajustar algunas funcionalidades de la plataforma, como el idioma, según tus necesidades y preferencias, además de consultar las características de tu cuenta.



Transkribus funciona mediante un sistema de pago basado en créditos que permite acceder a ciertas funcionalidades de la plataforma. Los créditos son una unidad de moneda virtual. En la opción **Usage**, podrás ver los créditos que has usado. Si tu cuenta es individual, tendrás derecho a 100 créditos cada mes- De esta cantidad te irán descontando créditos en función del uso de los diferentes modelos disponibles en la plataforma. Por ejemplo, la transcripción automática de una página usando un modelo HTR cuesta un crédito. Pero no te preocupes, aunque el saldo se te acabe, Transkribus te abonará otros 100 créditos el siguiente mes. La importación/exportación de documentos, así como las transcripciones manuales, no consumen créditos.

The screenshot shows the 'Usage' page in Transkribus. At the top, there are navigation tabs: 'Inicio', 'Colecciones', and 'Tags'. Below this, the 'Uso y límites' section displays the current plan as 'Individual' and the remaining credit balance as '600'. A table titled 'Cuadro de utilización del crédito' shows a list of credit operations:

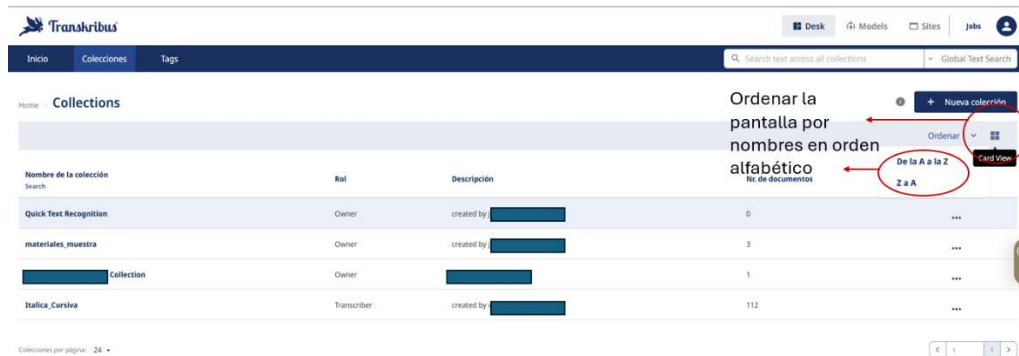
Descripción	Importe del crédito	Fecha	Expira en	Usuario	Acciones
Added by Admin '100 Free Credits'	+100	9 nov	9 dic	julieta1227@gmail.com	
Added by Admin '100 Free Credits'	+100	9 oct	9 nov	julieta1227@gmail.com	
Added by Admin '100 Free Credits'	+100	9 sept	9 oct	julieta1227@gmail.com	
Text Recognition (1 pages)	-1	9 ago	Never	julieta1227@gmail.com	...
Refund for failure in Text Recognition job (1.0 pages)	+1	9 ago	Never	julieta1227@gmail.com	...

En la sección **Desk**, el primer botón importante es el correspondiente a las **Colecciones**, donde encontrarás todos tus documentos. Las colecciones funcionan como las carpetas en tu computadora, con la diferencia de que las colecciones de Transkribus no pueden tener subcolecciones (subcarpetas).

The screenshot shows the 'Collections' page in Transkribus. The navigation bar at the top has 'Inicio', 'Colecciones', and 'Tags'. The 'Colecciones' tab is highlighted. Below the navigation bar, there is a search bar and a '+ Nueva colección' button. The main content area displays a list of collections:

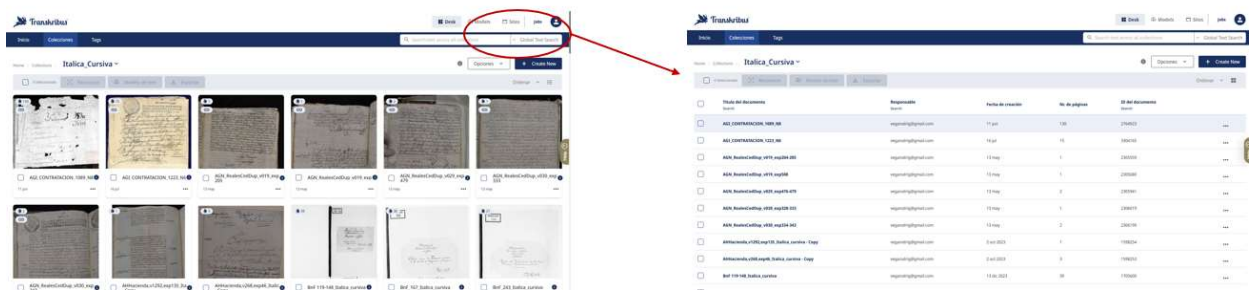
- Quick Text Recognition: 0 Documentos, Owner
- materiales_muestra: 3 Documentos, Owner
- Italica_Cursiva: 112 Documentos, Transcriber

Puedes cambiar la forma de visualizar tu pantalla de colecciones como se indica en la siguiente imagen:

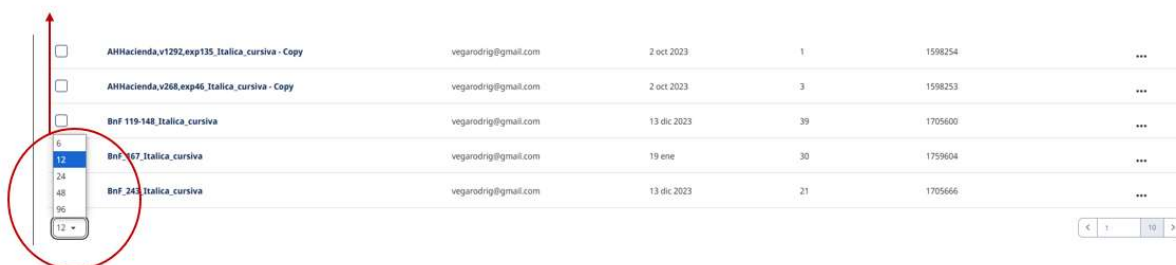


Esto te permitirá visualizar en forma de listado tus colecciones. También podrás organizarlas en orden alfabético para localizarlas fácilmente.

De igual forma, al ingresar a cualquier colección, los documentos se podrán visualizar tanto en forma de imágenes como de listado.



Al final de la página, podrás modificar el número de imágenes que ves por página, como se muestra a continuación:



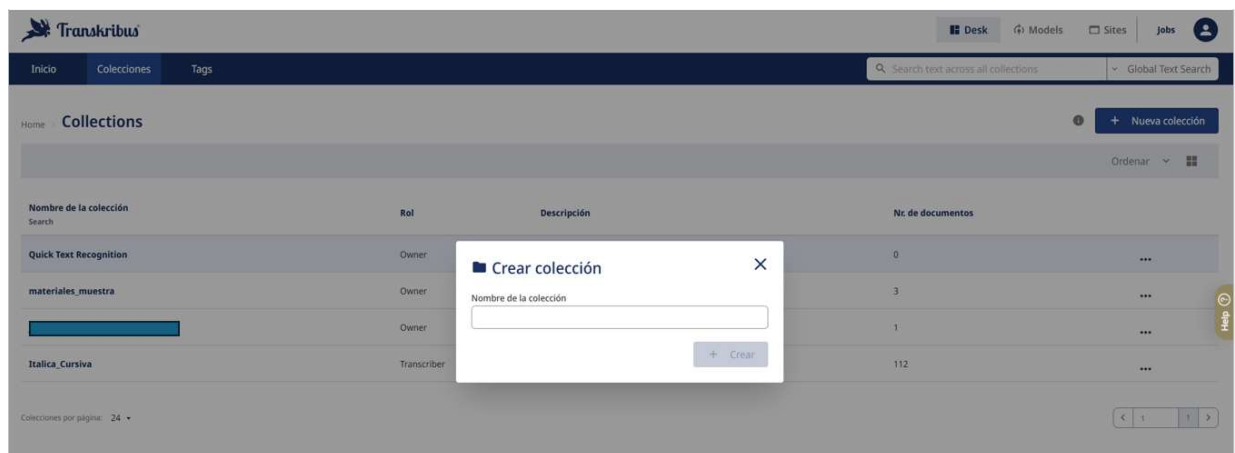
IMPORTACIÓN DE DOCUMENTOS

Ejercicio: Crea una colección e importa documentos.

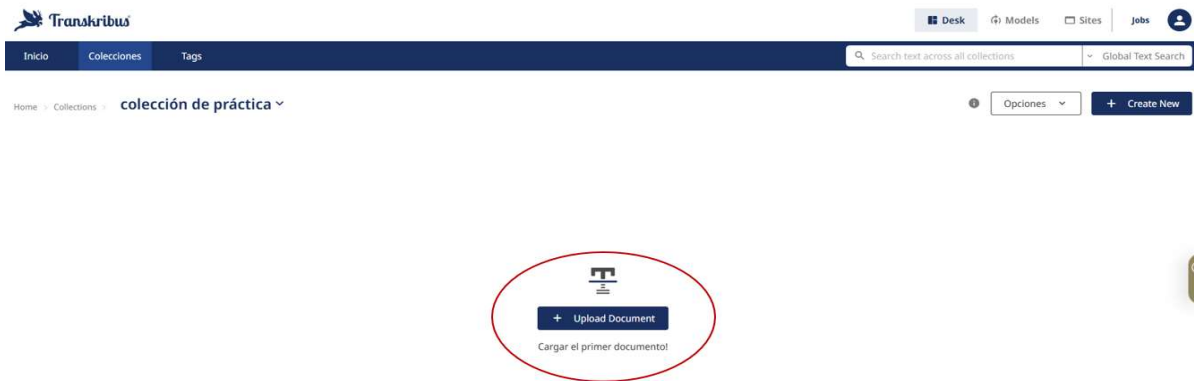
1. Puedes agregar documentos a colecciones nuevas o ya existentes. Para crear una nueva colección, necesitas ubicarte en la pantalla de colecciones y dar clic en el botón azul **“+Nueva Colección”**.



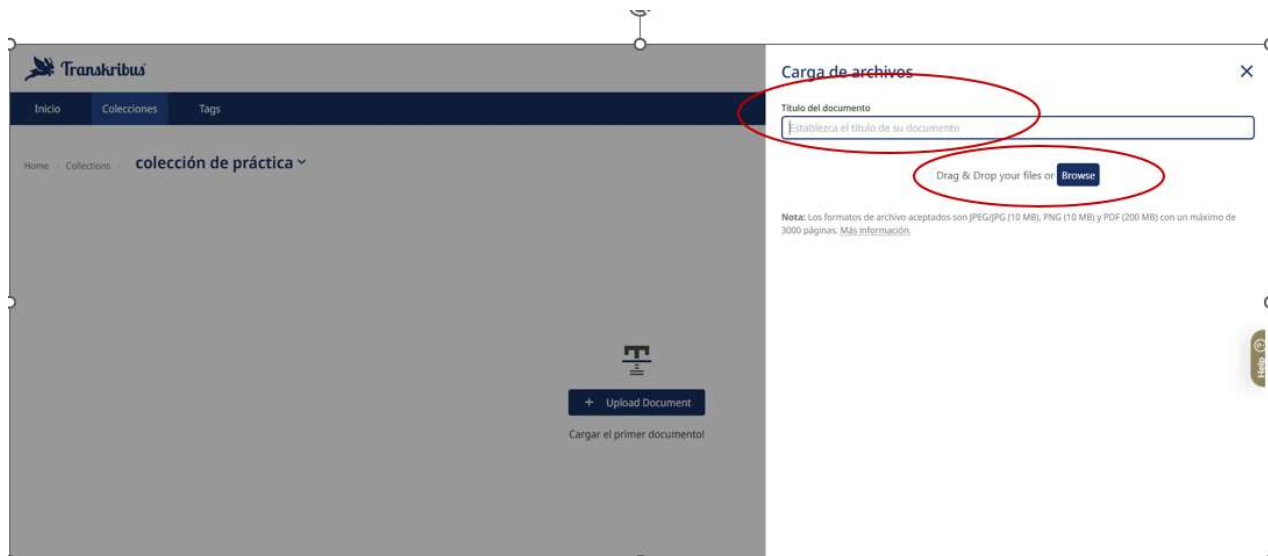
2. Una vez presionado el botón, te aparecerá el recuadro donde podrás ponerle nombre a tu nueva colección. En este caso, la nombraremos **“colección de práctica”**.



3. Al dar clic en el botón **“Crear”**, se creará y abrirá la nueva colección. Te aparecerá la siguiente pantalla, donde podrás subir documentos dando clic en el botón **“Upload Document”**.



