

ScanSoft®



OmniPage® SE FOR MACINTOSH

User's Guide

LEGAL NOTICES

©2002 by ScanSoft, Inc. All rights reserved. No part of this publication may be transmitted, transcribed, reproduced, stored in any retrieval system or translated into any language or computer language in any form or by any means, mechanical, electronic, magnetic, optical, chemical, manual, or otherwise, without prior written consent from the Legal Department at ScanSoft, Inc., 9 Centennial Drive, Peabody, Massachusetts 01960, U.S.A. Printed in the United States of America and in the Netherlands.

The software described in this book is furnished under license and may be used or copied only in accordance with the terms of such license.

IMPORTANT NOTICE

ScanSoft, Inc. provides this publication "as is" without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability or fitness for a particular purpose. Some states or jurisdictions do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you. ScanSoft reserves the right to revise this publication and to make changes from time to time in the content hereof without obligation of ScanSoft to notify any person of such revision or changes.

TRADEMARKS AND CREDITS

ScanSoft, OmniPage, OmniPage Pro, OmniPage Pro X, OmniPage SE, True Page, Direct OCR and Language Analyst are registered trademarks or trademarks of ScanSoft, Inc. in the United States and in other countries. Mac and Macintosh are registered trademarks of Apple Computer, Inc. in the U.S. and in other countries.

All other trademarks and trade names mentioned herein are hereby acknowledged and recognized as property of their respective owners.

ScanSoft Inc.

9 Centennial Drive
Peabody, MA 01960
U.S.A.

ScanSoft Europe BV

Randstad 22-139
1316 BW Almere
The Netherlands



CONTENTS

| | |
|--|-----------|
| Welcome | 7 |
| Chapter outline | 7 |
| Using this Guide | 8 |
| How to use online Help | 9 |
| Other online resources | 10 |
| New features in OmniPage Pro X | 11 |
| OmniPage SE for Mac and OmniPage Pro X | 12 |
| | |
| 1 Installation and setup | 13 |
| System requirements | 14 |
| Installing the software | 14 |
| Running the program under Mac OS 9 | 15 |
| Starting OmniPage Pro | 16 |
| Selecting your scanner | 17 |
| Registering OmniPage Pro | 20 |
| Removing OmniPage Pro | 20 |
| | |
| 2 Introduction | 21 |
| What is Optical Character Recognition? | 22 |
| Beyond OCR | 22 |
| Basic steps in the OCR process | 23 |
| The OCR Toolbar | 24 |
| The full OmniPage Pro interface | 25 |
| The Document window | 26 |
| The Thumbnail window | 26 |
| The Zone Info and Tools palettes | 27 |
| The Preferences dialog box | 28 |

| | |
|--|-----------|
| 3 Processing documents | 29 |
| Basic processing steps | 30 |
| Automatic processing | 30 |
| To prepare for automatic processing | 31 |
| To process a new document automatically | 32 |
| To process an existing document automatically | 33 |
| Manual processing | 34 |
| Steps for manual processing | 34 |
| Using automatic and manual processing together | 35 |
| Using the OCR Assistant | 36 |
| Bringing page images into OmniPage Pro | 38 |
| Scanning pages | 38 |
| Loading image files | 38 |
| Opening OmniPage Documents | 40 |
| Using drag-and-drop | 40 |
| Creating and modifying zones | 41 |
| Creating zones automatically | 42 |
| Specifying zone types | 43 |
| Drawing zones manually | 46 |
| Modifying zones | 48 |
| Table zones | 51 |
| Performing recognition | 52 |
| Performing OCR | 52 |
| Proofreading OCR results | 53 |
| Verifying recognized text | 55 |
| Color markers | 56 |
| Getting page information | 56 |
| Working with documents | 57 |
| Resizing a page display | 57 |
| Saving a document as you work | 58 |
| Moving to other pages | 58 |
| Reordering pages | 58 |
| Deleting a page | 59 |
| Undoing edits | 59 |
| Modifying images | 59 |
| Modifying text | 60 |
| Printing a document | 61 |

| | |
|--|-----------|
| Listening to a document | 62 |
| Closing a document | 62 |
| Quitting OmniPage Pro | 62 |
| Exporting documents | 63 |
| Saving an OmniPage Document | 63 |
| Saving images | 63 |
| Saving recognition results | 64 |
| Saving to Portable Document Format (PDF) | 66 |
| Copying a document to the Clipboard | 66 |
| Using drag-and-drop functionality | 67 |
| Direct OCR | 68 |
| Using Direct OCR | 69 |
| 4 Settings | 71 |
| OCR Toolbar options | 72 |
| Get Page options | 72 |
| Original Layout options | 74 |
| Style Set options | 75 |
| OCR options | 77 |
| Export options | 77 |
| Preference settings | 78 |
| Scanner settings | 78 |
| OCR settings | 82 |
| Spelling settings | 84 |
| Miscellaneous settings | 87 |
| 5 Customizing OCR | 89 |
| Specifying the style set | 89 |
| Specifying a global style set | 92 |
| Creating style sets | 92 |
| Applying and editing zone styles | 93 |
| Font mapping | 96 |
| Zone templates | 98 |
| Training OCR | 99 |
| User dictionaries | 103 |
| Settings files | 104 |

| | | |
|---|----------------------------------|-----|
| 6 | Technical information | 105 |
| | Troubleshooting | 106 |
| | Solutions to try first | 106 |
| | Low memory situations | 106 |
| | Low disk space situations | 107 |
| | Improving accuracy | 107 |
| | Improving fax recognition | 110 |
| | Interface problems and solutions | 111 |
| | System failure during OCR | 111 |
| | Supported languages | 112 |
| | Supported saving formats | 113 |
| | Supported image file formats | 114 |
| | Index | 115 |



Welcome

Welcome to OmniPage Pro X™, and thank you for buying our software!



OmniPage Pro is the professional OCR solution for Mac providing advanced OCR features. You may have the Special Edition, called OmniPage SE™ for Mac, which has been developed for distribution by scanner manufacturers. OmniPage SE for Mac contains a subset of the features of the professional product. This User's Guide and the online Help describe the features of the full product, using the SE icon shown on the left to document the differences between the two products. If you find the additional features of the professional product would be of benefit to you, you can use online facilities to upgrade your Special Edition to OmniPage Pro.

This User's Guide has been provided to help you get started and give you an overview of the program.

Chapter outline

Chapter 1, *Installation and setup*, tells you how to install and start the program and select a scanner. It lists the system requirements and provides guidance on registering the product.

Chapter 2, *Introduction*, explains the OCR process and how it forms part of the OmniPage Pro workflow. It also presents the program's main working areas and controls, starting with the OCR Toolbar.

Chapter 3, *Processing documents*, tells you how to do automatic and manual processing and how to combine them. It details processing steps: acquiring pages, zoning, recognizing, proofing and exporting.

Chapter 4, *Settings*, gives detailed information on each of the choices offered by the pop-up menus in the OCR Toolbar. It also guides you through the choices in the panels of the Preferences dialog box.

Chapter 5, *Customizing OCR*, provides information on some more advanced features, such as style sets and their zone styles, zone templates, training, user dictionaries and settings files.

Chapter 6, *Technical information*, gives troubleshooting advice and details the supported file formats and languages.

Using this Guide

This Guide supposes that you know how to work in the Macintosh® environment. Please refer to your Macintosh help resources if you have questions about how to use dialog boxes, menus, scroll bars, and so on. The following conventions are used in this Guide.

| Convention | Purpose |
|------------------------|---|
| <i>Italicized text</i> | <ul style="list-style-type: none">Emphasizes menu commands, dialog box options, button and file names: “Choose <i>Open...</i> in the File menu.”Names sections in this Guide.Emphasizes new terms the first time they are used. |
| Command key symbol (⌘) | Illustrates keyboard shortcuts. For example: ⌘C means hold the Command key down as you press the letter “c”. |
| Note or Tip | Introduces a tip or an item of note. |

How to use online Help

OmniPage Pro X has an extensive HTML-based online Help system. Click *Help Contents* or *Help Index* in the program's Help menu to open it. The Help system provides you with three tabbed panels:

- ◆ **Contents:** A three-level table of contents. Click a topic.
- ◆ **Index:** A two-level, alphabetical index. Enter a keyword or scroll to the desired location and click an entry.
- ◆ **Search:** Search keywords through the whole text of all help topics. It lists all topics containing the specified word(s).

For advice on other Help facilities, please consult the documentation for your HTML viewer.

Online help contains some topics not included in this User's Guide: an indexed glossary of terms, settings guidelines for a variety of document types, a Quick Start Guide for reading a sample image file, and documentation on Apple Event support and scripting.

▼ To get help on buttons and pop-up menus

Brief help is available without opening the online Help system. Hover the cursor over any button or pop-up list in the OCR Toolbar or the palettes. A concise description of the control appears in the status line along the base of the OCR Toolbar.

▼ To get help on topics and procedures

Select *Help Index* in OmniPage Pro's Help menu. Begin to type in a keyword you want to find. As you type in the first letters of a keyword, the Help system automatically shows you the first top-level index entry beginning with the letters typed in. OmniPage Pro's structured index helps you to quickly find answers for your questions.

Click an index entry to display its related topic. If an entry is linked to more than one topic, a pop-up list appears. Select the desired topic.

▼ **To browse through a series of topics**



Use the *Previous* and *Next* buttons top right of each topic. These allow you to view topics in the order they appear in the table of contents.

▼ **To view recently viewed pages**

Use the *Back* button to retrace your steps to your previously viewed topics.

▼ **To print a topic**

Select the *Print* button, specify a printer to be used and print settings.

Other online resources

Readme files, in HTML and PDF formats, are located on the installation CD. They contain last-minute information about OmniPage Pro X. Please read one of them before installing the application.

ScanSoft's web site www.scansoft.com includes a Scanner Guide with regularly updated information about supported scanners and related issues. Access the site from the online Help topic *Getting Help*.

New features in OmniPage Pro X

The family of OmniPage® products is now augmented by OmniPage Pro X for Macintosh. Here we summarize its most important new features compared to OmniPage Pro 8 for Macintosh.

