



# Tecnología de reconocimiento óptico de caracteres aplicada a la actividad pericial

**Los profesionales traductores sabemos que, en el desarrollo habitual de nuestras actividades, la tecnología es fiel compañera. Y esto no escapa a la hora de actuar como peritos. Veamos cómo podemos transformar voluminosos expedientes escaneados o fotografiados y archivos PDF en documentos editables útiles que nos permitan trabajar con un entorno informático pensado para los traductores.**

| Por la **Traductora Pública Sol Brienza**, integrante de la Comisión de Recursos Tecnológicos

Los avances tecnológicos actuales no solo nos permiten estar conectados a toda hora y en todo lugar, lo que facilita el trabajo remoto y la colaboración, sino también, y por sobre todas las cosas, optimizar nuestro trabajo, hacerlo más rentable y lograr un mayor nivel de calidad. En especial, los avances en la tecnología de reconocimiento óptico de caracteres (OCR) no se han quedado atrás. Mayor precisión, velocidad y facilidad de reconocimiento hacen que la interacción del traductor con estas herramientas sea mucho más sencilla. En este artículo, vamos a analizar de qué manera el *software* de OCR más avanzado y potente en el mercado —ABBY FineReader®— nos ayuda a la hora de actuar como peritos.

Uno de los principales beneficios de la conversión de expedientes escaneados, fotografiados o ya en formato electrónico no editable es poder trabajar en el entorno habitual actual del traductor, es decir, las herramientas de traducción asistida por computadora (TAC), y, por supuesto, aprovechar todas las ventajas que ofrecen: traducir con mayor rapidez y alimentar memorias de traducción para lograr uniformidad y usarlas en futuras pericias (o en pericias de una misma causa).

Es muy común que las pericias estén disponibles directamente en formato papel, por lo que un primer paso necesario es digitalizarlas. Cuando los documentos ya se entregan en formato electrónico, es muy probable que el original para la conversión sea un archivo PDF, aunque también puede suceder que simplemente sea una foto del expediente, y que esta se encuentre en formato JPG/JPEG o PNG. En todos los casos, el original está digitalizado como imagen, es decir, no es editable.

ABBY FineReader® permite convertir estos documentos escaneados, archivos PDF y archivos de imagen en documentos editables para poder trabajarlos. Y, a diferencia de otros sistemas de OCR que existen en el mercado, algunos gratuitos y otros no, le permite al traductor interactuar de forma eficiente y amigable con el archivo fuente que se desea convertir, para poder obtener resultados confiables rápidos y sencillamente.

Este *software* ofrece diferentes herramientas para procesar los archivos. Como primera medida, permite hacer un preprocesamiento automático del documento, es decir, determinar áreas de texto, imagen o tabla y aplicar correcciones según lo que la herramienta

crea necesario, en forma automática, y reconocerlas en consecuencia. Esto puede ser útil para archivos con una cantidad considerable de páginas o en los que la calidad del texto original sea muy buena.

Es importante aclarar que, como todo proceso automático, el preprocesamiento siempre requiere de verificaciones y ajustes posteriores. Siempre debe existir una valoración por parte del perito traductor acerca de cómo manejar el documento original, si resulta conveniente (o no) aplicar un preprocesamiento automático e, incluso, si resulta conveniente (o, mejor dicho, rentable) digitalizar y convertir su pericia. Todo dependerá del volumen, del plazo de entrega y de la predisposición del traductor para invertir tiempo en convertir los documentos y en «meter mano» en el procesador de texto luego de hacerlo.

Además, podemos «agrupar» diferentes archivos PDF en un solo documento convertido. Esto es, podemos seleccionar para la conversión muchos archivos (por ejemplo, que correspondan a diferentes fojas) y reunirlos en una misma conversión, para lograr un documento único, que será nuestro documento fuente al traducir.

Veamos la interfaz de trabajo habitual en ABBYY FineReader®:



La principal de las grandes funcionalidades que tiene esta herramienta es la posibilidad de editar nuestra conversión en el programa mismo, antes de enviarla a otras aplicaciones, como por ejemplo, MS Word. Si bien las opciones son más limitadas que en el procesador de texto, no dejan de ser útiles. En el panel de resultados del reconocimiento, a la derecha de la pantalla, podemos corregir cuestiones de fuente y formato, trabajar con texto oculto, ajustar palabras que tengan caracteres con confianza baja (que podemos visualizar en el panel de zoom) y agregar o eliminar lo que haga falta (por ejemplo, tabulaciones innecesarias). Toda esta edición es, como dijimos, previa al envío del documento editable a otras aplicaciones. Aquí encontramos otra pieza fundamental del *software*: la posibilidad de generar

documentos con diferentes grados de similitud con el original. Estamos hablando de lograr copias exactas del original (que respetan la diagramación del texto fuente; esta opción suele ser muy útil a la hora, por ejemplo, de generar archivos PPTX), textos con formato (donde, por ejemplo, se generan textos con viñetas automáticas, sangrías predefinidas, etcétera) y textos sin formato (esta es la opción más útil para los peritos, ya que en el ámbito pericial se trabaja con formato de traducción pública).