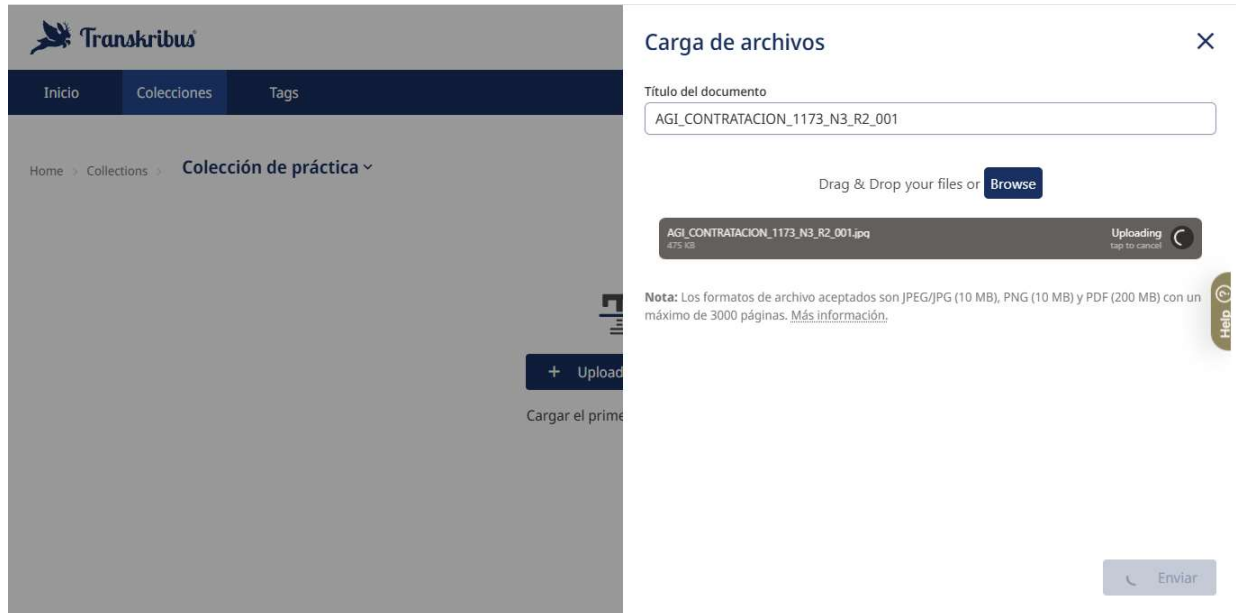
4. Del lado derecho de la pantalla, aparecerá la ventana de importación de imágenes. Puedes poner nombre al nuevo documento o dejarlo en blanco. Si lo dejas en blanco, la colección se nombrará automáticamente de acuerdo con el nombre de la primera imagen que subas. Lo importante es cargar el(los) archivo(s), ya sea arrastrándolo(s) o por medio de la opción “Browse”.



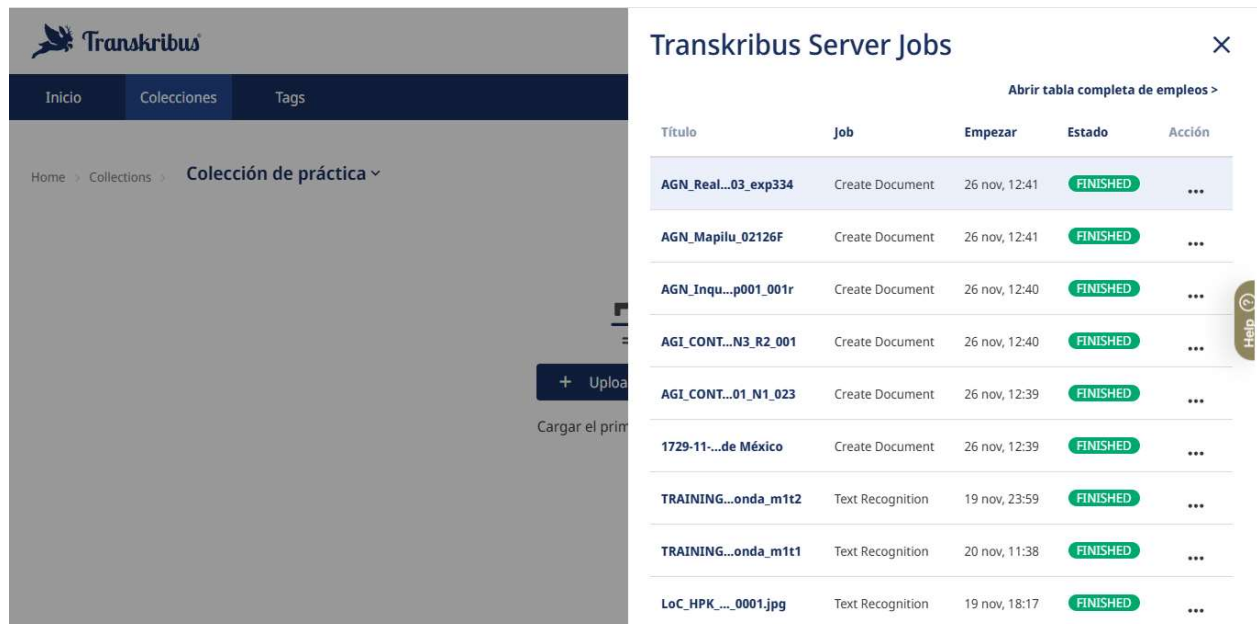
Es importante tener en cuenta que las imágenes se deben cargar con la orientación adecuada, pues la orientación del texto dentro de una imagen afecta a su reconocimiento y no es posible rotar las imágenes dentro de Transkribus. Antes de cargar tus imágenes, revisa si en alguna de ellas el texto tiene una orientación incorrecta y róta la hasta que el texto esté horizontal.

5. Agrega a la colección, **uno por uno**, los documentos que encontrarás en el [blogpost de materiales del curso \(https://wp.lancs.ac.uk/newspainfleets/?p=1126\)](https://wp.lancs.ac.uk/newspainfleets/?p=1126):

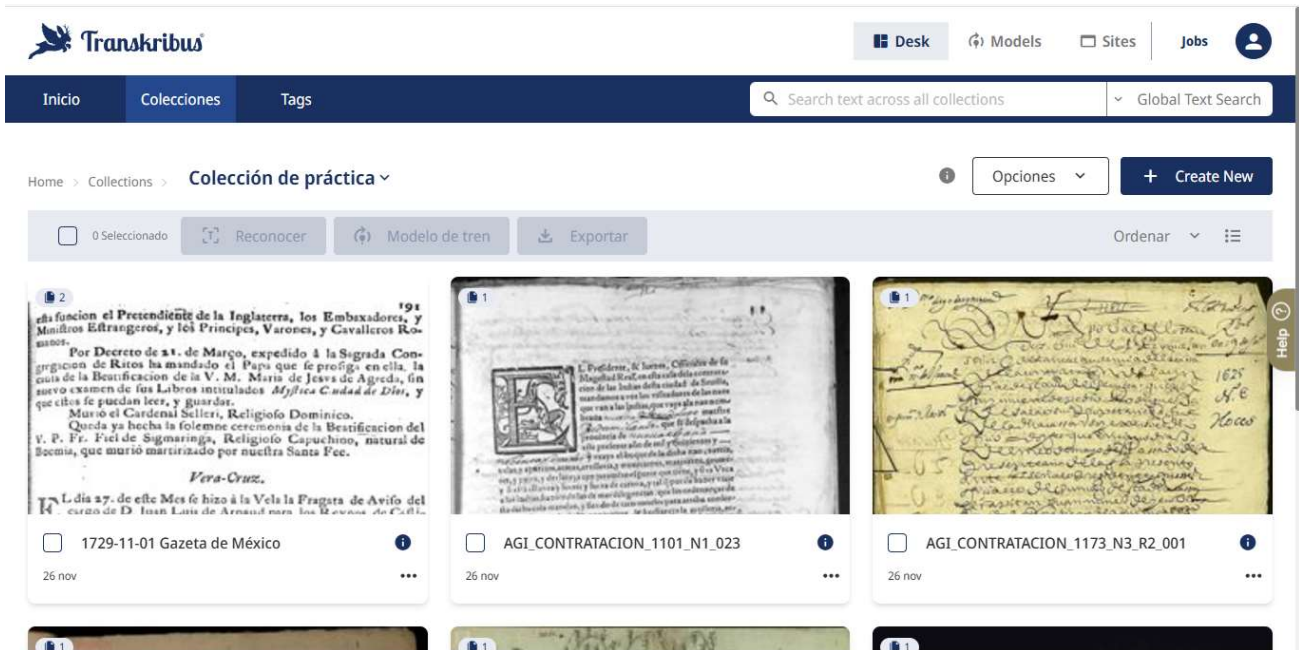
- (1) Mapa (AGN_Mapilu_02126F.jpg)
- (2) Periódico (1729-11-01 Gazeta de México.pdf)
- (3) Manuscritos (AGI_CONTRATACION_1173_N3_R2_001.jpg, AGN_Inquisicion_v173_exp001_001r.jpeg, AGN_RealesCedDup_v003_exp334.jpg)
- (4) Impreso/manuscrito (AGI_CONTRATACION_1101_N1_023.jpg)



6. Cuando des clic en “Enviar”, aparecerá la pestaña **Jobs** que muestra el avance en la importación de tu documento.

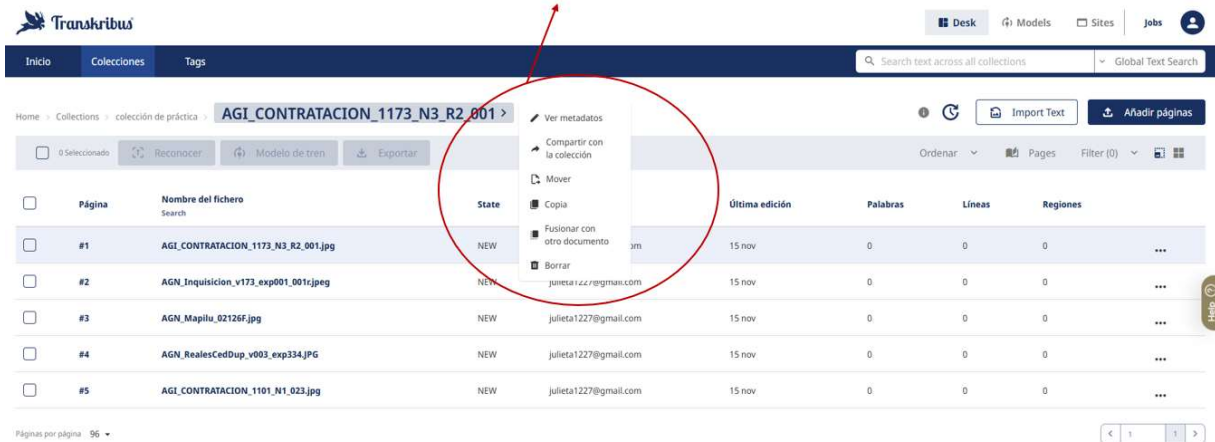



- Una vez cargados tus documentos, es decir, cuando su estado sea “Finished”, regresa a la colección cerrando la pestaña Jobs. Vuelve a cargar la página de la colección para que aparezcan todos los documentos recién cargados. Recuerda que puedes visualizarlos de dos formas, ya sea en formato de imágenes o como listado.