- ◆ A better recognition engine has been integrated, capable of delivering greater accuracy, particularly on degraded documents.
- ◆ Support for the Mac® OS X operating system. A revised user interface exploits the improved display techniques of the new system. Support is maintained for Mac OS 9.
- ◆ A new Assistant facility provides interactive step-by-step guidance for users new to the world of OCR processing.
- ◆ Improved parsing of page elements to retain the formatting and layout of the original pages, in particular better retention of color graphics and smarter text/graphics detection.
- ◆ Better auto-detection and handling of tables and spreadsheets.
- ◆ Detection and recognition of reverse text (white or pale letters on black or dark backgrounds).
- ◆ Portable Document Format (PDF) files can be opened and their contents transformed to editable text.
- ◆ Recognized pages can be saved to Portable Document Format (PDF) files, ready for display, use on the Web or for file transfer.
- ◆ Export support added for MS Word 98, 2001 and X and MS Excel 98, 2001 and X.
- ◆ Improved export support for HTML (upgraded to HTML 4.0).
- ◆ Voice read-back facility for texts in English and Spanish.

The intelligent voice read-back facility and support of the HTML 4.0 and PDF formats are not included in OmniPage SE for Mac.



OmniPage SE for Mac and OmniPage Pro X



This list documents benefits which are not incorporated in OmniPage SE for Mac, but which can become available by upgrading to OmniPage Pro X:

- ◆ Significant improvement in recognition accuracy.
- ◆ Access to the training facility.
- ◆ Ability to load and read the contents of PDF files.
- ◆ Ability to save recognized documents to PDF format.
- ◆ Ability to load GIF image files.
- ◆ Handling Photoshop PSD, PNG and Windows BMP image files for input and output.
- ◆ Support for WYSIWYG HTML 4.0 output.
- ◆ Access to text-to-speech software, allowing recognized texts in English and Spanish to be read aloud.
- ◆ Support of Zone Templates containing user-defined zones for documents having the same layout.

For more information or to upgrade, please visit www.scansoft.com.

Chapter 1



Installation and setup

This chapter provides information on installing OmniPage Pro X and selecting a scanner to use with it.

Please consult the *Readme* file which provides the most up-to-date information on installing and running the program. *Readme* is supplied in HTML and PDF formats. These files are copied from the CD to the OmniPage Pro X folder (OmniPage SE folder for the Special Edition) during installation.

This User's Guide is also supplied in PDF format. It is copied to the sub-folder *User's Guide*. The Mac OS X operating system includes a PDF viewer. Under Mac OS 9, please use Adobe Acrobat. The PDF files can be navigated easily using the bookmarks (table of contents), page thumbnails and hyperlinks on cross references and index entries.

Please continue reading this chapter for the following information:

- ◆ System requirements
- ◆ Installing the software
- ◆ Running the program under Mac OS 9
- ◆ Starting OmniPage Pro
- ◆ Selecting your scanner
- ◆ Registering OmniPage Pro
- ◆ Removing OmniPage Pro

System requirements

The minimum system requirements for OmniPage Pro X are:

- ◆ iMac, iBook, PowerBook, Power Macintosh or PowerPC compatible computers with at least a G3 processor
- ◆ Mac OS 9.1 or later, Mac OS X (10.1 or above) and QuickTime 4.1 or later (this is normally included in OS X)
- ◆ 128 MB of memory (RAM) on Mac OS X; 64MB on Mac OS 9 with 32 MB allocated to OmniPage Pro (or 64 MB allocated to handle full-page color images with more than 256 colors)
- ◆ 80 to 100 MB of free hard disk space
- ◆ A color monitor with at least 256 colors and 800x600 pixel resolution
- ◆ A Macintosh-compatible pointing device
- ◆ A supported and correctly installed scanner, if you plan to scan documents.

Performance and speed will be enhanced if your computer's processor, memory and available disk space exceed minimum requirements.

Installing the software

▼ To install OmniPage Pro X:

1. Insert the OmniPage Pro CD in the CD-ROM drive.
2. Double-click *OmniPage Pro X Setup* (*OmniPage SE Setup* for OmniPage SE for Mac).
3. Select a language and then click *Continue*. This language will be used for installation and also as the program's interface language.

4. Read the license agreement. If you click *I Agree*, you can continue installation.
5. Personalize your copy in the dialog box that appears.
Type in your name, the name of your company and the serial number. You will find the serial number on the CD case.
(For OmniPage SE for Mac you may get a serial number automatically).
6. Click *OK*.
7. Click *Install* in the next dialog box to proceed. A further dialog box lets you choose where the OmniPage Pro files will be installed. Select a drive and optionally a folder location (using *Open* or *New*) and click *Choose*. The program will be installed in a folder named *OmniPage Pro X* (*OmniPage SE* for the Special Edition). If you want to keep a previous OmniPage version, install your new version to a different location.

All the program files will be copied to the chosen drive and location. Some sub-folders will be created, including Components, Help, Sample Files, Training Files, User Dictionaries, User's Guide, and Zone Templates.

Note Under Mac OS 9 you may get a warning message if you have no CarbonLib installed on your machine. In this case double-click the CarbonLib Setup. The required CarbonLib will be installed, the computer will then restart and the OmniPage Pro installation will start automatically.

Running the program under Mac OS 9

This User's Guide and the online help describe the use of the program under the Mac OS X operating system. Some dialog boxes have a slightly different appearance under Mac OS 9. Mac OS X supports an Application menu: it includes *Preferences...* which is in the Edit menu under Mac OS 9 and *Quit* which is in the File menu in Mac OS 9.



Online Help highlights all differences between Mac OS X and Mac OS 9 with an OS 9 icon.

The Help menu under Mac OS 9 allows you to show or hide balloon help. This relates to system-wide balloon help, which can appear within OmniPage Pro X under OS 9.

Starting OmniPage Pro



There are several ways of starting OmniPage Pro®:

- ◆ Open the *OmniPage Pro X* folder and double-click the OmniPage Pro X icon.

The program launches and the OCR Toolbar will be displayed. For quicker access, place an alias program icon on your Desktop.
- ◆ Drag and drop one or more image files onto the OmniPage Pro X icon.

The program launches and loads the dropped image files. It does not immediately recognize them.
- ◆ Drag and drop an OmniPage Document icon onto the OmniPage Pro X icon or double-click an OmniPage Document icon.

The program launches and opens the previously created OmniPage Document. See page 58 and *Saving an OmniPage Document* on page 63.
- ◆ Use the Direct OCR feature. See *Direct OCR* on page 68.



Selecting your scanner

Before you can select a scanner in OmniPage Pro X, its driver must already be installed on your system. It should also be tested, to be sure it is working properly with the scanning software supplied by its manufacturer. Consult the documentation supplied with your scanner.

You can select a scanner type in the Select Scanner dialog box. If you cannot find your scanner model in the scanner list in this dialog box, OmniPage Pro allows you to select a driver from one of the two general scanner driver types supported by the program. You can select either a *Photoshop plug-in* or a *TWAIN driver* depending on your scanner.

For specific scanner types which work with a TWAIN driver, you can choose whether to use their own interface or use OmniPage Pro's interface. For scanners using a Photoshop plug-in driver, its interface is always displayed while scanning.

Each scanner driver provides a different user interface, so the available options may vary.

Tip See an overview table in the online Help topic *Selecting a scanner*. This summarizes the user interface differences depending on which type of scanner driver is chosen.

▼ To select a scanner:

1. Switch on your scanner and start OmniPage Pro.
2. Choose *Preferences...* from the Application menu (Mac OS 9: Edit menu) then click the Scanner icon to display the Scanner panel.
3. Click the *Select...* button to get the Select Scanner dialog box.
4. Select a scanner manufacturer under *Manufacturer* in the Select Scanner dialog box.
5. Select a scanner model under *Scanner*.

6. Check the driver name under *Driver*. If you have more than one driver, select the one you want to use.
7. Click *Verify* to be sure the selected scanner is correctly configured.
8. Click *OK* to close the Select Scanner dialog box.
9. Click *Save* in the Preferences dialog box.

If the displayed scanner list does not contain the manufacturer or type of your scanner, you have two more choices under *Manufacturer*: **(Photoshop plug-in)** and **(TWAIN driver)**. To decide which of these general scanner drivers your scanner supports, refer to the documentation supplied with your scanner. See the next two sections for more details on selecting **(TWAIN driver)** or **(Photoshop plug-in)**.

Tip If you do not have a scanner at all, you can select **(Test)** under *Manufacturer* in the Select Scanner dialog box to simulate scanning.

▼ **To access a scanner through a TWAIN driver:**

1. Follow instructions 1-3 from the previous section, *To select a scanner*.
2. Select **(TWAIN driver)** under *Manufacturer*.
3. Select a driver name under *Scanner*.
4. Check that your scanner driver delivered by the manufacturer has appeared under *Driver* and select it, if it is not already selected.
5. Click *Verify* to check the functioning of your scanner.
6. Decide which user interface you want to use for your scanner: the driver's own interface or OmniPage Pro's interface. See the overview table in the online Help topic *Selecting a scanner* which summarizes the user interface functioning for different scanner drivers.
 - Select *Show TWAIN User Interface* if you want to use the user interface of your scanner driver.

- Deselect *Show TWAIN User Interface* if you want to start scanning from OmniPage Pro using the scanner settings in the Scanner panel of the OmniPage Pro Preferences dialog box.

7. Click *OK* to close the Select Scanner dialog box.
8. Click *Save* in the Preferences dialog box.

▼ **To access a scanner through a Photoshop plug-in:**

1. Copy your scanner driver from the Plug-Ins folder of the Adobe Photoshop program to the OmniPage Pro X: Components: Scanner Support: Plug-Ins folder.

It is assumed that the scanner driver delivered by the manufacturer has already been copied to the Adobe Photoshop program's Plug-Ins folder during scanner installation. If you do not have Adobe Photoshop installed or you do not know how to find your scanner driver plug-in, please see the *Support* section in www.scansoft.com for information and tips on locating Adobe Photoshop plug-ins.

2. Follow instructions 1-3 from the section *To select a scanner*.
3. Select **(Photoshop plug-in)** under *Manufacturer*.
4. Select the driver just copied under *Scanner*. Check the driver name under *Driver*.
5. Click the *Verify* button if you want to display the info panels. The driver's info panel will appear first, then the Scanner Info panel. Inspect and then close them.
6. Click *OK* to close the Select Scanner dialog box.
7. Click *Save* in the Preferences dialog box.

