

# Tips for Making Accessible Scans

The Canon copiers at Bryn Mawr College automatically apply **Optical Character Recognition (OCR)** to scanned documents. In theory, this means that any scans you make will be saved as screen-readable PDFs. However, the OCR process only works *if the original and your scan are of decent quality*. This document contains advice for creating scans that will be successfully converted.

## What is Optical Character Recognition?

Ordinarily, a scan is simply a photograph of a page. The textual elements on that page are not editable, searchable text – they are simply patterns of light and dark. In order for a document to work with screen reader software, text-to-Braille software, highlighting and annotation tools, and other assistive technologies, these textual patterns need to be converted to actual text — that is, to a string of characters that can be edited, deleted, searched, etc. — through a process called *optical character recognition (OCR)*. The OCR software looks at the patterns of lights and darks and uses algorithms to determine which patterns are characters and which characters they are.

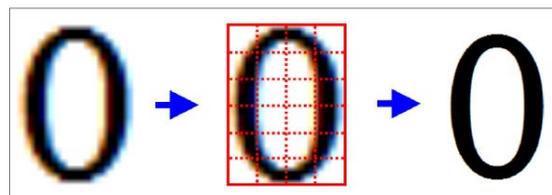


Fig. 1. Optical Character Recognition (OCR)

## Making Successful OCR Scans

Bryn Mawr's Canon copiers have been set to apply OCR to all scans automatically, so that all new scanned PDFs will be accessible. *However, the output of an OCR conversion is only as good as the input.* Originals of poor quality or bad scanning techniques can prevent the OCR process from working properly. If the OCR process can't correctly identify and interpret characters, the text it generates will be nonsense and your PDF will not be accessible.

Here are some tips for making sure that your scans will convert correctly:

- Start with a clean original. Highlighting, underlining, and page damage are primary culprits in preventing the OCR process from properly recognizing text.
- Scan with all pages oriented in the same direction and as close to horizontal or vertical as possible. The OCR process can correct for slight skewing, but text on a page that is very tilted will not be interpreted correctly.
- Avoid cutting off text or blocking it with your hands, bookmarks, etc. (If the text isn't visible, it can't be recognized!)
- For best results, scan one page at a time. The OCR process will usually recognize text documents scanned two-up — that is, with two facing pages in a book or journal scanned the same time — as having two columns of text. However, two-up scanning often creates shadows and distortions that can prevent parts of the text from being correctly interpreted. If each page of your original has multiple columns of text, you should absolutely scan one page at a time.

## Documents that Will Need Special Attention

Documents containing unusual fonts (i.e., scripts, Blackletter) or languages other than English may not OCR correctly, even if they are well scanned. Depending on a reader's disability, documents that contain mathematical or scientific notation, images, tables, or graphs may not be fully accessible even after the text has successfully been converted to a screen-readable format. If your documents have these elements, check with the reader in question or with Deb Alder, Bryn Mawr's Access Services Coordinator (x7351, [dalder@brynmawr.edu](mailto:dalder@brynmawr.edu)), to determine whether additional steps need to be taken to make those elements accessible.