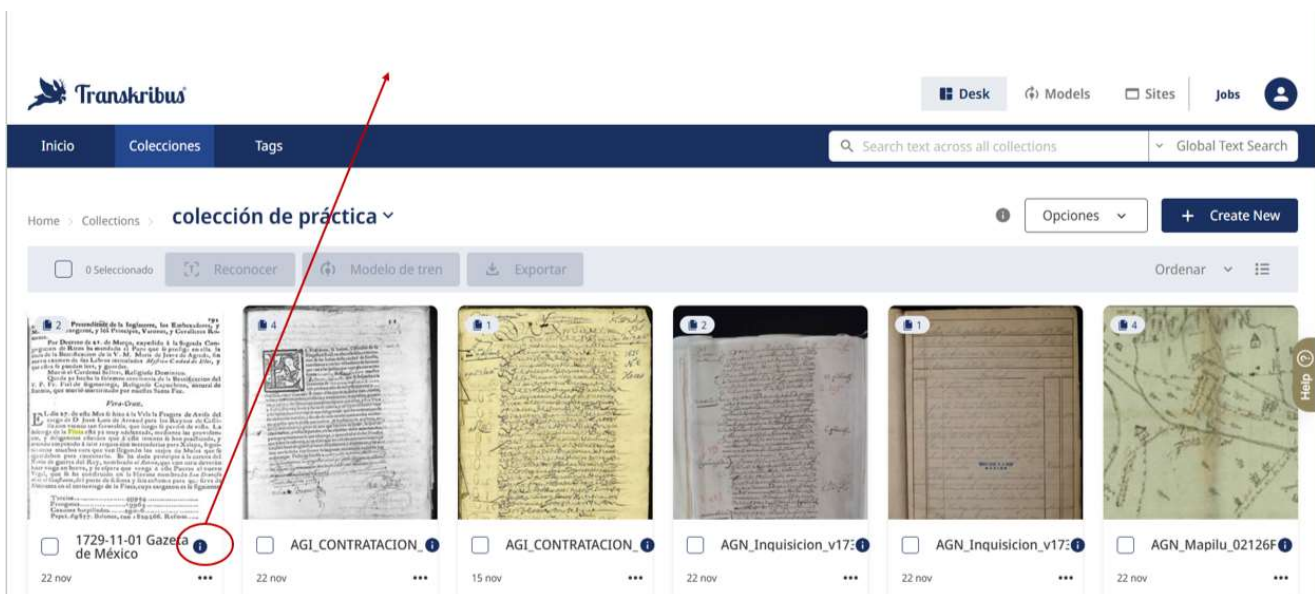


Editar metadatos

En la pestaña de Metadatos de un documento, puedes agregar o cambiar la información básica del mismo (como el nombre) e incluir datos adicionales de autoría o género literario. Para editar los metadatos de tus documentos, debes dar clic en el nombre documento y, después, en la opción “Editar metadatos” del menú contextual que se despliega a continuación, como se muestra en la imagen:



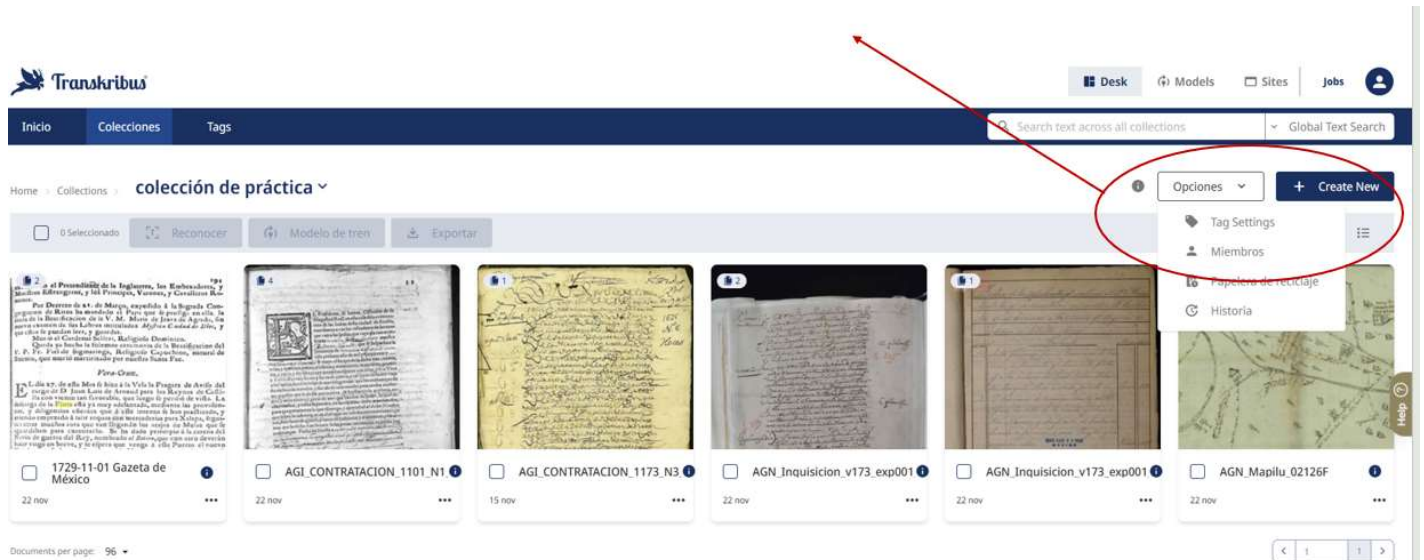
Además de editar los metadatos, desde este menú podrás compartir el documento con otra colección, moverlo, copiarlo y desecharlo. También puedes acceder a los metadatos de tu documento haciendo clic en el icono  ubicado junto al nombre del documento, como se muestra en la siguiente imagen.



Los metadatos de los documentos que te proporcionamos los puedes encontrar en el mismo [blogpost de materiales del curso](#), en el archivo de texto nombrado `Metadatos_materiales_practica_Transkribus.docx`.

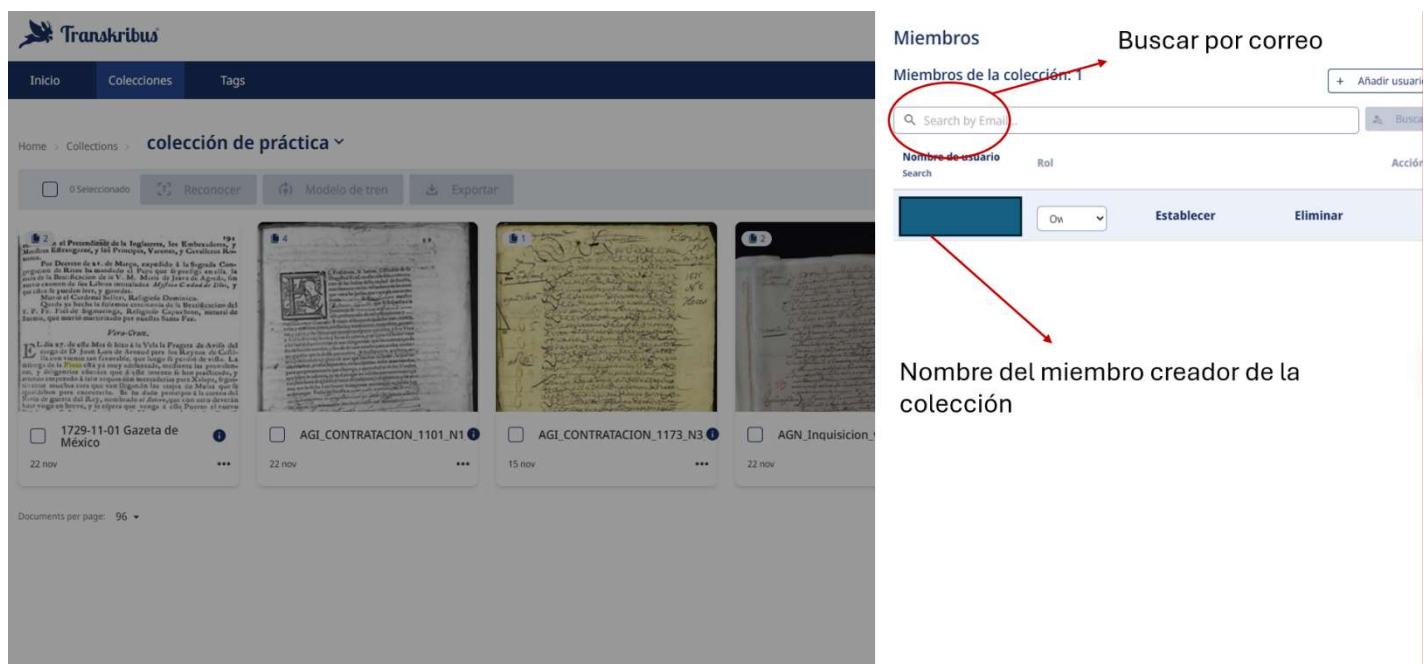
Compartir una colección

Para compartir una colección, es necesario ubicar el cuadro de **Opciones** en tu pantalla, como se indica en la imagen.



A continuación, tendrías que hacer clic en **Miembros** y escribir el correo electrónico de la persona con la que quieres compartir tu colección. Ojo, si la persona con la que deseas compartir la colección no se ha dado de alta, será imposible compartírsela.

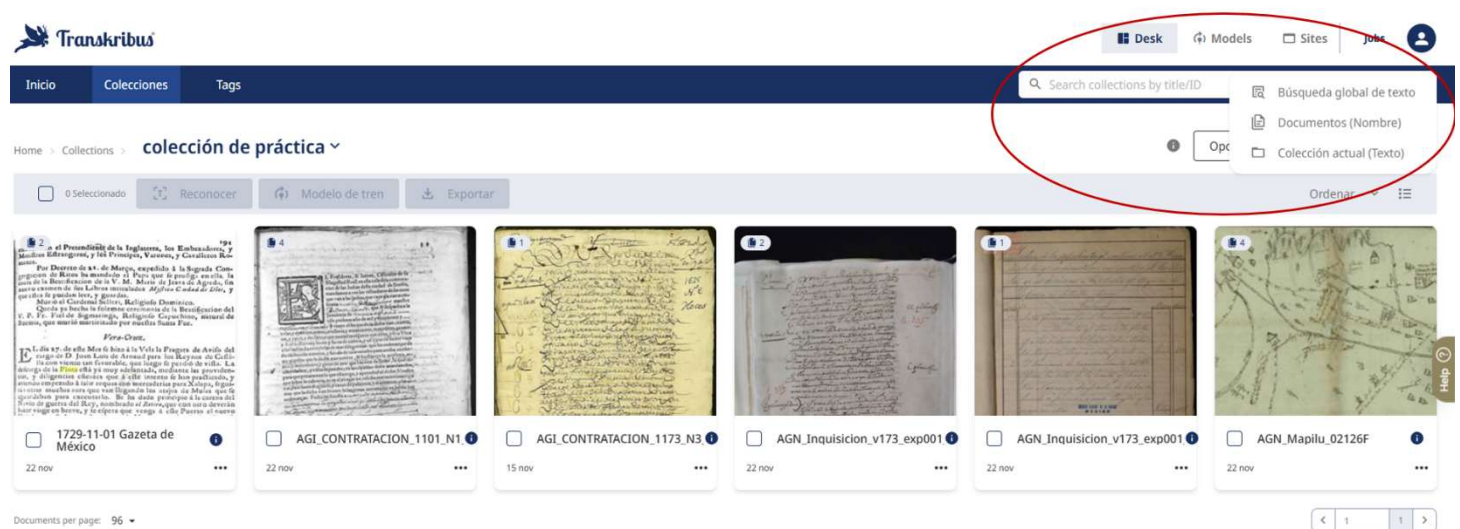
No se pueden compartir documentos individuales, sólo se puede compartir la colección completa.



A cada usuario de una colección se le puede dar el rol de OWNER, EDITOR o TRANSCRIBER. Cada uno de estos roles tiene diferentes permisos, los cuales puedes consultar en: <https://help.transkribus.org/managing-users>

Búsqueda de documentos

Existen tres formas de buscar un documento: 1) por texto, 2) por nombre del documento o 3) por nombre de la colección. Para hacer una búsqueda, es necesario ubicarse en la barra de búsqueda, en la esquina superior derecha de la pantalla, y seleccionar la forma deseada, como se observa en la imagen:

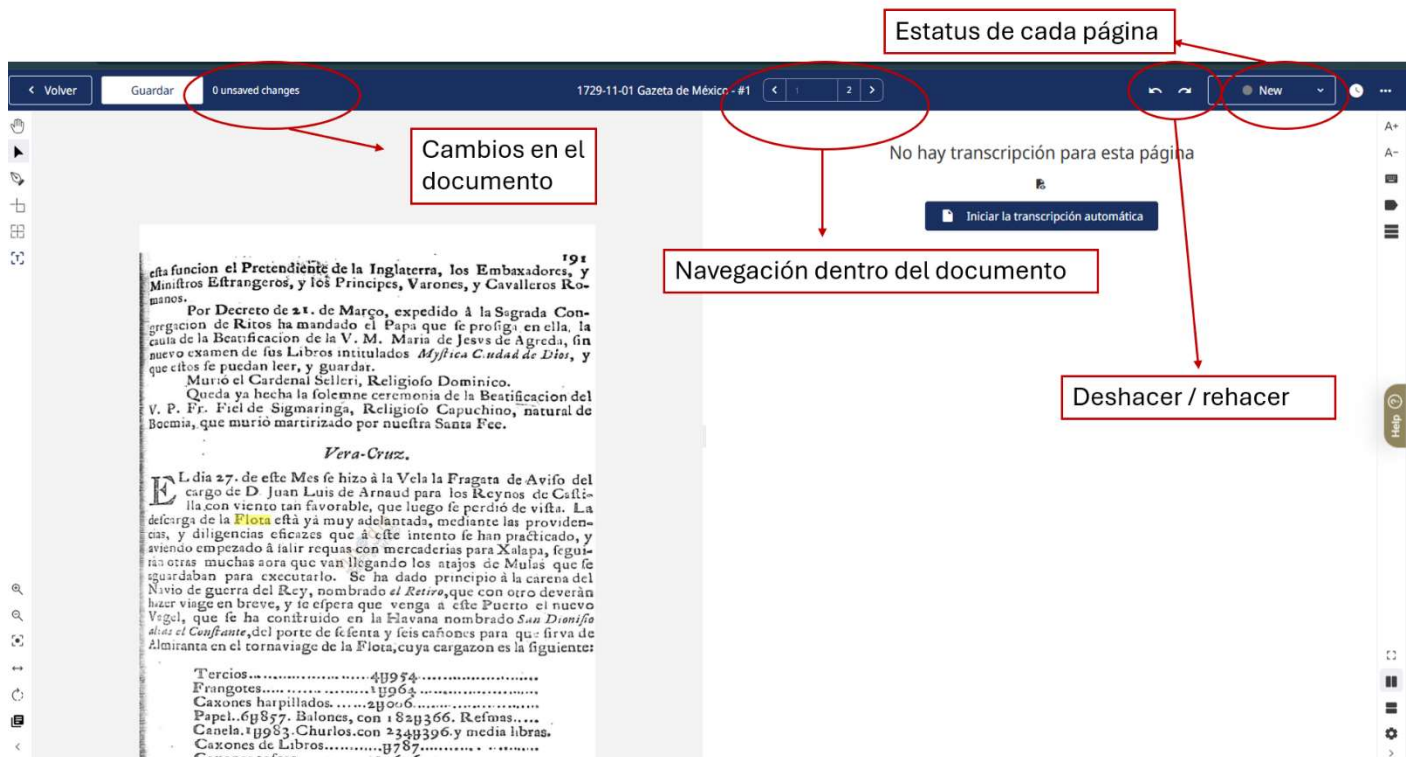


INTERFAZ DE TRANSCRIPCIÓN

Una vez dentro de un documento, te encontrarás con la siguiente pantalla, cuyo contenido exploraremos a continuación.