▼ **To scan in the Classic Environment:**

- Select *Scan in Classic Mode* in the Select Scanner dialog box if it is not already selected. Please wait while the program compiles a scanner list.

This option enables you to scan pages even if your scanner has

a driver for Mac OS 9 only. If the option is selected, scanning will be performed in the Classic Environment. If the option is deselected, scanning can only be performed with a scanner driver developed for Mac OS X.

The *Scan in Classic Mode* option is not available for selection under Mac OS 9.

Registering OmniPage Pro

ScanSoft's electronic registration runs at the end of installation. It will search an Internet connection to immediately perform the registration online. We provide an easy electronic form that can be completed in less than five minutes.

If you did not register the software during installation, you can do it later. You can go to register.scansoft.com or you can choose *Register OmniPage Pro X...* from the Application menu (Mac OS 9: Apple menu) to register online.



For a statement on the use of your registration data, please see ScanSoft's Privacy Policy.

Registering OmniPage SE for Mac may proceed slightly differently.

Removing OmniPage Pro

Move or copy any files you want to keep from the OmniPage Pro X folder. These might be settings, training, template, user dictionary, export or OmniPage Document files. Then drag the folder to the Trash.

Chapter 2



Introduction

You probably do business correspondence and other written projects on your computer. However, certain sources of information may not be immediately available for use. For example, if you want to incorporate part of a magazine article into a document in your word processor, you somehow have to get its text into your computer. Painstakingly retyping the article is not an appealing solution.

OmniPage Pro X offers a smart solution to increase your productivity. Its *optical character recognition (OCR)* technology accurately and easily converts text from scanned pages and image files into editable form for use in your favorite computer applications. You do not have to retype whole texts — OmniPage Pro does it for you.

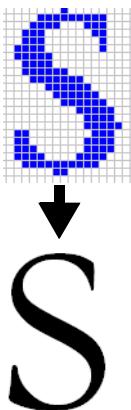
Please continue reading this chapter for information on these topics:

- ◆ What is Optical Character Recognition?
- ◆ Basic steps in the OCR process
- ◆ The OCR Toolbar
- ◆ The full OmniPage Pro interface

The OCR Toolbar is the control center for the program. The other main working areas appear when a document is started:

- ◆ **Thumbnail view:** this displays small images of each page.
- ◆ **Image view:** this displays an image of the current page.
- ◆ **Text view:** this displays the recognition results of the current page.

What is Optical Character Recognition?



Optical character recognition (OCR) is the process of extracting text from images. Images can result from scanning paper documents or opening image files. Images do not have editable text characters; they have many tiny dots (*pixels*) that together form character shapes. These present a picture of the text on a page.

During OCR, OmniPage Pro analyzes the character shapes in an image and determines character solutions to produce editable text. In other words, the OCR program 'reads' the page.

After OCR, you can export the recognized text to a variety of word-processing, desktop publishing, and spreadsheet applications.

Beyond OCR

In addition to text, OmniPage Pro X can retain the following elements in a document after OCR for display and export.

▼ Graphics

Photos, logos and drawings are examples of graphics. The program cannot recognize handwriting, but signatures can be saved as graphics.

▼ Text formatting

Font types, sizes, and styles (such as **bold** or *italic*) are examples of character formatting. Indents, tabs, margins and line spacing are examples of paragraph formatting.

▼ Page formatting

Column structure, paragraph spacing, and placement of graphics are examples of page formatting.

The elements that are retained depend on settings you select before OCR and on the capabilities of the saving format you choose. See chapter 4, *Settings*, for more information.

Basic steps in the OCR process

There are three main steps in OmniPage Pro's OCR process. They correspond to three large numbered buttons in the OCR Toolbar. Documents can be processed automatically or manually. In automatic processing, the *Start* button takes all specified document pages through the whole process (1-2-3) without a stop. Processing is done according to settings selected in pop-up menus on the OCR Toolbar and in the Preferences dialog box. In manual processing, each step can be performed separately and settings can be modified between each step. The three basic steps are:

1. Acquire page images

Scan pages or load one or more image files. See page 38. A miniature image of each page appears in Thumbnail view, the image of one page appears in Image view.

A *layout description* assists auto-zoning and a *style set* defines a formatting level for the recognized pages. When processing manually, zones should be drawn and styled at this point.

2. Perform OCR

Pages can be recognized with or without proofing. See page 53. During recognition, zones are automatically created on all pages without existing zones. On pages with zones, auto-zoning can be requested. OmniPage Pro performs OCR on text zones and can transfer graphics zones. Recognition results appear in Text view.

3. Export the document

The document can be saved to a specified file name and format, or copied to Clipboard. The document remains open in OmniPage Pro after its first export, allowing text to be further edited and pages added or re-recognized with changed settings and zoning. The document can be saved repeatedly, also to different saving formats.

It can be saved as an *OmniPage Document*, allowing it to be reopened later in OmniPage Pro X. See page 40, 58 and page 63.

See the topics *Automatic processing* and *Manual processing* at the beginning of chapter 3.

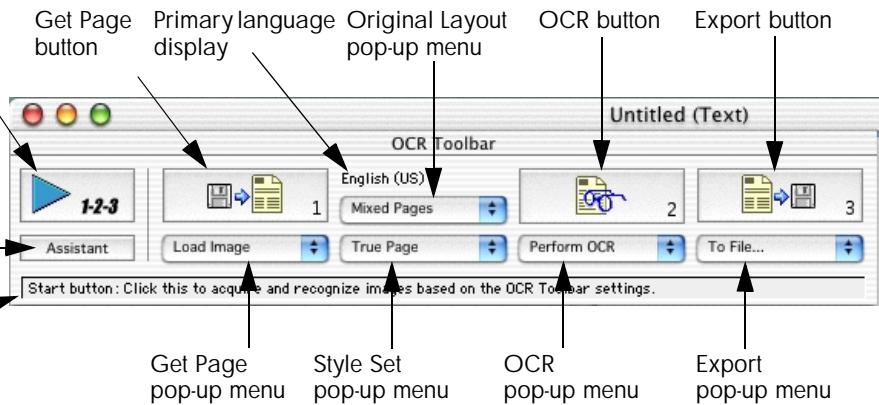
The OCR Toolbar

The OCR Toolbar appears when you first start the program. It is the control center for all document processing. The OCR Toolbar can be minimized under Mac OS 9.

Start button: Use this to start and re-start automatic processing, and to stop any processing.

Assistant button: Guides you to select settings and launches automatic processing.

The status line reports the current operation or the operation you can do next.



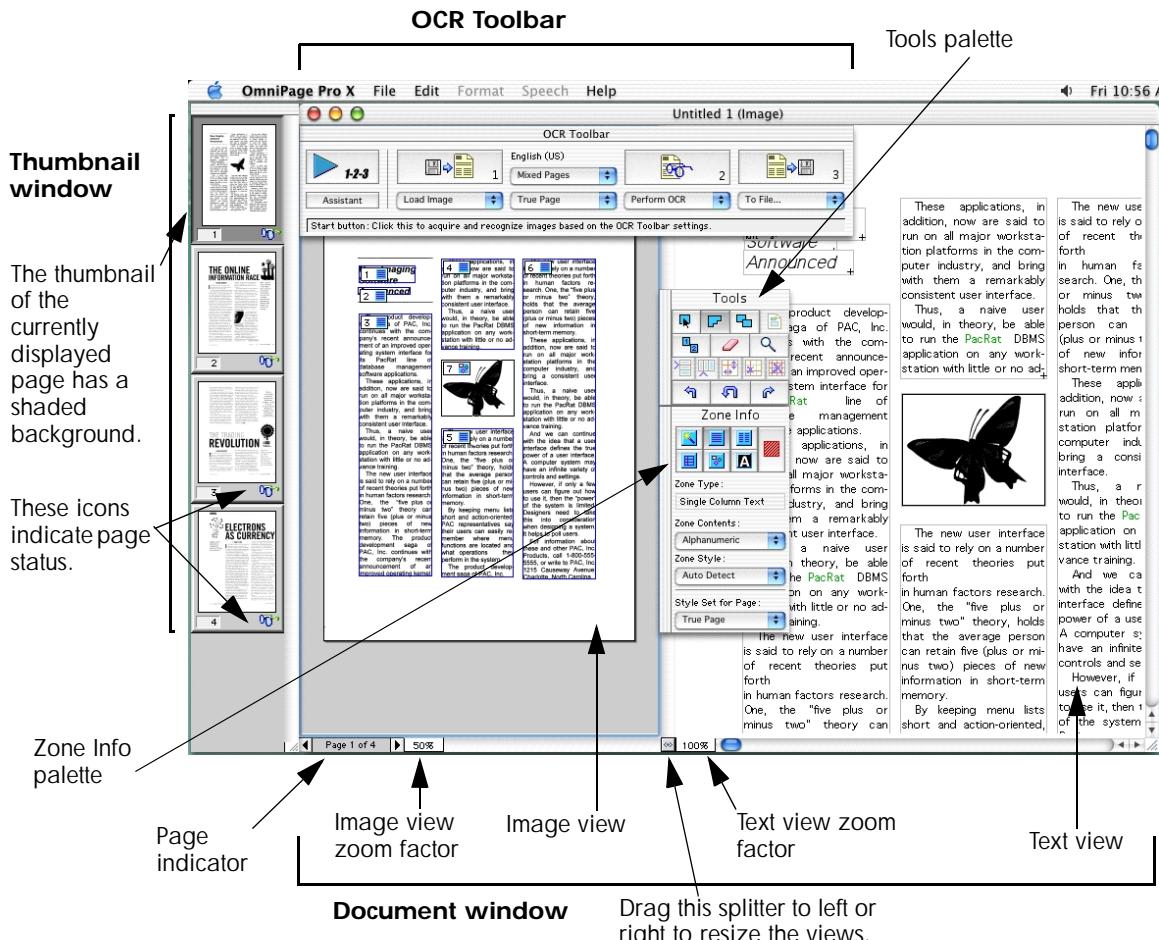
- ◆ The *Start* button lets you activate or re-activate automatic processing. When processing is in progress, it displays *Stop*.
- ◆ The *Get Page*, *OCR* and *Export* buttons are for manual processing. They allow each step to be performed separately, as follows:
 - The *Get Page* button lets you acquire one or more images from file or by scanning with the specified mode.
 - The *OCR* button lets you send the current page to recognition, or re-recognition, with or without proofing automatically started. It also allows training to be done. The last sentence does not apply to OmniPage SE for Mac.
 - The *Export* button lets you save results from all recognized pages in the document to file or copy them to Clipboard.
- ◆ The five pop-up menus let you select options. Processing is done according to the selected options. Before starting automatic processing, you must ensure all these options are suitable.
- ◆ The current primary recognition language is displayed. Three dots after the language name denote that at least one secondary language is also selected.