Una vez terminado el preprocesamiento automático, o luego de abrir el documento para hacer un procesamiento manual (véase más abajo), es necesario determinar o ajustar las áreas sobre las que el programa reconocerá el texto.

Las áreas son:



Al trazar áreas de texto e imagen, podemos, por ejemplo, ajustar los recuadros mediante las opciones de cortar o agregar partes de área, para hacer firme a las particularidades del original (por ejemplo, firmas solapadas sobre el texto principal), o incluso ajustar la orientación del texto para su reconocimiento.

Al trazar áreas de tabla, podemos ajustar la cantidad de filas y columnas al agregar o eliminar separadores verticales u horizontales, y también combinar o separar celdas para mejorar el reconocimiento que haya hecho el programa.

La opción de imagen de fondo permite trabajar con imágenes con texto. Lo que hace es mantener la imagen y convertir el texto en editable.



La herramienta «Reordenar» nos permite indicar al programa el orden en que debe leer o reconocer las diferentes áreas, que corresponderá al orden en que veamos la información en el documento de destino.

Tecnología de  
reconocimiento  
óptico de caracteres  
aplicada a la  
actividad pericial

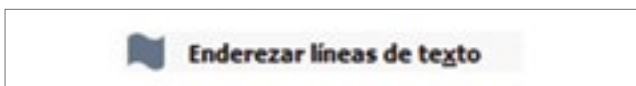
Si decidimos hacer un reconocimiento manual, es necesario desmarcar la opción de preprocesamiento automático al cargar el archivo. Esto puede ser útil cuando la calidad del original no es del todo adecuada, o bien cuando consideramos que la herramienta podría reconocer elementos en forma errónea y nos demandaría más tiempo ajustar lo que reconozca que trazar las áreas nosotros mismos.

Veamos las posibilidades que ofrece ABBYY FineReader® para editar el original en forma manual, con vistas a mejorar la calidad y precisión del reconocimiento. Con la posibilidad de aplicar cambios en la página que se muestre en pantalla, en páginas pares o impares o en todas las páginas de un documento, entre las opciones de edición de imágenes, podemos encontrar las siguientes:

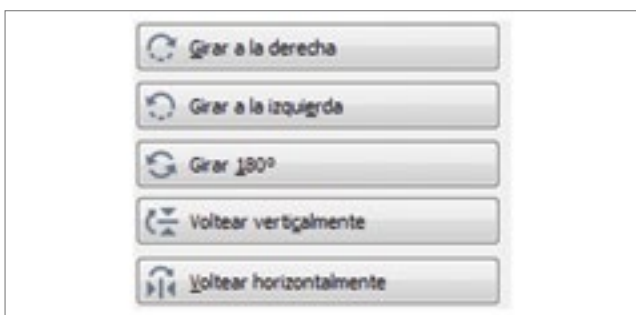
#### Herramientas que permiten corregir los defectos de texto:



Permite mejorar la inclinación de las líneas de texto.



Permite mejorar líneas de texto curvas.



Permiten indicar al programa la orientación real del texto.

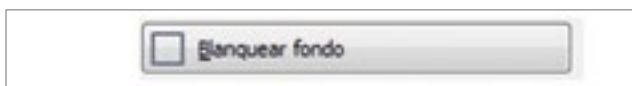


Permite eliminar marcas hechas con bolígrafo o resaltador.



Permite borrar áreas no deseadas que podrían afectar el reconocimiento (solo fondo blanco).

#### Herramientas que permiten corregir los defectos de imagen:



Permite transformar el fondo de la imagen en color blanco, para aumentar el contraste.



Permite invertir los colores de texto y fondo.



Permite mejorar los niveles de brillo y contraste de la imagen.

Con todas estas herramientas y opciones básicas (y no tanto) que ofrece ABBYY FineReader®, podemos lograr que nuestro trabajo como peritos sea más uniforme, rentable y de mejor calidad, además de saber «real» que deberemos traducir. Con estos objetivos en mente, es importante no solo que el traductor integre una herramienta de OCR en su entorno habitual de trabajo, sino también que se interese por optimizar el uso que hace de ella y aprender cada día un poco más. □

*Nota: Este artículo se basa en la versión 14 de ABBYY FineReader®.*