En la parte superior de la pantalla encontrarás los siguientes controles:

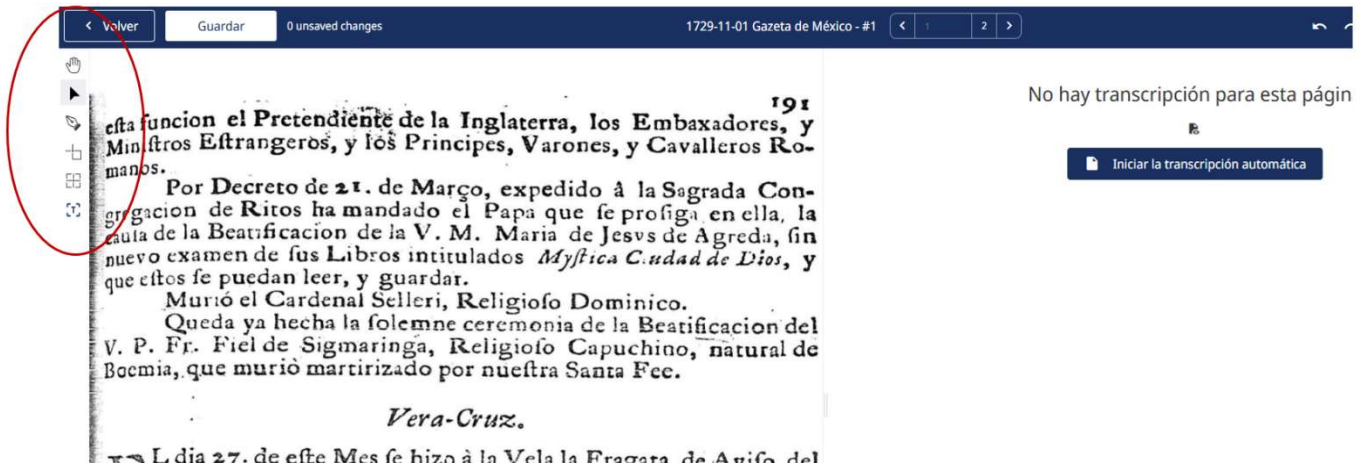
- Botón “Guardar”, junto con el número de cambios que has hecho en el documento sin salvar.
- Controles de navegación entre páginas de tu documento.
- Botones deshacer y rehacer.
- Estatus de cada página.



El botón de estatus (estado) de la página, en la parte superior derecha de la pantalla, permite saber, y hacerle saber a los demás, el progreso de tu trabajo. Los diferentes estados que puede tener una página se muestran en la imagen siguiente. Estos pueden ser: “En curso” (anaranjado), muestra que la página se está trabajando; “Hecho” (amarillo) señala que el trabajo está terminado y listo para revisión; “Final” (verde claro) indica que la página está terminada y revisada; “La verdad sobre el terreno” (una terrible traducción de *Ground Truth*, en verde más oscuro) refiere a que la página está lista para usarse en el entrenamiento de modelos HTR.










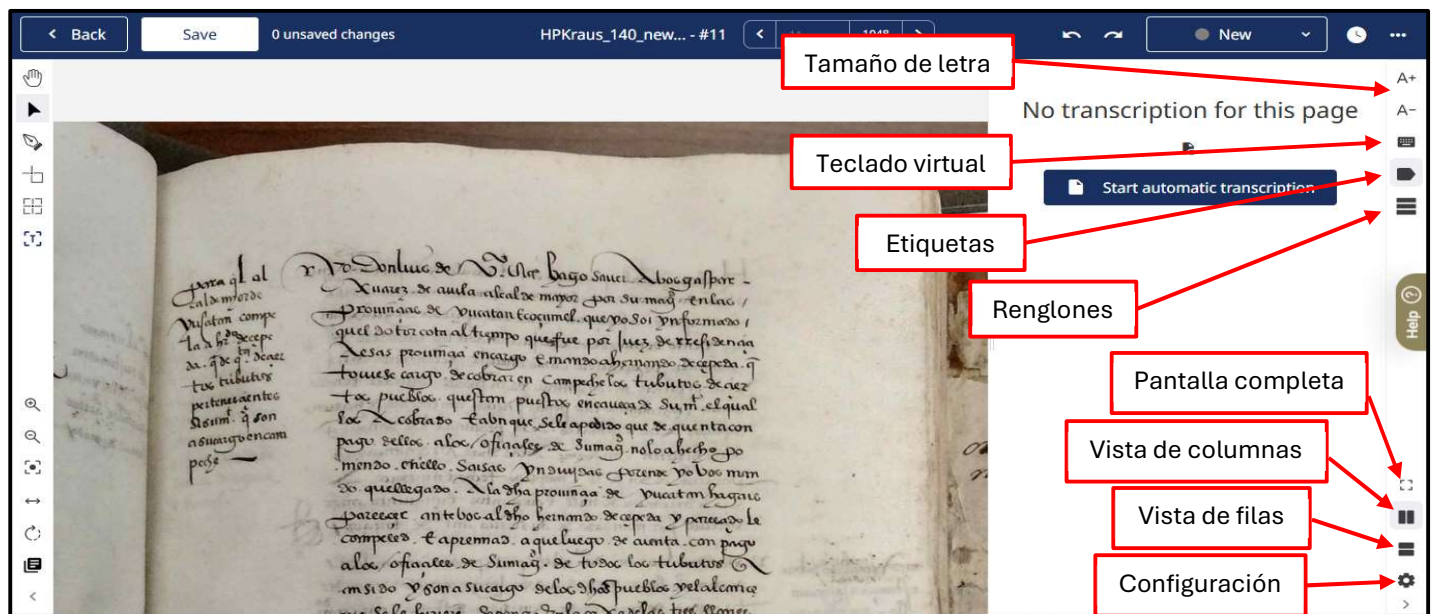
El siguiente grupo de botones que debes conocer se encuentra en el lado izquierdo de tu pantalla.



- **Drag Mode** 🖱️ : mover la página a donde uno prefiera.
- **Modo selección** 📏 : se utiliza para seleccionar un elemento (línea, región o tabla).
- **Añadir línea** ➕ : agregar renglones (líneas base).
- **Añadir región** 📐 : agregar regiones.
- **Añadir tabla** 📄 : agregar tablas.
- **Reconocimiento** 🗣️ : inicia la transcripción automática de la página usando modelos de reconocimiento de texto.

La barra de controles laterales de la derecha incluye botones para:

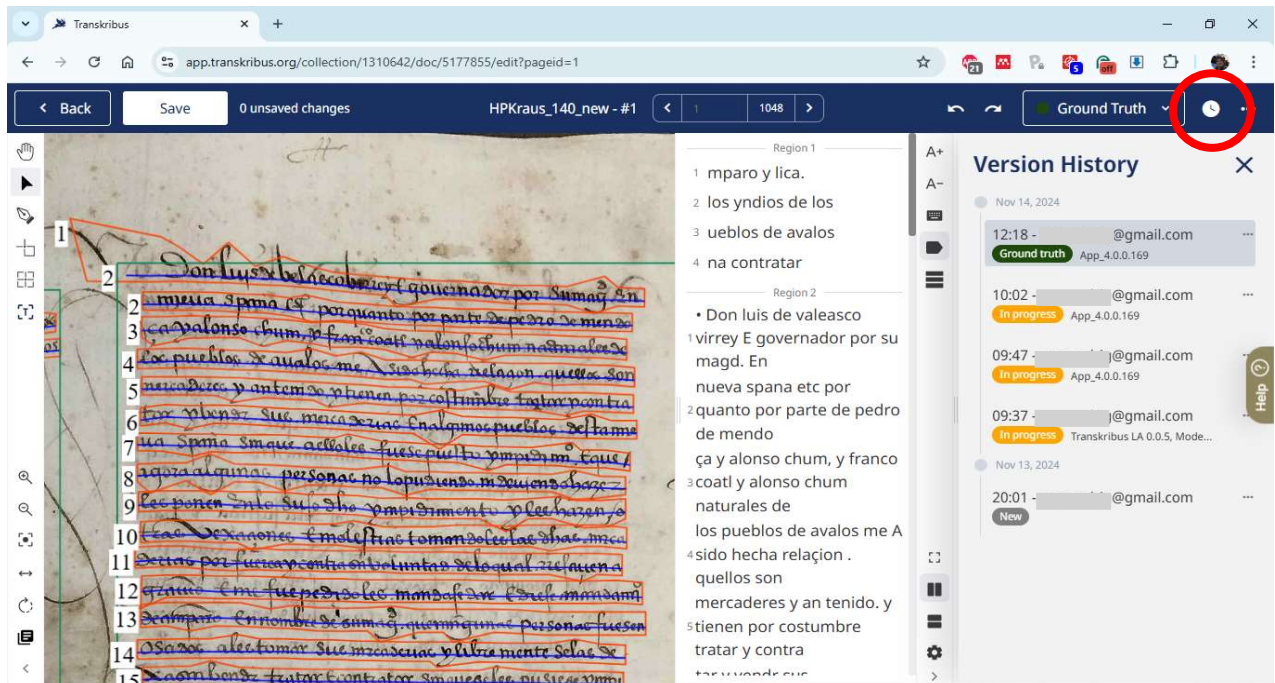
- **Tamaño de letra:** permite aumentar A^+ o disminuir A^- el tamaño de letra de la transcripción.
- **Teclado virtual** : permite activar o desactivar el teclado virtual en pantalla.
- **Etiquetas** : permite activar o desactivar la posibilidad de agregar etiquetas o anotaciones al texto de la transcripción.
- **Renglones (Diseño)** : permite activar o desactivar la posibilidad de modificar el orden de los renglones de la transcripción.
- **Pantalla completa** : entra/sale del modo pantalla completa.
- **Vista de columnas** : ubica la transcripción a un lado de la imagen.
- **Vista de filas** : ubica la transcripción debajo de la imagen.
- **Configuración** : muestra el menú de configuración de textos, imágenes, teclado virtual y etiquetas.



Control de versiones

Todos los cambios realizados en una página se guardan en el historial de versiones. A este se accede en el icono del reloj, en la esquina superior derecha de la pantalla. Cada una de las

versiones incluye la fecha y hora, el correo del usuario que la guardó, el estatus de la página y la plataforma o el modelo con el que se hizo el trabajo.