SE

The full OmniPage Pro interface

The full OmniPage Pro X interface appears when you start a document. The main screen areas of the interface are:

- ◆ The OCR Toolbar
- ◆ The Document window (with Image view and Text view)
- ◆ The Thumbnail window
- ◆ The Zone Info and Tools palettes
- ◆ The Preferences dialog box



The Document window

The Document window allows you to view and work with pages in the current document. You can drag this window to different locations. Original page images are displayed in Image view and recognition results are displayed in Text view. A highlight-colored border denotes which view is active. Click inside a view area to activate it.

Both views have scroll bars if the current page cannot be fully displayed. Click on the zoom control at the bottom left corner of a view to change its zoom factor. Choose from fixed or variable values (Zoom to Width and Zoom to View).

The splitter button at the bottom of the window lets you change the amount of space available for each view. To hide Image view completely, drag the splitter to the left edge of the Document window. To restore Image view, drag it to the right.

The Document window can be minimized and restored. Closing the document window closes the current document (with a warning if unsaved changes exist).

The Thumbnail window



The Thumbnail window appears vertically on the left of the desktop to provide Thumbnail view. This displays numbered miniature pictures (thumbnails) of all pages in the current document. You can use thumbnails to move to other pages, reorder or delete pages. An icon at the bottom right of a page indicates that the page has been recognized.

You can import one or more images to a defined location inside a document by drag-and-drop. You can also use a thumbnail to drag a copy of a page image from a document to the Desktop, a file location or into other applications.

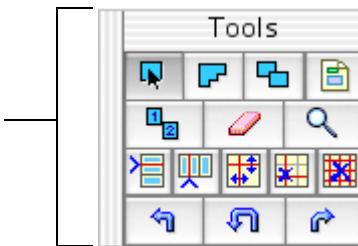
The Thumbnail window has a scroll bar and can be dragged to other locations. The window cannot be closed, under Mac OS 9 it can be minimized.

See *Working with documents* on page 57 for more information on using thumbnails for page operations.

The Zone Info and Tools palettes

The Zone Info and Tools palettes are displayed whenever Image view is active. You can drag them to different locations. Under Mac OS 9, they can be minimized and restored.

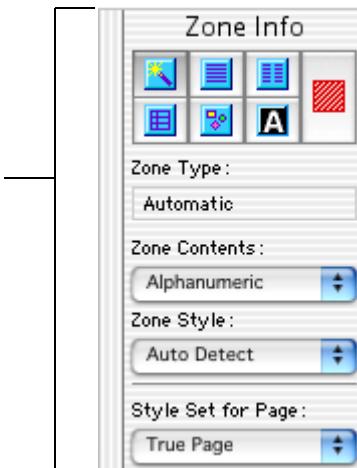
Use the Tools palette to draw regular or irregular zones, modify zones, apply a zone template (not in SE), reorder zones, erase parts of the image, zoom in or out on the image, handle table zones, or rotate an image.



See *Drawing zones manually* on page 46 for guidance on using each of these buttons.

Hover the cursor over any button in the palettes to read a description of its function in the status line at the base of the OCR Toolbar.

Use the Zone Info palette to select zone types, zone contents, zone styles, and a style set for the current page.



See *Specifying zone types* on page 43 and *Applying and editing zone styles* on page 93 for guidance on using these buttons and pop-up menus.

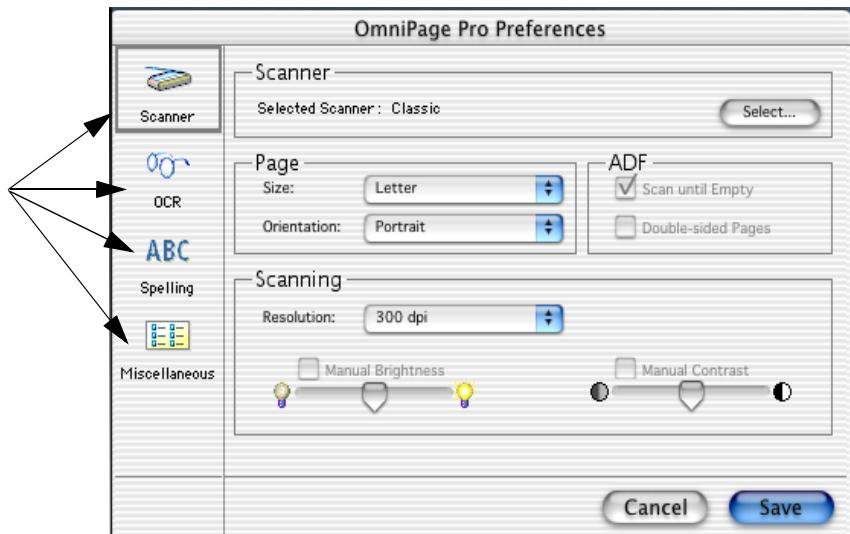
The style set True Page® lets you conserve the original page layout.

The Preferences dialog box

This dialog box is the central location for all OmniPage Pro settings not accessible through the OCR Toolbar. To open it, choose *Preferences...* in the Application menu (Mac OS 9: Edit menu).

The Preferences dialog box has four sections: Scanner, OCR, Spelling and Miscellaneous. Each section can be displayed by clicking its icon on the left.

Click each icon to view and select different groups of settings.



Guidance on selecting settings in each section is provided in chapter 4. You can save your set of preference settings to a Settings file, as described on page 104.

Note Online Help has a *Quick Start Guide*. This provides step-by-step instructions for reading a sample image file supplied with the program. The resulting document can be viewed in a target application and serves as a benchmark. You should be able to get similar accuracy from comparable documents of your own.

Chapter 3



Processing documents

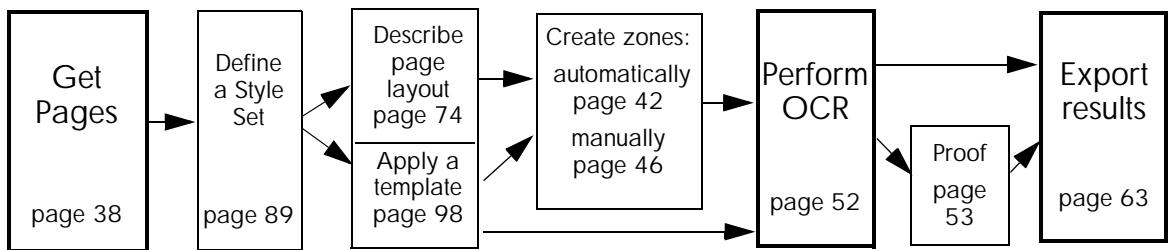
This chapter describes how to process documents in OmniPage Pro from start to finish. It tells you how the basic steps of OCR are linked during automatic and manual processing. It explains how you can exploit the advantages of each type of processing within a single document. The chapter also provides instructions for performing each OCR step and for other tasks you can do with your documents.

Please continue reading this chapter for information on these topics:

- ◆ Basic processing steps
- ◆ Automatic processing
- ◆ Manual processing
- ◆ Using automatic and manual processing together
- ◆ Using the OCR Assistant
- ◆ Bringing page images into OmniPage Pro
- ◆ Creating and modifying zones
- ◆ Performing recognition
- ◆ Working with documents
- ◆ Exporting documents
- ◆ Direct OCR

Basic processing steps

The following diagram summarizes how the basic steps are linked, and directs you to a page in this Guide. This workflow is broadly valid for both automatic and manual processing. The steps performed by the three basic OCR Toolbar buttons have a darker border.



Automatic processing

You can use the *Start* button to process a new document from start to finish or to finish processing an open document. The operations that occur when you click *Start* depend on the options selected in the OCR Toolbar's pop-up menus.



For example, OmniPage Pro can scan a stack of pages from a scanner's automatic document feeder (ADF), create zones on all pages, recognize the pages, offer the results for proofing, and then let you save the recognition results to file.

During automatic processing, auto-zoning always runs, unless you specify a zone template file. If you want to draw or modify zones manually, you can do this after recognition and first export are finished, and then re-recognize those pages afterwards.

To prepare for automatic processing



- 1. Select the source for one or more page images.**
Choose *Load image* to open one or more page images from file.
Choose *Scan in B&W* to scan in black-and-white.
Choose *Scan in Gray* to scan in grayscale.
Choose *Scan in Color* to scan in color (with a color scanner).
See page 38 and page 72 for information on these choices.
- 2. Select a style set.**
Choose a style set to define the formatting level and page layout you want applied to the recognition results.
See page 74 and page 75 for information on these choices.
- 3. Select a page layout description.**
Choose a page layout description to influence the auto-zoning.
Choose from Single Column, Multiple Column, Spreadsheet or Mixed Pages. Or choose a zone template if you have one.
The last sentence does not apply to OmniPage SE for Mac.
- 4. Select the type of recognition you want.**
Choose *Perform OCR* to have recognition without proofing. You can still proof the text later, after its first export. See from page 52.
Choose *OCR & Proof* to have proofing started as soon as all pages are recognized. See page 53.
Train OCR is not available in OmniPage SE for Mac.
- 5. Select an export target for the document.**
You can direct your document to be saved to a file whose name, location and type you define, or have the recognition results copied to the Clipboard. See page 66.
- 6. Ensure all other settings are in order.**
Settings are located in the Preferences dialog box (see chapter 4). These include recognition languages, user dictionaries and scanner settings. If you are scanning, place your page(s) correctly in the scanner. To scan multiple pages from an ADF, select *Scan Until Empty* in the Scanner Panel of the Preferences dialog box.
- 7. Click the *Start* button to launch automatic processing.**
Automatic processing proceeds as described in the next topic.