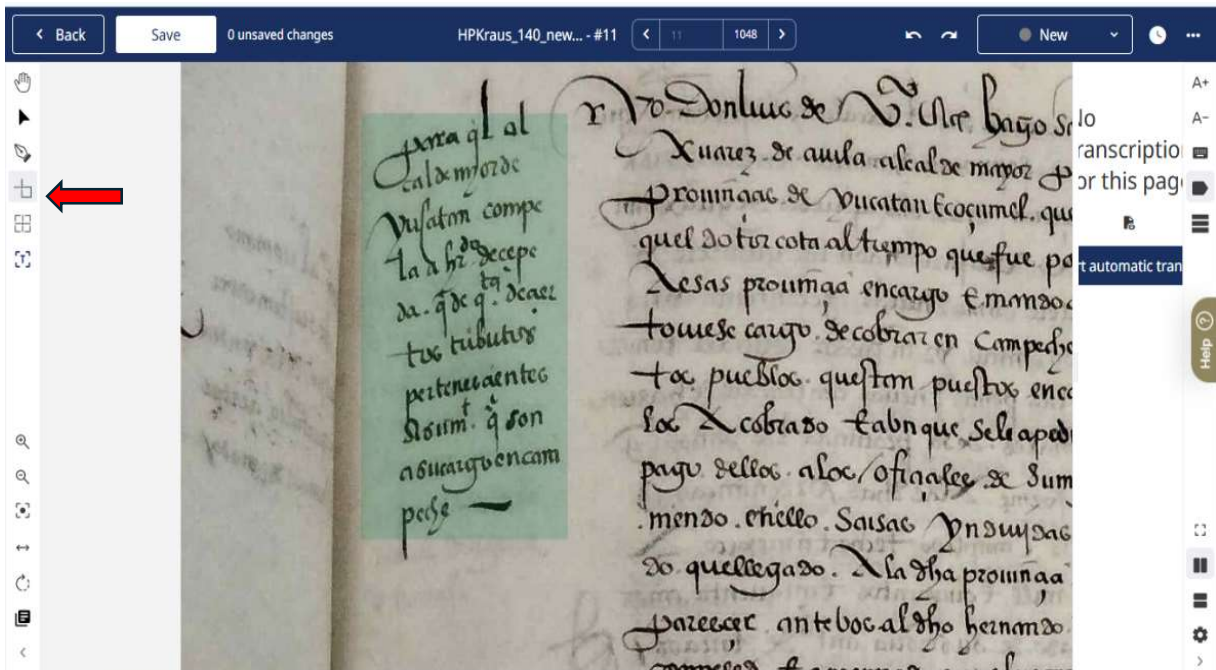


TRANSCRIPCIÓN MANUAL

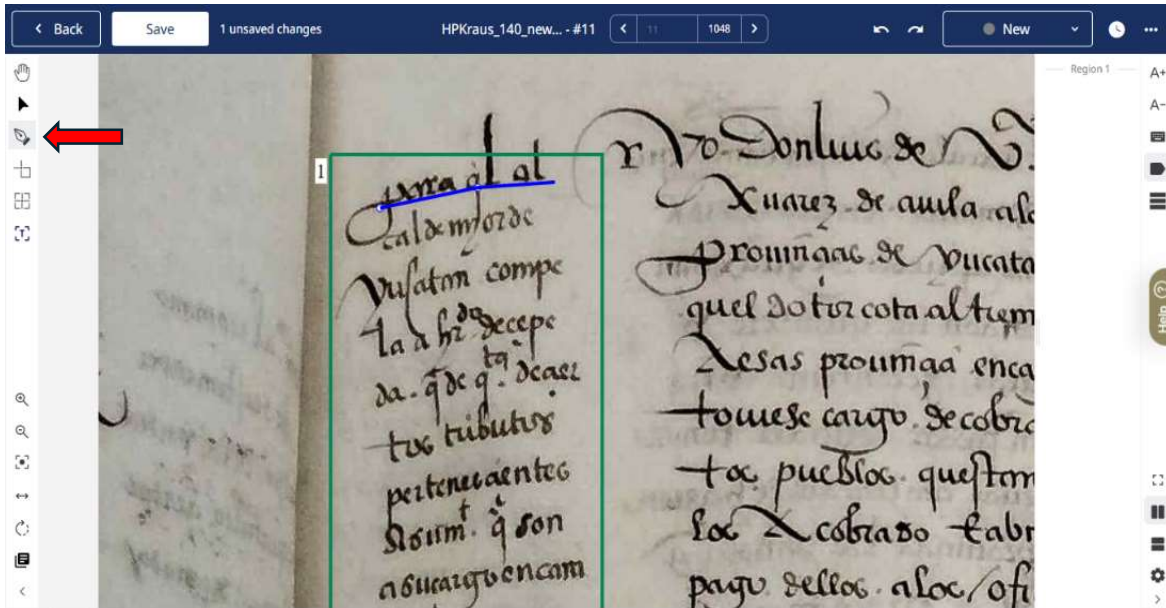
Para llevar a cabo la transcripción del texto, ya sea manual o automática, es necesario indicarle al programa en qué partes de la imagen se encuentra el texto. Esto se hace por medio de regiones, líneas base y polígonos. Una **región** es un área cuadrangular que encierra el texto manuscrito contenido en la imagen; puede haber varias regiones en una misma imagen (por ejemplo, encabezado, márgenes, cuerpo, etc.). Una **línea base** (*baseline*) es una polilínea que corre a lo largo de la parte inferior de la línea de texto y le indica al programa la orientación del texto. A partir de las líneas base, se crean automáticamente los **polígonos** de línea, que encasillan todo el texto contenido en un mismo renglón.

Para crear manualmente renglones para la transcripción, debes:

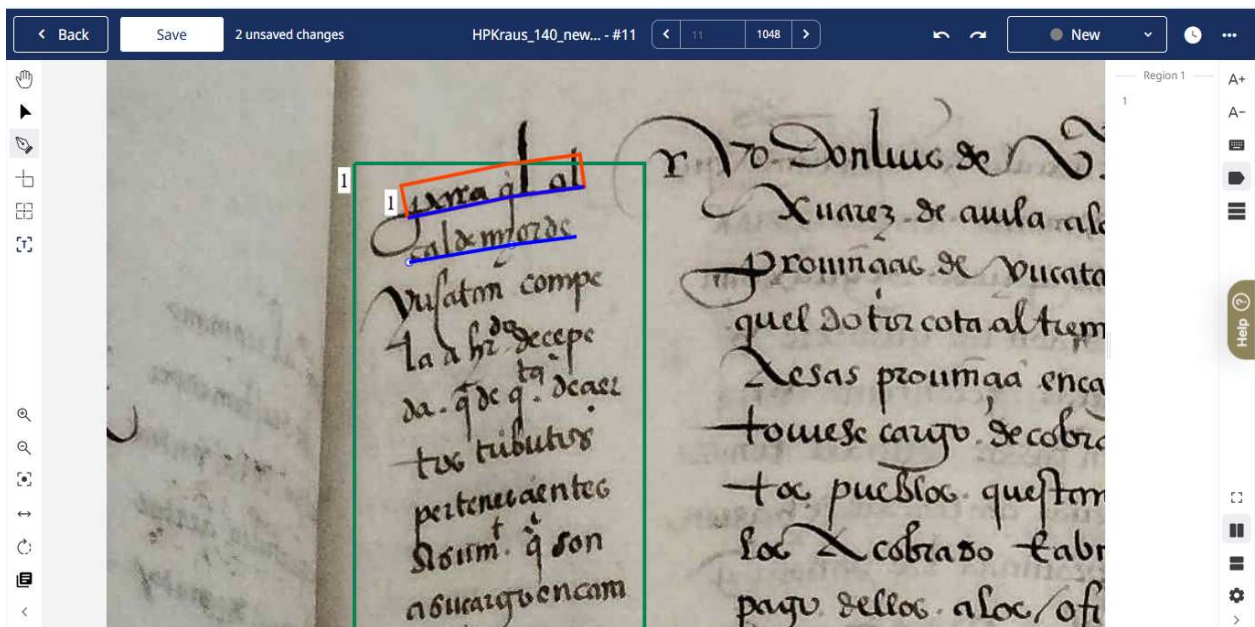
1. Agregar una región (recuadro verde) que abarque el área del texto donde estarán los renglones. En una misma página se puede agregar una o varias regiones.



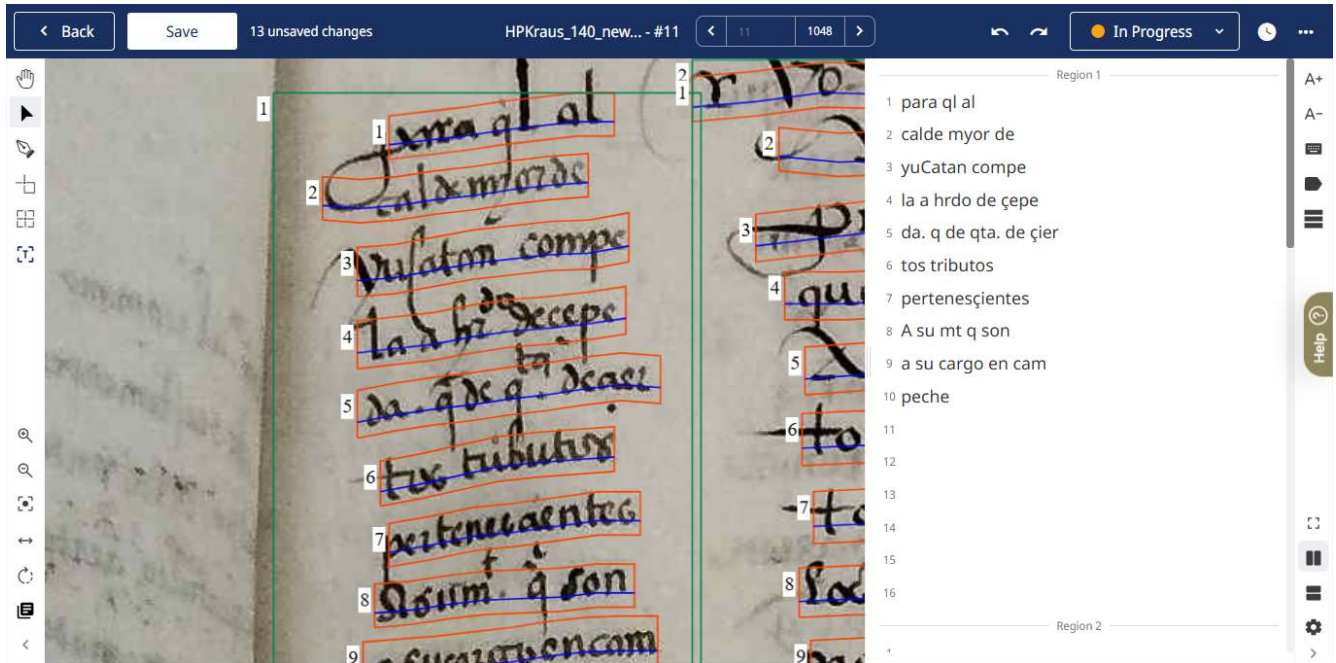
2. Dentro de la región, agregar una línea base (azul), haciendo clic en puntos en la base de las letras a lo largo de todo el renglón.





3. Una vez agregada la línea base, automáticamente se crea el polígono del renglón (rojo). Continuar agregando líneas de la misma forma para los demás renglones.



4. A cada renglón que agregues en la imagen, Transkribus asocia automáticamente un renglón de texto en el área de la derecha de la pantalla, donde se llevará a cabo la transcripción. Una vez agregados todos los renglones, se puede comenzar con la transcripción manual.



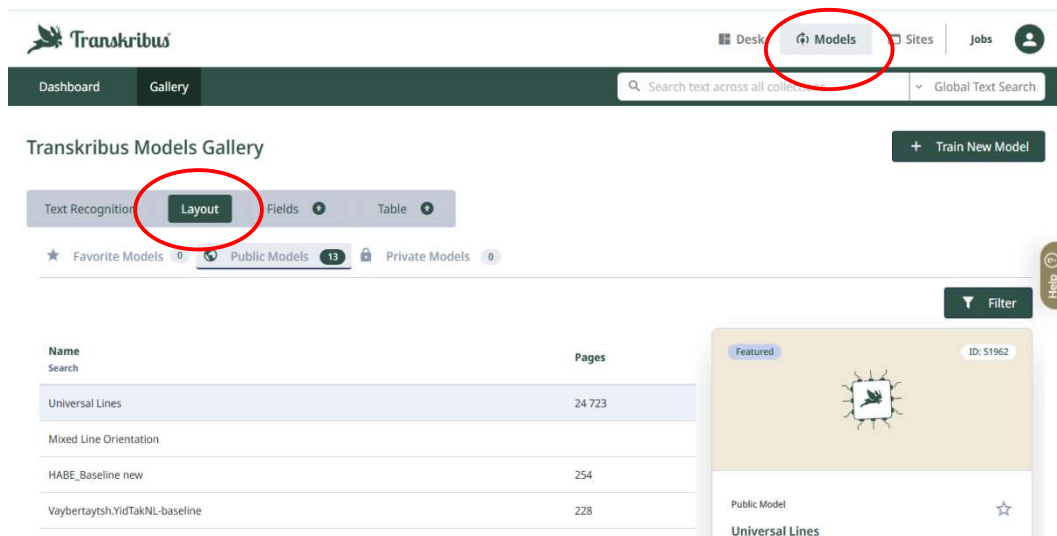
Recuerda que puedes ubicar el área de transcripción al lado derecho de la imagen o debajo de ésta, según prefieras, usando los botones “Vista de columnas”  y “Vista de filas”  de la barra de controles derecha.