To process a new document automatically

We assume you have started OmniPage Pro X and can see the OCR Toolbar, but you have no document open and all settings are ready.



1. Click the *Start* button to launch automatic processing.
2. All specified pages are scanned or the Load Images dialog box lets you select image files. The status line reports progress as images are acquired. Page images appear briefly in Image view.
3. A miniature image of each page appears in Thumbnail view as it is acquired. Image view displays each page; when all pages are acquired, it displays the first acquired page.
4. Recognition starts with a progress monitor in the OCR Toolbar status line. Automatic or template (not in SE) zoning is done, text is detected and recognized on one page after the other.
5. The first image appears again in Image view with zones. Its recognition results appear in Text view.
6. If proofing was requested, it starts from the top of the first page. Make corrections as desired. Click in Text view to interrupt proofing. Then you can edit or verify the recognized text, move to other pages or change settings. The proofreading button *Ignore* becomes *Start*. Click this to resume proofreading. Click *Done* to finish proofing before the end of the document.
7. The Export dialog box appears if you chose export to file. Define a folder, file name and saving format, and choose other export options. If you chose *Save and Launch*, the recognition results will appear in the target application. If you chose export to Clipboard, a message tells you when the recognition results have been placed. The document remains open in OmniPage Pro for further editing. Pages can be re-recognized with changed zoning or settings. New pages can be added. The document can be saved repeatedly.



During processing, the *Start* button becomes a *Stop* button. Click it to stop processing. The current processing step is discarded but the results of all completed steps remain. For example, if you click *Stop* during OCR, there will be no recognized text but the image remains.

To process an existing document automatically

You can also click *Start* to perform automatic processing when you have a document open. It does not matter whether its pages were processed automatically or manually. To scan new pages into the document, place them in the scanner correctly. When you click *Start*, the OCR Instructions dialog box offers you the following choices.

◆ Load and Process Additional Pages

If the selected source is from file, the Load Images dialog box appears, allowing you to specify files. Otherwise, scanning will start immediately. If *Scan Until Empty* is selected, all pages in the ADF will be scanned one after the other. All specified pages enter the document and are recognized. Existing pages remain unchanged, even if some of them were unrecognized. If the current page was the last in the document when you clicked *Start*, the new pages are appended to the end of the document. If not, the Acquire Images dialog box lets you specify where to place the new pages. When recognition (and optionally proofing) are completed, the whole document is exported: sent to Clipboard or saved to file through the Export dialog box.

◆ Process All Unrecognized Pages

Recognition (and optionally proofing) is performed on all unrecognized pages. No new pages can be added if this option is selected. When processing is finished, or if there are no unrecognized pages, export starts, to Clipboard or file as specified. When saving to file, the Export dialog box appears. All changes to all pages are saved, not just the pages recognized by this command.

◆ Reprocess All Pages

All recognition results for all recognized pages in the document will be discarded, and all images will be (re-)recognized. Any image without zones is auto-zoned. If any zones exist, the Zoning Instructions dialog box lets you choose to use current zones only, to discard all zones and have auto-zoning, or to run auto-zoning in addition to existing zones. Your choice will be applied to all pages containing manually drawn or modified zones.

Manual processing

You can use manual processing when you want greater control over the OCR process. Processing proceeds step-by-step. This allows you to view and manually zone images before you send them for recognition. It also lets you modify settings between each processing step or from page to page. That can be important if some pages in the document need different settings from others.

During manual processing you can acquire multiple pages with each click of the *Get Page* button. Similarly, the *Export* button is for exporting recognition results from all recognized pages in the document. By contrast, the *OCR* button is used to have only the current page processed.

Steps for manual processing

Three OCR Toolbar buttons let you control the process step-by-step:

1. Acquire images

Define the image source in the Get Page pop-up menu. Choose to scan pages or to load one or more image files. Click the *Get Page* button (number 1). A miniature image of each page appears in Thumbnail view, the image of one page appears in Image view. Recognition does not start. See *Bringing page images into OmniPage Pro* on page 38 and *Get Page options* on page 72.

2. Create zones on the images

Draw zones in Image view using the Tools palette. Zones are areas that define which parts of a page image should be recognized. You can also load template zones and draw zones in addition to the zones placed from the template. See *Creating and modifying zones* on page 41 and *Zone templates* on page 98. OmniPage SE for Mac does not support Zone Templates.

3. Perform OCR

Specify to have recognition, with or without proofing, or to do training in the OCR pop-up menu. Click the *OCR* button (number 2). Choose to use existing zones only or to allow auto-

zoning on all unzoned parts of the page. Any page without zones will be auto-zoned. You will see a progress indicator as the current page is recognized. After OCR, recognition results appear in Text view. If you requested proofing and there are suspect words on the page, proofing begins immediately. If you did not request proofing, you can view, edit and verify the recognized text or start proofing from any point in the text.

See *Performing OCR* on page 52 and *Training OCR* on page 99.

4. Export the document

Specify an export target in the Export pop-up menu. You can save recognition results to one or more files, or have them copied to the Clipboard. Click the *Export* button (number 3). If you are saving to file, specify the file name, format and location.

See *Exporting documents* on page 63 for more information.

Using automatic and manual processing together

Automatic processing provides speed and efficiency. After you have selected settings, many pages can be processed from start to finish without user intervention. Manual processing demands more attention, but gives the user greater control over the recognition results. It is possible to tap into both benefits while processing a single document. Suppose you have a long document, ideally suited to automatic processing, except for a few pages needing separate zoning or settings. We provide two examples of how you could proceed.

- ▼ **To start automatically and finish manually:**
 1. Prepare settings and then process all pages automatically.
 2. Export the document to protect it, maybe as an OmniPage Document.
 3. Examine the recognition results, especially on pages you think will need individual attention. Identify which changes are needed to zoning or settings.

4. Make the required changes on a page and reprocess it manually by clicking on the *OCR* button.
5. Specify a choice in the Zoning Instructions dialog box.
6. Repeat steps 4 and 5 until all pages are adequately recognized.
7. Export the finished document as required.

▼ **To start manually and finish automatically:**

1. Prepare settings and acquire all the images for the document by clicking the *Get Page* button.
2. Examine the images for suitable brightness, orientation and content. Rescan or rotate unsuitable images. Use the eraser tool or zoning to remove or exclude spotty and degraded areas. Reorder pages as desired.
3. Manually zone pages needing special attention. Place pictures or diagrams in *Graphics* zones and areas you do not want recognized in *Ignore* zones. Draw and specify text zones.
4. Click the *Start* button and choose *Process All Unrecognized Pages* in the OCR Instructions dialog box.
5. Make a choice in the Zoning Instructions dialog box for all pages. Choose *Use Only Current Zones* or *Keep Current Zones and Find Additional Zones*.
6. After proofing (if requested), you can export the document.

Using the OCR Assistant

The OCR Assistant is a useful guide to users new to OmniPage Pro. It takes you through six panels, using questions and advice to help you choose suitable settings. It then launches automatic processing.

The OCR Assistant can be started only when no other document is open. It offers the choices currently set in OmniPage Pro. Some settings are not offered by the OCR Assistant; these should be selected in the Preferences dialog box before starting. They are as follows:

- ◆ **Scanner:** All settings. Be sure to turn on *Scan Until Empty* if you want to scan multiple pages from an ADF.
- ◆ **OCR:** A training file and options for saving graphics.
- ◆ **Spelling:** A user dictionary and Language Analyst® options.
- ◆ **Miscellaneous:** Retain or drop table grids.

Assistant

Click the *OCR Assistant* button to start moving through the six steps:

Step 1, Acquiring images: Choose one of the scanning modes (black-and-white, grayscale or color) or to load image files. If you are scanning pages, place them in the scanner.

Note You can scan pages only if you have previously selected a scanner through the Preferences dialog box. If you are scanning through the TWAIN interface, use it to choose the scanning mode.

Step 2, Language choices: Choose a primary language and, if desired, one or more secondary languages. Press the command key as you click to make or remove multiple selections.

Step 3, Proofreading: Choose to proofread text immediately after recognition or to proceed to first export without proofing.

Step 4, Original layout: Choose an option that best describes your incoming pages to guide the auto-zoning process.

Step 5, Format retention: Choose how much formatting you want in your exported document.

Step 6, Export: Choose to save to file or copy to Clipboard.

Click *Finish* to launch automatic processing, as already described.

The document remains in OmniPage Pro after first export. Pages can be added or re-recognized with changed settings. It can be exported repeatedly, to the same or other file formats.

Settings changed in the OCR Assistant remain valid in OmniPage Pro. If you have another document to process which needs the same settings, you do not have to run the OCR Assistant again. Just click the *Start* button to have it automatically processed.

Bringing page images into OmniPage Pro

This section describes the different methods for acquiring images:

- ◆ Scanning pages
- ◆ Loading image files
- ◆ Opening OmniPage Documents
- ◆ Using drag-and-drop

Scanning pages

You can scan a paper document to generate an electronic image. See *Starting OmniPage Pro* and *Selecting your scanner* in chapter 1.

▼ To scan pages into OmniPage Pro:

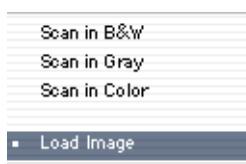
1. Place a page in your scanner. You can scan a stack of pages if you have an automatic document feeder (ADF).
2. Select one of the scanning modes in the *Get Page* pop-up menu.
3. Choose *Preferences...* in the Edit menu and open the *Scanner* panel to make sure the appropriate settings are selected for your page. See page 78. If you want to sequentially scan all pages in an ADF, make sure that *Scan Until Empty* is selected. Otherwise, you must click the *Get Page* button to scan each subsequent page.
4. Click the *Get Page* button in the OCR Toolbar.

Pages are scanned in order and the resulting images appear in Thumbnail view. The first page is displayed in Image view.

Loading image files

An image file is an electronic picture of text, such as a fax or scanned image, that is saved in an image file format. You can load more than one file at once. You can also load selected or all pages from multi-page image files. See *Supported image file formats* in chapter 6.