LAYOUT ANALYSIS

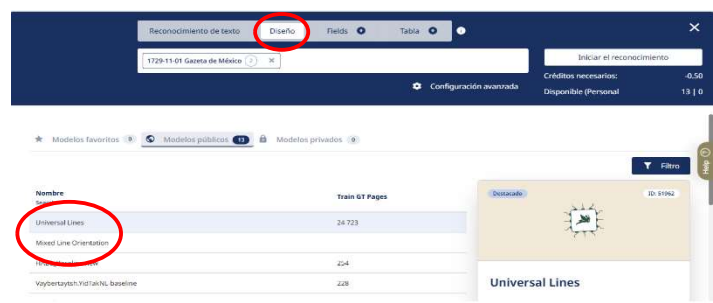
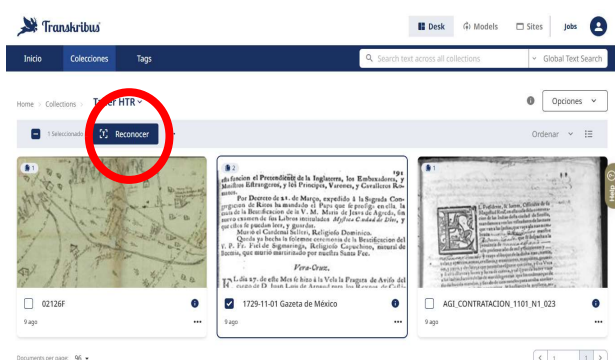
El *Layout Analysis* (análisis del diseño o estructura de la página) permite a Transkribus identificar automáticamente las regiones de la imagen en donde se encuentra el texto. Los modelos de *Layout Analysis* realizan de manera automática el proceso de creación de regiones, líneas base y polígonos que llevaste a cabo manualmente en la sección anterior.

Ejercicio: Corre el *Layout Analysis* a un documento.

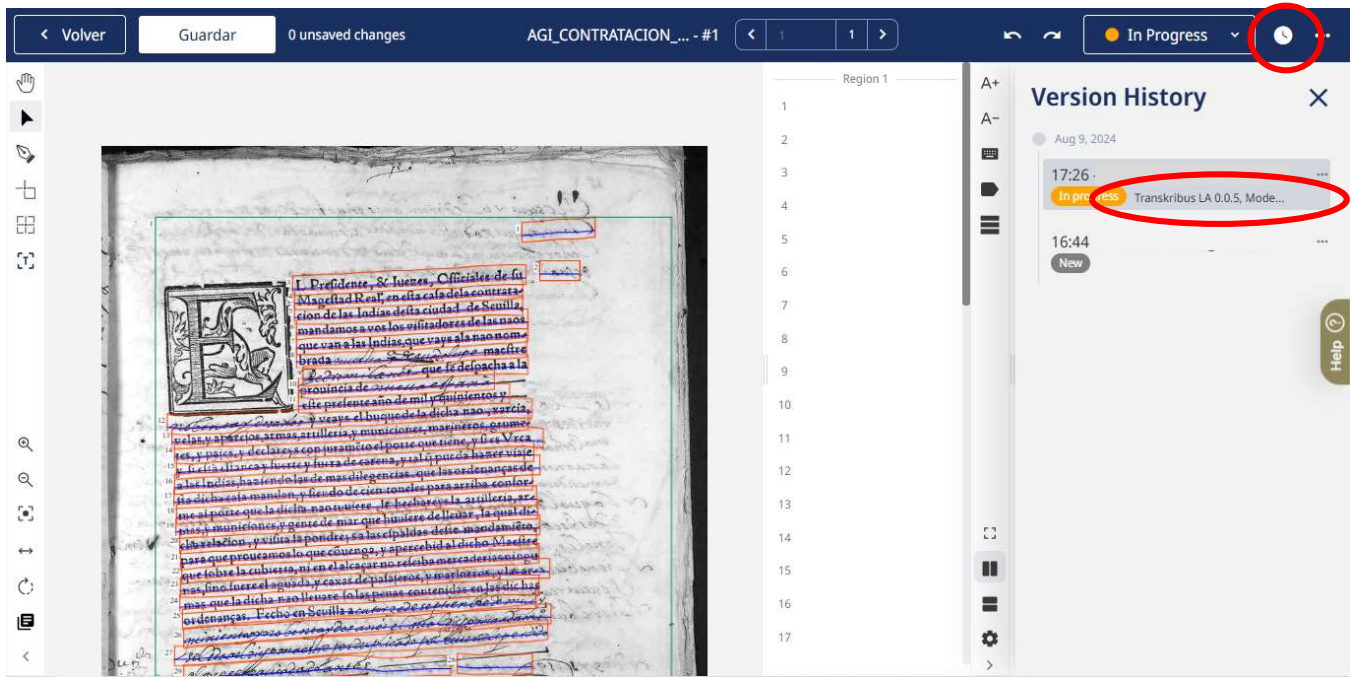
- 1) Ve a la sección Galería de modelos, luego a la sección *Layout* (Diseño). Revisa la descripción de algunos de los modelos públicos para saber qué hacen y cómo fueron entrenados.



- 2) Regresa a tu colección de documentos. Selecciona una página de alguno de tus documentos o de los documentos de la carpeta de materiales del curso y da clic en el botón **Reconocer**. En la pantalla de modelos, ve a los de *Layout* (Diseño) y corre el reconocimiento usando el modelo público *Universal Lines*. Luego repite el proceso usando los modelos *Mixed Line Orientation* y *Newspapers (NewsEye)*.



3) Cuando hayan terminado los reconocimientos, regresa a tu documento. En el “Control de Versiones” (icono del reloj en la barra superior) encontrarás los diferentes trabajos que se han hecho a la misma página. Los diferentes *Layout Analysis* aparecerán nombrados como “Transkribus LA... Model ###”. Las siglas “LA” corresponden a *Layout Analysis* y el número después de *Model* es el identificador del modelo que se usó: Universal Lines=51962, Mixed Line Orientation=49272, Newspapers (NewsEye)=49277.



4) Compara los resultados que te arrojó cada modelo en los diferentes documentos:

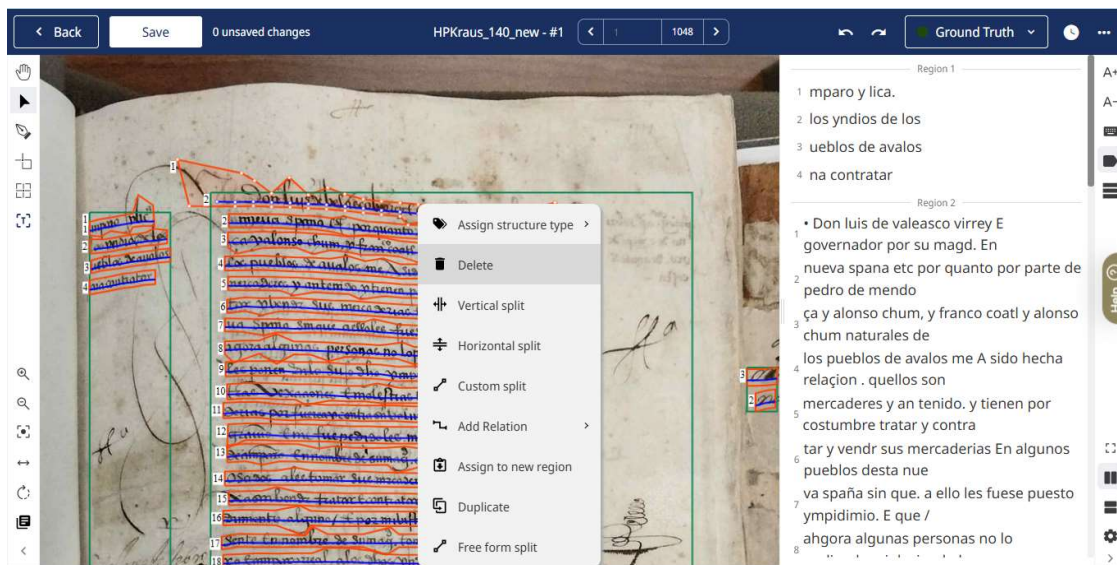
- ¿Qué modelo funcionó mejor para qué documento? ¿Habría sido más fácil/rápido crear los renglones a mano? ¿Sería necesario entrenar un modelo específico para el tipo de página que analizaste?

Limpieza del layout (diseño)

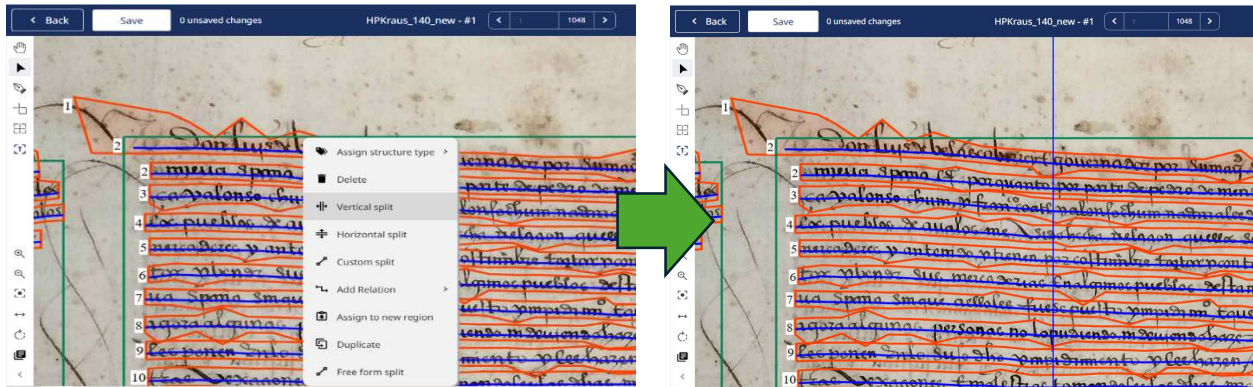
Los modelos de *Layout Analysis* (LA) suelen dar buenos resultados con páginas de texto relativamente limpias y estructuradas. No obstante, siempre puede haber errores. Por ello, una vez terminado el LA, es importante revisarlo y, de ser necesario, “limpiarlo”, ya sea borrando, dividiendo, uniendo o reordenando renglones, o modificando los polígonos de estos.

Para llevar a cabo las siguientes acciones, primero hay que dar clic en el botón **Modo selección** (barra de controles izquierda).

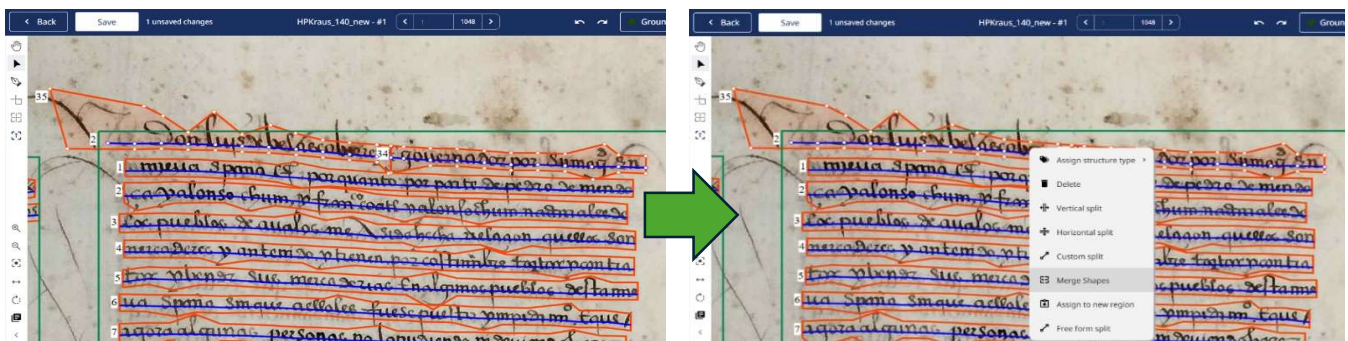
Borrar renglones: Hay dos opciones para borrar un renglón: 1) Da clic en el renglón que deseas borrar para seleccionarlo y luego presiona la tecla Suprimir de tu teclado. 2) Da clic derecho dentro del polígono del renglón; en el menú contextual que aparece, selecciona la opción “Delete”.



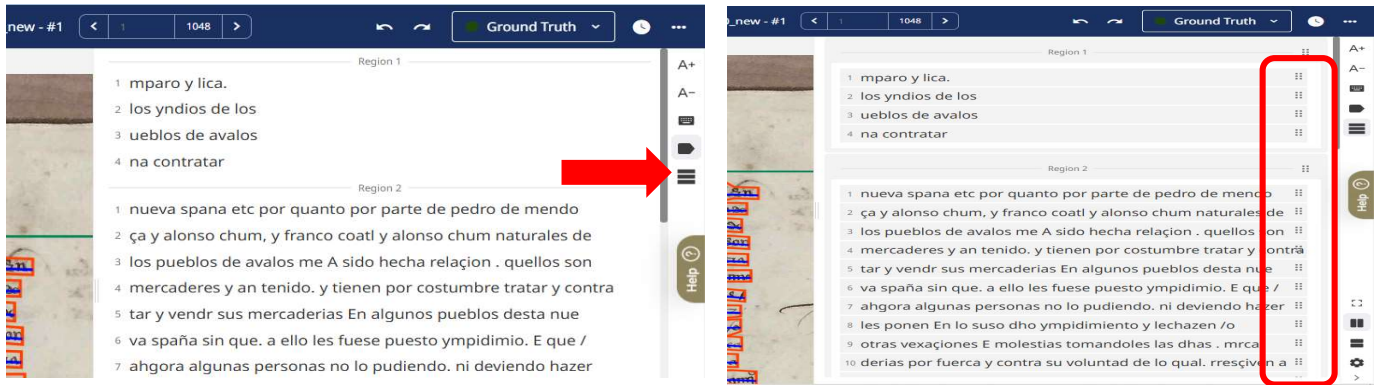
Dividir renglones: Para dividir un renglón, hay que dar clic derecho dentro del polígono del renglón y en el menú contextual que aparece, seleccionar la opción “Vertical split”. Aparecerá una línea vertical azul que puedes mover con el ratón. Coloca esta línea en el sitio donde quieras dividir el renglón y da clic. Los dos nuevos renglones que se creen conservarán el texto completo de la transcripción que haya tenido el renglón original.