▼ **To load a single page image file:**

1. Select *Load Image* as the option in the Get Page Pop-up menu.
2. Click the *Get Page* button. The Load Images dialog box appears. It is a standard Macintosh dialog box.
3. Specify in the *Show* pop-up menu which files should be listed: All image files, or only files with a single format.
4. Select the folder containing your file with the *From* pop-up menu.
5. Select the file you want to load and then click *Open*. Or, double-click the file name.

The image from the file is displayed in miniature in Thumbnail view and at the specified magnification in Image view.

▼ **To load multiple images from file:**

1. Select *Load Image* in the Get Page pop-up menu and click the *Get Page* button. Select which file types should be listed.
2. Under the OS X operating system, select files as follows:
 - Files listed together: Shift+click the first and the last file names. These files and all in between will be selected.
 - Non-adjacent files: Command+click each file. Command+click a selected file to deselect it.
3. Click *Open* after you have selected all the files you want to load. Image files are loaded in the order they are listed and combined into one working document.
4. When opening a multi-page image file (TIFF or PDF), you can select which pages to open. Miniature page images appear in Thumbnail view and the first page is displayed in Image view.
5. Drag page images to new locations in Thumbnail view if the pages do not appear in the desired order.

Note If you scan or load pages while a document is currently open with its last page displayed, new pages are appended to the end of the document. If the last page is not the active one, you will be asked where to place incoming pages.

Opening OmniPage Documents

You can open an *OmniPage Document* using the *Open* command in the File menu. An OmniPage Document (OPD) is a file in OmniPage Pro's proprietary format. OPDs contain original page images, zones, settings and recognition results (if any). Each piece of recognized text remains linked to the image it came from, so text can still be proofed and verified when the OPD is reopened. You can also make editing changes to recognized text, re-recognize pages and add further pages to the document. You can save recognition results from the OPD more than once, for instance to different file formats.

Note OmniPage Pro can only have one working document open at a time. If you try to open another file while you have a document open, you are prompted to close the current document. However, you can add pages to your current document using the *Get Page* button.

▼ To open an OmniPage Document:

1. Choose *Open...* in the File menu.
The Open OmniPage Document dialog box appears.
2. Open the folder where your OmniPage Document is located.
3. Double-click a file name or select the file and click *Open*.
The OmniPage Document opens with one thumbnail image for each page. The original image of the first page appears in Image view and its recognition results (if any) in Text view. Some settings from the OPD are activated.

Note For advice on saving OmniPage Documents, see page 58 and page 64.

Using drag-and-drop

You can import images into an open document by drag-and-drop from the Desktop or Finder. Use Shift-clicks to select multiple files. You can import multi-page image files; the Select Pages dialog box allows you to specify which of the file's pages to open.

If you drag and then drop the image icon on Image view, the page or pages are appended to the end of the document.

If you drop the image icon on Thumbnail view, you can choose where to have the page(s) placed. As you drag the icon over the pages, a black bar appears between two pages. Drop the icon to have the new page(s) placed immediately below the bar.

The first of the imported pages becomes the current page.

You can launch OmniPage Pro X and load one or more images to start a new document. Drag an image file icon from the Desktop or Finder onto the OmniPage Pro X icon.

If you drag an image file icon onto the OmniPage Pro icon when you have the program running with a document, the new image is appended to the document if its last page was active, otherwise a dialog box lets you specify where to place the new image(s).

You can also launch the program by dragging the icon of an OmniPage Document onto the program icon, or by double-clicking the OPD icon. You cannot drag an OPD file into an open document. In this case, you will be invited to save any changes to the current document before it is closed and the OPD opened.

Note To use drag-and-drop to export recognition results, see page 67.

Creating and modifying zones

Page images are displayed in Image view. This is where *zones* can be manually created before OCR. Zones are bordered areas that identify parts of a page that will be recognized as text, retained as graphics or ignored. Any part of a page not enclosed by a zone is ignored during OCR, unless you specify that auto-zoning should run.

Note You can create zone templates (not in SE) to use when you process documents with the same zoning requirements. Zone templates remember the shape, position, order, type, contents, and style of zones. See *Zone templates* on page 98.

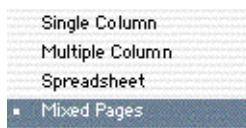
This section presents the following topics:

- ◆ Creating zones automatically
- ◆ Specifying zone types
- ◆ Drawing zones manually
- ◆ Modifying zones

Creating zones automatically

OmniPage Pro can create zones automatically for you. To do so, it uses the selected page layout description to find blocks of text and graphics on the page, place these in zones and decide a reading order.

▼ To run auto-zoning during automatic processing:



1. Choose a setting in the Original Layout pop-up menu that most closely matches the layout of your page or pages. Select Single Column, Multiple Column, Spreadsheet, Mixed Pages, or a template of your own (not in SE). See *Original Layout options* on page 74 for more information on these settings.
2. Check all other settings, then click the *Start* button to begin automatic processing. This will include auto-zoning (unless you applied a template and chose *Use Only Current Zones*). After recognition, the automatically detected zones are displayed in Image view. Each zone has a number indicating the order in which it was recognized. The zone icon next to the number indicates the zone type. If the zone locations, types or order are not suitable, change the zoning and then re-recognize the page.

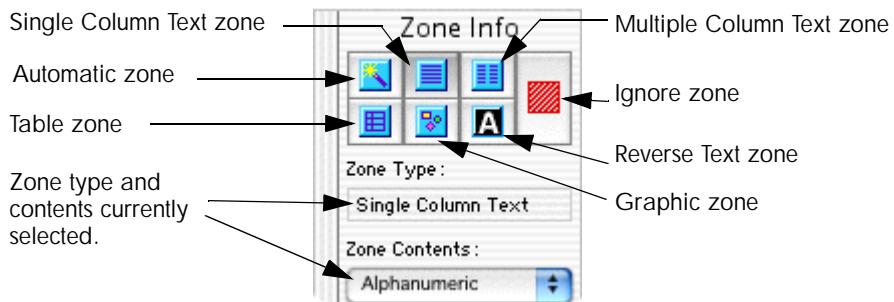
▼ To run auto-zoning during manual processing:

1. Choose a setting in the Original Layout pop-up menu that most closely matches the layout of your page or pages.
2. Click the *OCR* button to have the current page zoned and recognized. If there are no zones on the page, OmniPage Pro will automatically create zones and display them after recognition. If the page has at least one zone, the Zoning Instructions dialog box offers the following choices:

- Use Only Current Zones (auto-zoning will not run)
- Discard Current Zones and Find New Zones
- Keep Current Zones and Find Additional Zones.

Specifying zone types

All zones are identified as a particular type. This determines the way they are treated during OCR. You can specify zone types using the tools at the top of the Zone Info palette. This palette always appears when Image view is active.



The Zone Type display box tells you the zone type of the currently or last selected zone. The corresponding zone type tool has a 'pushed-in' appearance. When multiple zones with different types are selected, the display box will show 'Mixed Zone Types'.

Click a tool to change the zone type. This will apply to all currently selected zones (if any) and to new zones drawn from now on. Here are the properties of the different zone types:

▼ **Automatic zone type**



This zone type gives OmniPage Pro the right to make its own decisions on how to handle the contents of the zone. It decides whether the zone contains text or graphics. It decides whether text is in columns or not and reversed or not. Any side-by-side columns detected are treated as flowing text (moving top to bottom, then left to right). *Automatic* zones have purple borders. After recognition, the automatic zone may be replaced by a set of smaller zones.

▼ **Single Column Text zone type**



OmniPage Pro treats all contents as one block of text; it does not look for columns or detect graphics. Tabs are inserted between any side-by-side columns detected within a zone, so this zone type can be used for tables or texts in columns you do not want decolumnized or placed in a table grid. These zones have blue borders (denoting a zone containing text).

▼ **Multiple Column Text zone type**



OmniPage Pro tries to find columns within the zone area. If it finds them, the text is decolumnized (unless True Page is selected as the style set). After recognition, each column is likely to have its own zone. Graphics will not be detected inside the zone area. These zones also have blue borders.

▼ **Table zone type**



OmniPage Pro will treat the zone contents as a table. The contents will be placed in a table grid or in tab-separated columns, as requested in the Miscellaneous panel of the Preferences dialog box. These zones have orange borders and dividers. They must be rectangular (not irregular).

▼ **Graphic zone type**



OmniPage Pro treats all contents as a graphic area; it will not extract text from the zone. If *Retain Graphics* is selected, it copies the image area and transfers it to Text view. If True Page is selected as the style set, the graphics areas appear in frames in their original locations. In all other cases, the graphics are placed at the end of the recognized text from the page. These zones display a graphic icon and have black or white borders, depending on the background color.

▼ **Reverse Text zone type**



If the page contains reverse text (white or pale letters on a black or dark background), place this in a separate reverse text zone. The text will be recognized and displayed as normal text. If you want the text

reversed in your output document, do this in your target application. These zones have black or white borders, depending on the background color.

▼ **Ignore zone type**



OmniPage Pro ignores the zone entirely during auto-zoning. This is useful if you want OmniPage Pro to draw zones automatically but first want to identify areas to be ignored. By excluding complex tables or areas of line-art you do not need, you can speed up processing considerably. These zones have red borders and stripes.

Tip You can change the zone type of individual zones any time before OCR. For example, suppose auto-zoning placed a Single Column Text zone over two columns of text. If you do not want tabs inserted between the two columns, you can change the zone type to *Automatic* or *Multiple Column Text*. The columns will then be recognized separately and text will flow from one column to the next.

▼ **To specify a zone type:**



1. Click the Draw/Select Zones tool in the Tool palette if it is not already selected.

If the Tools palette is not visible, check that Image view is active and (in Mac OS 9) that the palette has not been minimized.

2. Select the zone you want to identify by clicking it.
 - Shift-click to select additional zones.
 - Double-click the Draw/Select Zones tool or choose *Select All* in the Edit menu to select all zones on the current page.

3. Click the desired zone type in the Zone Info palette. The zone type of all selected zones will change accordingly. This value will also be used for new zones that you draw.