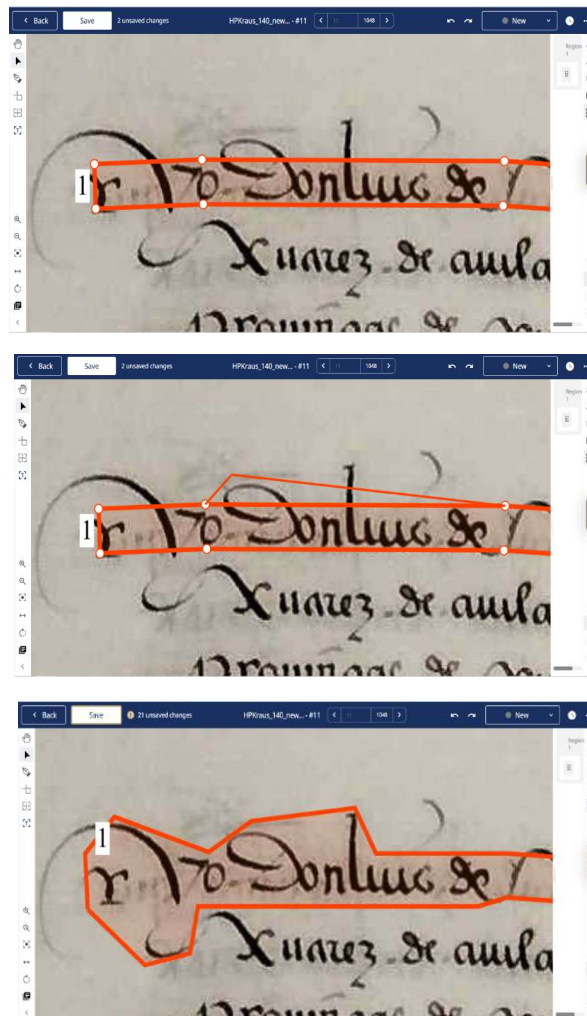
Unir renglones: Para unir dos renglones, manteniendo presionada la tecla Control del teclado, dar clic en ambos renglones para seleccionarlos. Una vez seleccionados, soltar la tecla Control y dar clic derecho dentro del polígono de uno de los renglones. En el menú contextual que aparece, seleccionar la opción “Merge shapes”. El nuevo renglón que se crea conservará el texto completo de la transcripción que hayan tenido ambos renglones originales.



Reordenar renglones: Para cambiar el orden de los renglones, primero hay que activar la opción “Diseño” (Layout) dando clic en el botón de la barra de controles de la derecha de la pantalla. Esto cambiará de blanco a gris el color los renglones en el área de transcripción y agregará un grupo de puntos en el extremo derecho de cada renglón. Dando clic en estos puntos en el renglón que se desee mover, podemos arrastrar el renglón hacia arriba o hacia abajo a su nueva posición. Al soltar el renglón en la nueva posición, automáticamente se modifica la numeración de todos los demás renglones.



Modificar polígonos: Se puede modificar la forma del polígono de un renglón moviendo, agregando o eliminando los vértices que lo delimitan. Al dar clic en el renglón para seleccionarlo, los vértices aparecen como círculos blancos. Para agregar un nuevo vértice, basta con hacer clic en la línea del polígono y arrastrarla hacia donde se quiera crear el vértice.

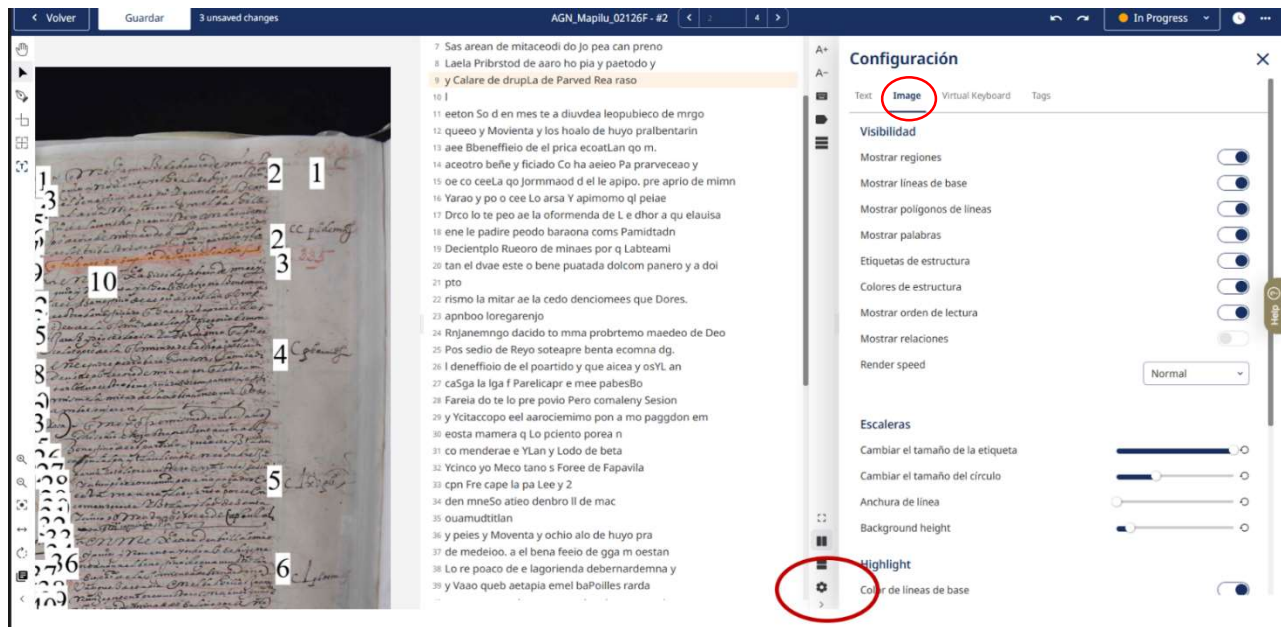


¡No olvides guardar tus cambios! Sobre todo, en esta parte de limpiar el documento y modificar los polígonos.

IMPORTANTE: Si tu computadora se suspende o está inactiva por un tiempo mientras está abierto Transkribus, al regresar a trabajar es muy probable que no puedas guardar los cambios. Tendrás que volver a cargar la página, volver a la colección o cerrar sesión y entrar nuevamente.



Finalmente, recuerda que puedes modificar a tu gusto el tamaño de las etiquetas, los colores y el grosor de las líneas, polígonos, áreas, etc. Estas opciones se encuentran dando clic en el botón “Configuración” (esquina inferior derecha) y después en la sección “Image”.

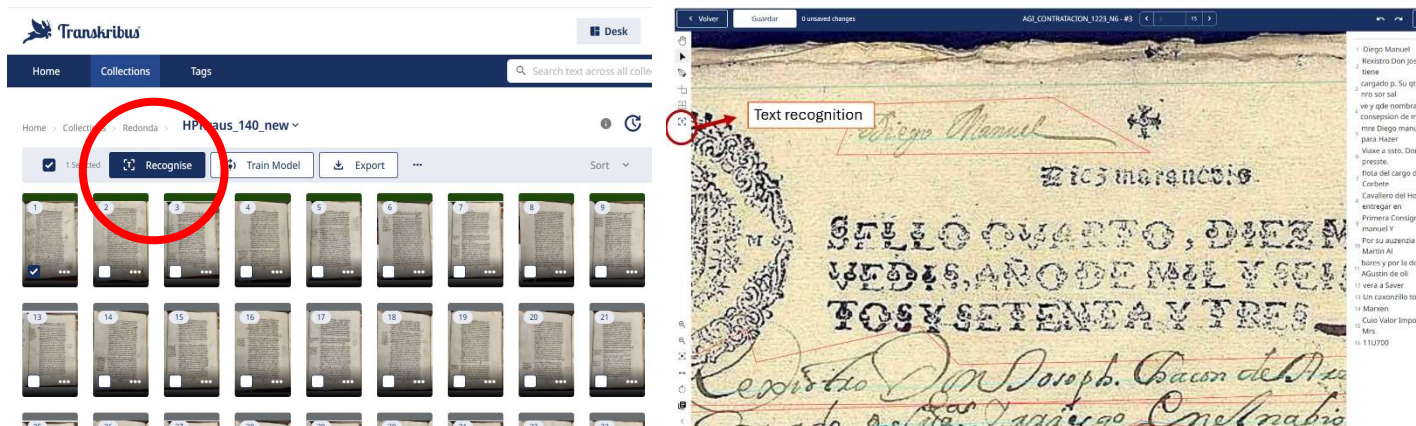


TRANSCRIPCIÓN AUTOMÁTICA

Ejercicio: Lleva a cabo la transcripción automática de una o varias páginas de uno de tus documentos usando un modelo de reconocimiento de texto manuscrito (HTR).

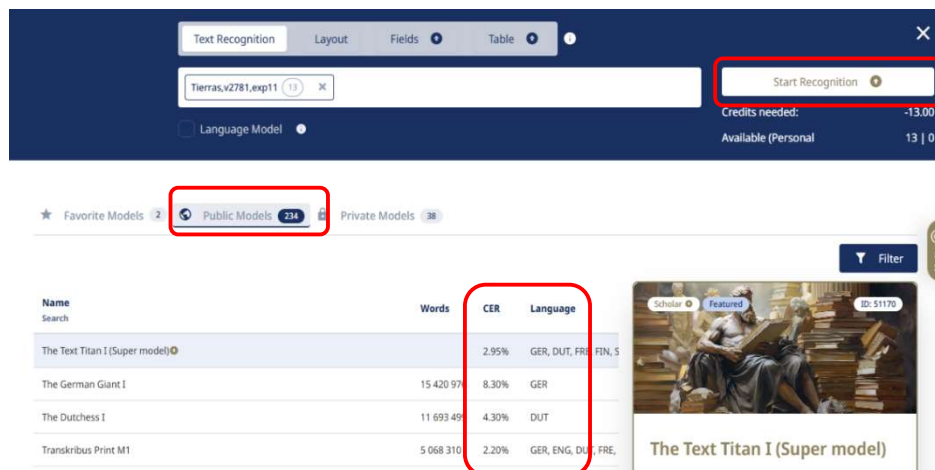
1) Dentro del documento, selecciona **una** página que quieras analizar. Da clic en el botón azul “Reconocer”.

Alternativamente, si estás dentro de la página (en la interfaz de transcripción), puedes dar clic al símbolo T (*text recognition*) en la barra de controles situada a la izquierda de tu pantalla.



Nota: Es posible analizar múltiples páginas a la vez, pero esto hace el proceso más lento. Para fines de este ejercicio sólo analizaremos una página.

2) Selecciona el modelo con el que quieras hacer el reconocimiento de texto. Toma en cuenta el idioma, el tipo de documentos con los que fue entrenado el modelo y su tasa de error (CER / *Character Error Rate*). Estos datos los puedes ver en la descripción del modelo.



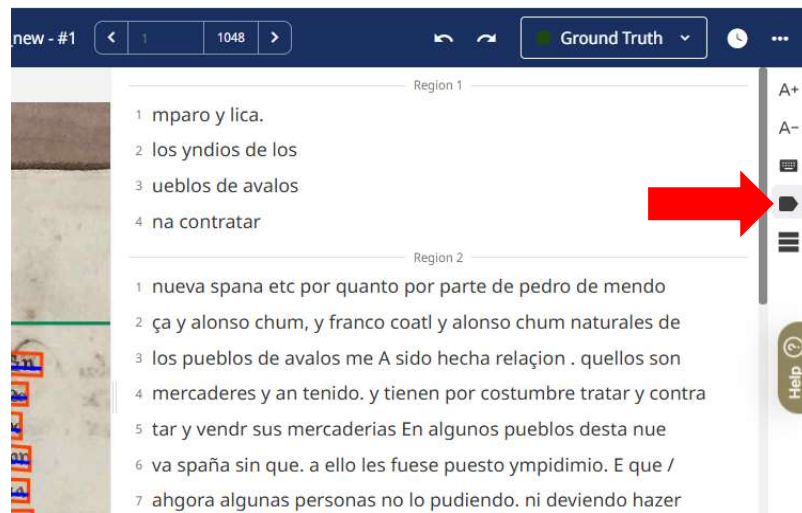
- 3) Da clic en “Start Recognition”. El proceso de reconocimiento y transcripción automática tardará unos segundos. Cuando esté listo, el estatus de la página cambiará a “En curso” (color anaranjado) y, al abrir la página, podrás ver la transcripción automática.
- 4) Ahora prueba a hacer el reconocimiento en esa misma página usando otro modelo diferente.
- 5) Revisa y compara las transcripciones automáticas que te arrojaron ambos modelos:
 - ¿Fueron buenas, regulares, malas?
 - ¿Qué tanto habría que corregirlas?
 - ¿Habría sido más rápido hacer la transcripción a mano desde el principio?
 - ¿Un modelo funcionó mejor que el otro? ¿A qué crees que se debe?
 - ¿Valdría la pena entrenar un modelo nuevo?