▼ **To specify zone contents:**

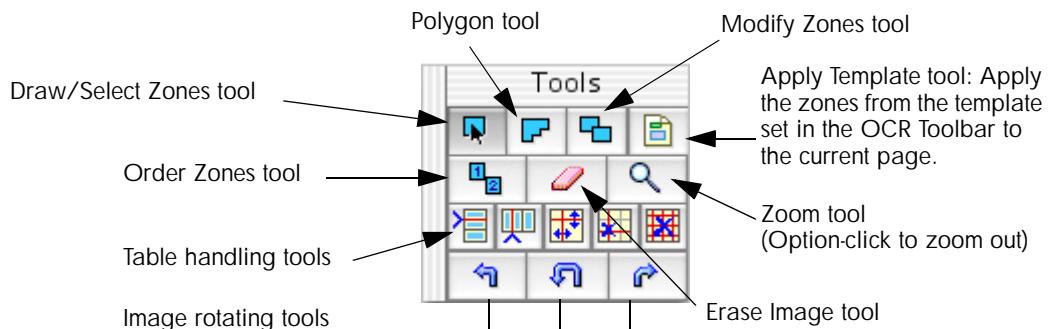
1. Select a zone whose zone contents you want to modify.

Zone contents can be specified only for text zones, that is for Automatic, Single Column Text, Multiple Column Text, Table or Reverse type zones.

2. Select Alphanumeric or Numeric in the Zone Contents pop-up menu.

Drawing zones manually

You can draw and modify zones using tools in the Tools palette. If the Tools palette does not appear, check that Image view is active and the palette is not minimized (Mac OS 9 only).



You can use the tab key to cycle through the zone tools when Image view is active.



The Apply Template tool is not available in OmniPage SE for Mac.

▼ To draw a rectangular zone:



1. Click the Draw/Select Zones tool in the Tools palette if it is not already selected. The mouse pointer becomes a drawing tool.
2. Make sure no existing zones are selected.
3. Click the appropriate zone type in the Zone Info palette. For example, click the *Graphic* type to draw a zone around a photo. See *Specifying zone types* on page 43.
4. Enclose an area of the image you want as a zone by holding down the mouse button and dragging the drawing tool to form a rectangular box.
5. Release the mouse button when you are done.
After drawing a zone, you can resize it by dragging its handles.

6. Repeat steps 3–5 until you have finished drawing zones around each area that you want to process.

You can draw up to 64 separate zones. Draw zones in the order you want them processed. A number at the top left of each zone indicates the reading order.

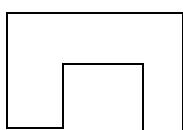
If you draw a zone over an existing one, the borders of the new zone will wrap *around* the existing zone. The zones will not overlap.

▼ **To draw an irregular zone:**

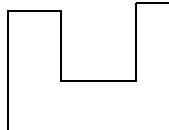


1. Click the Polygon tool in the Tools palette. The mouse pointer becomes a drawing tool in Image view.
2. Make sure no existing zones are selected.
3. Click the appropriate zone type in the Zone Info palette.
4. Position the drawing tool where you want to start drawing the first side of the zone and click the mouse button once.
5. Move the drawing tool to form the first side of your zone.
6. Click the mouse button again when the dotted line has the desired line length. The line becomes solid.
7. Draw a perpendicular line in either direction and then click to form the next side of the zone.
8. Repeat step 7 to finish drawing each side of your zone.
9. Double-click to close the shape.

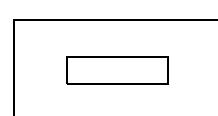
You will not be allowed to draw a line if it constitutes a restricted shape. The following zone shapes are restricted:



Indented along the bottom



Indented along the top



Hole in the middle

If you draw an irregular zone when the zone type is set to *Table*, it will change to *Single Column Text*. You cannot change the zone type of an irregular zone to *Table*.

Modifying zones

Zones can be modified before OCR takes place. You can move, copy, resize, reorder, extend, connect, divide, and delete zones. If you modify zones after recognition, you will have to re-recognize the page for the modifications to take effect.

The Modify Zones tool is for adding and subtracting zone areas.

Typically, this results in irregular zones, so it is not available for table type zones. This tool is also for connecting and dividing zones.

▼ To move zones:



1. Click the Draw/Select Zones tool in the Tools palette if it is not already selected.
2. Place the mouse pointer inside a zone.
3. Hold down the mouse button and drag the zone where you want to move it. Or use the arrow keys. Only the zone borders are moved. The contents of the page image remain as is.

▼ To resize zones:

1. Click the Draw/Select Zones tool if it is not already selected.
2. Select the zone you want to resize by clicking it. Handles appear on the zone border.
3. Select a handle, hold the mouse button down, and drag the mouse pointer in the direction you want to enlarge or reduce the zone.
4. Release the mouse button when you are done. The zone border changes to display the modified zone area.

▼ To reorder zones:



1. Click the Order Zones tool. The numbers in the zones disappear.
2. Click within the zone you want to have recognized first. The number 1 appears in the zone.
3. Click within the next zone you want recognized. The number 2 appears in the zone.

4. Continue until all the zones are appropriately ordered.

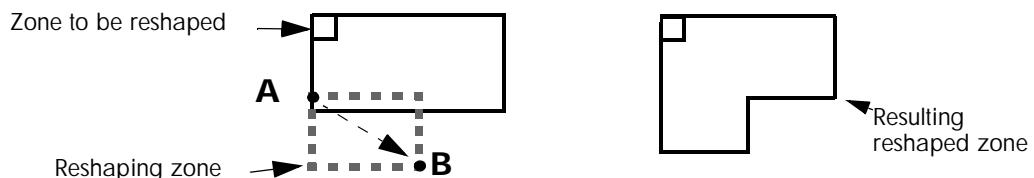
If you do not number all the zones, they will be automatically numbered when you select another tool or start OCR. Unless you are using the True Page style set, the order of zones determines the order in which text will be placed on a recognized page.

▼ **To add an area to a zone:**



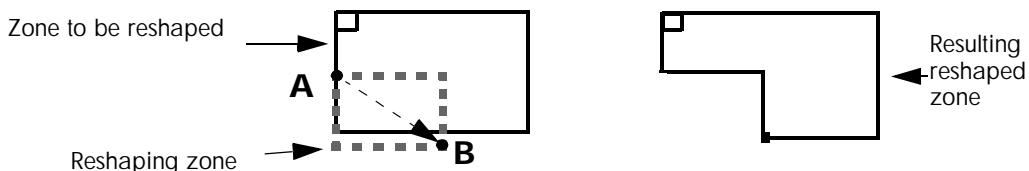
1. Click the Modify Zones tool in the Tools palette.
2. Position the mouse pointer inside the existing zone at one corner of the area you want to add to the zone. (Point A in the example below).
3. Hold down the mouse button and drag the mouse pointer to the opposite corner of the area you want to add. (Point B in the example).
4. Release the mouse button.

The reshaping zone you have defined (shown with a dotted line in the example) does not appear, but the existing zone takes on its new shape.



▼ **To subtract an area from a zone:**

To remove an area from a zone, use the above procedure, but hold down the Command key (⌘) as you draw the reshaping zone.

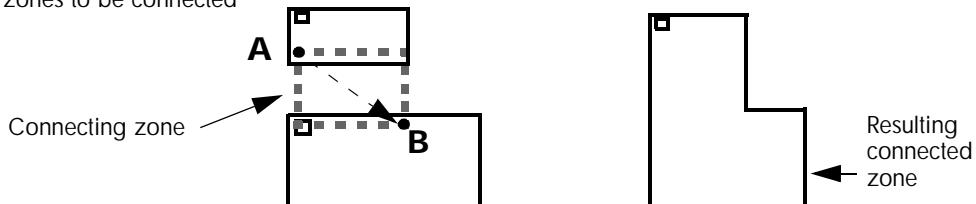


▼ **To connect two or more zones:**



1. Click the Modify Zones tool in the Tools palette.
2. Position the mouse pointer in one of the zones you want to connect.
3. Hold the mouse button down and drag the mouse pointer onto the zone(s) you want to connect. Enclose the whole area you want included in the new connected zone.
4. Release the mouse button when you are done.
The zone borders change to display the new connected zone.

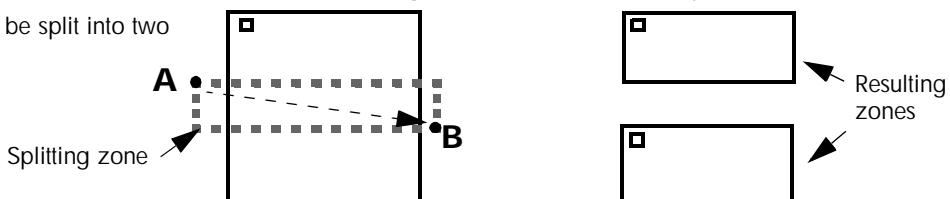
Two zones to be connected



▼ **To divide a zone:**

1. Click the Modify Zones tool in the Tools palette.
2. Position the mouse pointer at the point where you want to divide the zone.
3. Hold down the Command key (⌘) and the mouse button while dragging the mouse pointer over the area where you want the separation to occur.
4. Release the mouse button when you have completely cut through the zone. The original zone is replaced by two zones.

Zone to be split into two



▼ **To delete zones:**



1. Click the Draw/Select Zones tool in the Tools palette if it is not already selected.

2. Select the zone you want to delete by clicking it. Handles appear on the selected zone.
 - Shift-click to select additional zones.
 - Double-click the Draw/Select Zones tool or choose *Select All* in the Edit menu to select all zones on the current page.
3. Press the Delete key or choose *Clear* in the Edit menu.
The selected zones disappear, but the page image itself remains. If you do manual zoning and select *Use Only Current Zones*, any part of an image not enclosed by a zone is ignored during OCR.

Table zones

Table zones must be rectangular. During auto-zoning, the program automatically places row and column dividers. The table tools in the Zone Info palette become active if the current page contains at least one table zone. Use the tools to modify dividers in table zones:



Insert rows: Click this, then move the mouse pointer into a table zone. It will appear \triangleright . Each click inserts a horizontal row divider.



Insert columns: Click this, then move the mouse pointer into a table zone. It will appear \triangleleft . Each click inserts a vertical column divider. Press Control and click to insert a divider only in the current row.