ANOTACIÓN

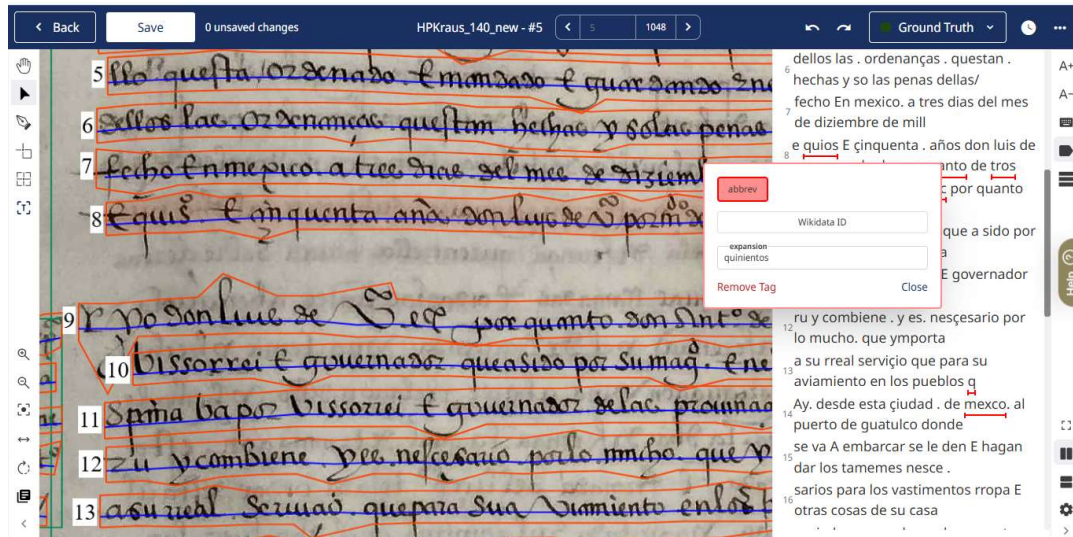
Transkribus permite hacer anotaciones, es decir, agregar etiquetas, tanto al texto de las transcripciones como a los elementos de la estructura del *layout*. Además de verlas en pantalla, estas anotaciones se exportan como etiquetas asociadas al texto en formato de metadatos.

Ejercicio: Agrega etiquetas de abreviatura a tu transcripción.

1. Para poder hacer anotaciones (agregar etiquetas), primero hay que activar la opción “Tags” dando clic en el botón de la barra de controles de la derecha de la pantalla.

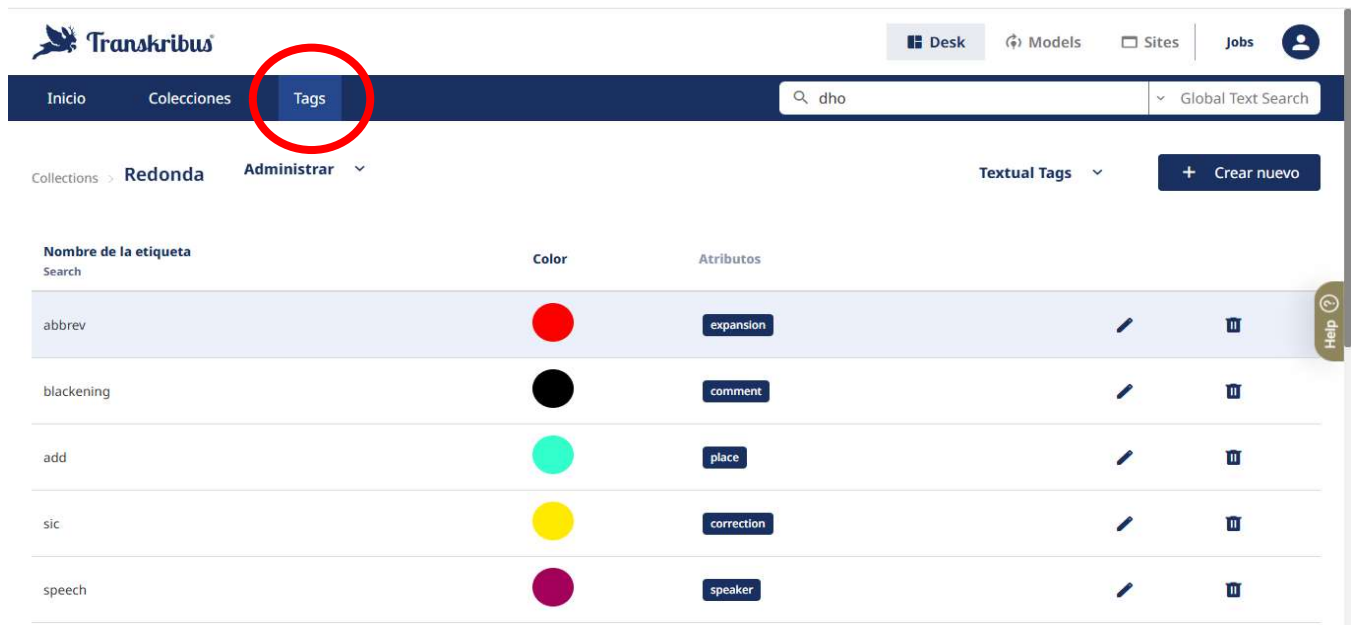


2. Ubica alguna palabra en el texto de tu transcripción que sea una abreviatura.
3. Selecciona la palabra, ya sea arrastrando el cursor o haciendo doble clic sobre ella. Aparecerá un menú contextual con las diferentes etiquetas que puedes agregar (las que tienes habilitadas).
4. Selecciona la etiqueta “abbrev” (abreviatura).
5. En el nuevo menú contextual que aparece, desarrolla la abreviatura dentro del recuadro “expansion” y da clic en “Cerrar”. La nueva etiqueta se agrega como una línea de color debajo de la palabra seleccionada, en el texto de la transcripción.



- Si quieres eliminar una etiqueta, haz clic en la línea de color y, luego, en el recuadro selecciona “Remove Tag”.

Hay diferentes etiquetas que se pueden usar como anotaciones y, también, se pueden agregar etiquetas personalizadas. Además, las etiquetas pueden variar entre colecciones. Todo esto se configura en el menú **Tags**.



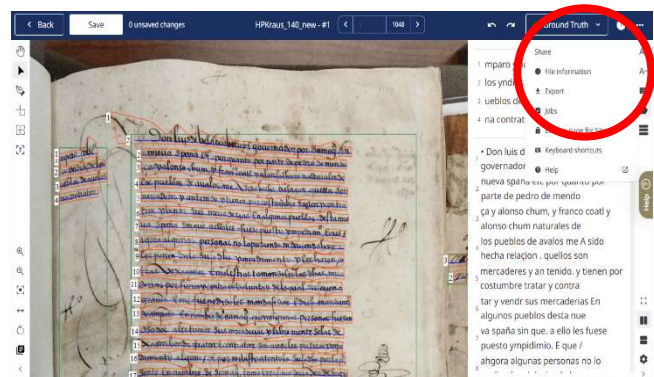
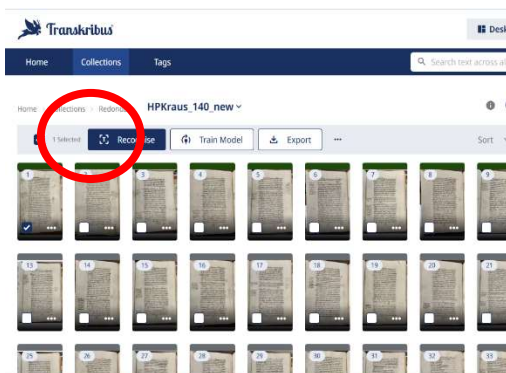
EXPORTACIÓN

Los resultados del trabajo de transcripción (manual o automática) y anotación, se pueden exportar en varios formatos: texto (en formato DOCX o TXT), texto enriquecido (en formato PDF o XML), o imagen (en diferentes resoluciones). La exportación se puede hacer para una página, varias páginas, un documento completo o varios documentos completos.

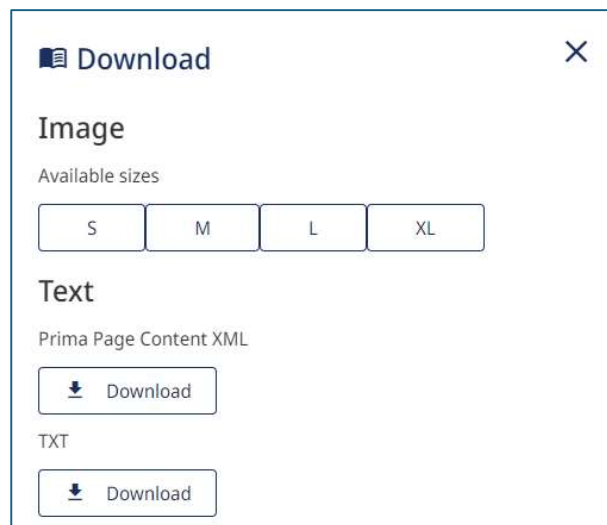
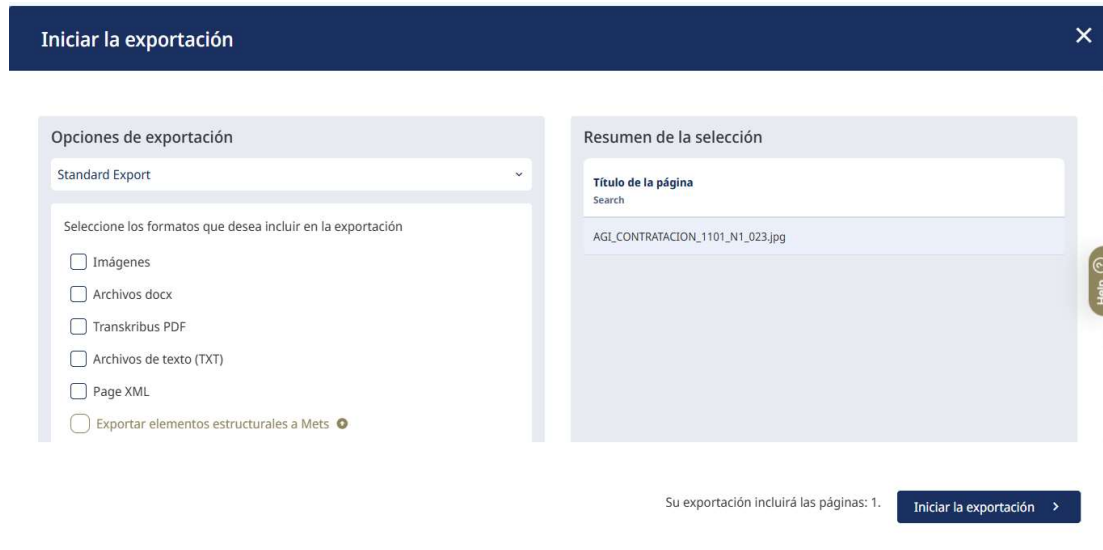
Ejercicio: Exporta los resultados de una transcripción automática.

1. Selecciona una página a la que le hayas hecho el reconocimiento de texto con un modelo HTR.
2. Si estás en la vista del documento, selecciona la página haciendo clic en su recuadro blanco y luego da clic en el botón “Exportar”.

Si estás en la vista de página (dentro de la interfaz de transcripción), da clic en los tres puntos blancos en la esquina superior derecha (junto al icono del reloj) y, luego, en “Exportar”.



3. En la siguiente pantalla, selecciona TODAS las diferentes opciones de exportación que te aparecen. Esto te permitirá compararlas.



4. El documento exportado se te enviará por medio de un enlace a tu correo electrónico. Al hacer clic en el enlace, se descargará a tu computadora un archivo comprimido (zip).
5. Abre el archivo comprimido y revisa los diferentes archivos que incluye.

Nota: Si escogiste como una de las opciones de exportación “Transkribus PDF”, el archivo PDF incluirá una capa con la imagen del documento y otra con la transcripción. Para poder verlas, debes habilitar el menú de capas en tu visor de PDFs. La siguiente imagen es un ejemplo de Acrobat Reader.