Move dividers: Click this, then move the mouse pointer into a table zone. When it reaches a divider it appears as \div or $\cdot\div$. Click and drag the pointer to move the selected divider. You cannot drag a divider beyond its neighbor. Avoid placing dividers very close together and do not let them cut through texts.



Remove dividers: Click this, then move the mouse pointer into a table zone. When it reaches a divider it appears as \times or $\cdot\times$. Click to delete the indicated horizontal or vertical divider.



Remove/Replace All: Click this, then move the mouse pointer into a table zone. It appears as \times . Click to remove all dividers in the table. The mouse pointer becomes $\#$. Click again to have dividers automatically redetected in the table zone.

Performing recognition

Performing recognition involves analyzing character shapes found in an image and generating editable text from them. This is also referred to as *performing OCR*. After OCR, you can proofread for recognition errors and misspelled words before you export the text to another application.

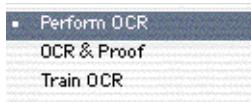
This section describes the following procedures:

- ◆ Performing OCR
- ◆ Proofreading OCR results
- ◆ Verifying recognized text
- ◆ Color markers
- ◆ Getting page information

Performing OCR

Before performing OCR, make sure the current zones and settings are appropriate for your document. For example, to transfer the contents of graphic zones to have them embedded in the recognition results, you must select *Retain Graphics* in the OCR panel of the Preferences dialog box. See *OCR settings* on page 82.

▼ To perform OCR on a single current page:



1. Select *Perform OCR* or *OCR & Proof* in the OCR button's pop-up menu. *OCR & Proof* prompts you to check for errors after OCR. *Train OCR* is not available in OmniPage SE for Mac.
2. Click the *OCR* button.

The page is recognized according to the current zones and settings. If there are no zones on the page, zones are created automatically or with a currently selected zone template. Recognition results appear in Text view.

To recognize more than one page at a time, you must use automatic processing (see page 33).

Proofreading OCR results

Recognized text appears in Text view after OCR so you can check for errors and misspellings in the text before exporting it.

Error checking (proofing) starts automatically after OCR if you chose *OCR & Proof* as the OCR option. It starts from the first recognized page and continues through all recognized pages in the document. If you chose *Perform OCR* you must start proofing by choosing *Proofread OCR...* in the Edit menu as described below. Then, proofing starts from the current cursor position.

You can select main and secondary recognition languages, a user dictionary and whether to use a Language Analyst or not in the Spelling panel of the Preferences dialog box. See *Spelling settings* on page 84 for more information. See also *User dictionaries* on page 103.

▼ To check and correct errors in recognized text:

1. Choose *Proofread OCR...* in the Edit menu.

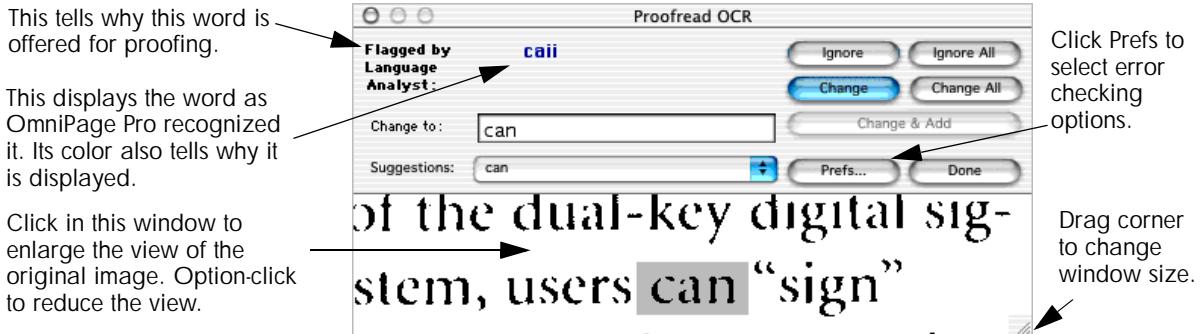
Proofing stops on words containing an unrecognizable character and displays them **red**. An unrecognizable character is replaced by a red reject character; a tilde (~) by default.

If a Language Analyst is enabled, proofing will also stop on:

- Words containing one or more characters recognized with a lower degree of certainty (words displayed **green**)
- Words flagged by the Language Analyst, for instance for not being found in a main or user dictionary (displayed in **blue**)

You can choose whether or not to stop on acronyms, abbreviations and proper names in the Spelling panel of the Preferences dialog box.

When OmniPage Pro stops on a word, it highlights the word in Text view. These words will also have color markers if *Show Markers* is enabled in the Edit menu. The Proofread OCR dialog box shows the original image of the word (also highlighted) in its context on the original page.



2. Select one of these options for the word:

- Click *Ignore* to allow the word to remain as recognized.
- Click *Ignore All* to skip all instances of the word as recognized, during the current proofing session. (The word will not be skipped if it contains a suspect character).
- Click *Change* to replace the recognized word with the word in the *Change to* edit box. Either type a word into the edit box or click to open the *Suggestions* pop-up menu and select a word.
- Click *Change All* to replace all instances of the word with the word in the *Change to* edit box.
- Click *Change & Add* to replace the word with the word in the *Change to* edit box and to add this word to the current user dictionary. You cannot add a word with a reject symbol.

After you select an option for the word, OmniPage Pro finds the next doubted word. As you proof each word, its colored marking is removed.

3. To interrupt proofing, click in Text view. Then you can make editing changes, verify text, modify settings and even jump to other pages. The proofreader button *Ignore* becomes *Start*. Click this to restart proofing. If you remained on the same page, proofing restarts from the point where it was interrupted. If you have jumped to another page, it starts from the top of that page.

4. Click *Done* or close the Proofread OCR dialog box to save all changes and exit proofing before the end of the document is

reached. The program informs you when the end of the document has been reached; all your changes are saved automatically.

Note OmniPage Pro can only perform a spelling check on words that it has recognized. It cannot check words that you have manually typed into Text view.

Tip To delete unneeded characters (for instance generated by 'noise' on the image), clear the edit box and click *Change*. If the program mistakenly splits a word into two, maybe at the end of a line, type in the whole correct word when the first part of the word is displayed, then empty the edit box when the second part appears.

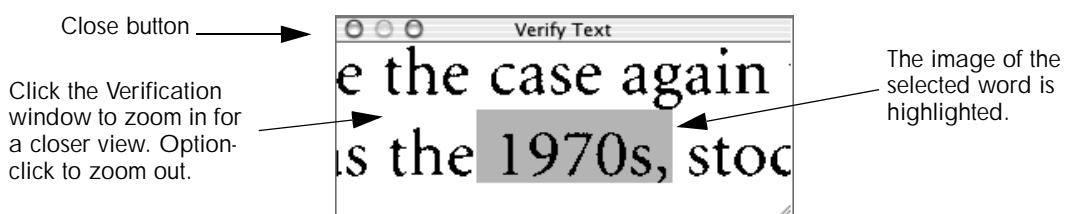
Verifying recognized text

You can compare recognized text against its original image to make sure that text was recognized correctly.

▼ **To verify text against its original image:**

1. Make sure Text view is active.
2. Hold down the Option key and double-click the word you want to verify. Or, select the word and choose *Verify Text* in the Edit menu, or press **⌘Y**.

The Verification window opens and shows a clear close-up of the original word and its surrounding area in the image.



You can type in a new word to replace the selected recognized word.

3. Click the standard *Close* button to close the Verification window.

Color markers

Words to be stopped on during proofing may appear in color (red, green or blue) in Text view and in the Proofread OCR dialog box.

To temporarily hide color markers in recognized text, make Text view active and choose *Hide Markers* in the Edit menu. The coloring is removed from all marked words in the current document, and no marking is placed on new pages or documents. To show markers again, choose *Show Markers* in the Edit menu. Proofing will still stop on all suspected words and display them in the appropriate color, even when markers are hidden in Text view.

Proofing always stops on red words. If *Use Language Analyst* was enabled in the Spelling panel of the Preferences dialog box at recognition time, proofing will also stop on the green and blue words and these will be available for marking in Text view.

Changing the *Use Language Analyst* setting has no effect on text which has already been recognized.

Color markers are not retained when you export a document to another application.

Getting page information

After OCR, you can choose *Show Page Info* in the File menu (or press **⌘I**) to get the following information for the current page:

- ◆ Source of the OCR, whether a scan performed by OmniPage Pro or a file that you have loaded (with the file name and folder).
- ◆ Resolution of the scanned image, in dpi (dots per inch).
- ◆ Image Size, in pixels and inches or centimeters.
- ◆ Color depth and resolution for color images.
- ◆ Number of words and characters on the page (including spaces).
- ◆ Recognition time in minutes and seconds. This excludes time for scanning, drawing manual zones and writing data to disk.
- ◆ Number of reject and suspect characters.
- ◆ Recognition rate in characters per second and words per minute.

Working with documents

The Thumbnail window gives an overview of all pages in the document and allows you to perform page-level operations. The Document window allows you to work with each page one after the other. This section describes the following procedures:

- ◆ Resizing a page display
- ◆ Saving a document as you work
- ◆ Moving to other pages
- ◆ Reordering pages
- ◆ Deleting a page
- ◆ Undoing edits
- ◆ Modifying images
- ◆ Modifying text
- ◆ Printing a document
- ◆ Listening to a document
- ◆ Closing a document
- ◆ Quitting OmniPage Pro

Resizing a page display

You can enlarge (zoom in) or reduce (zoom out) the view of a page displayed in Image view or Text view.

▼ To resize a page display:

1. Click the view that you want to resize (Text or Image) to make that the active view.
2. Click the box that displays the zoom percentage located in the Info line, along the bottom of the Document window. Select the desired zoom setting in the pop-up menu.



In Image view you can also click the Zoom tool in the Tools palette and then click the area of the image you want to enlarge. Option-click to reduce the view.

Saving a document as you work

If you are working with a long or important document, or want to reopen the document in OmniPage Pro in a future session, you should save it as an OmniPage Document soon after beginning your work.

To save the document to disk for the first time, choose *Save* or *Save As...* in the File menu. The Save As OmniPage Document dialog box appears, allowing you to choose a location and specify the file name. The recommended extension for an OmniPage Document is .opd.

If the file has already been saved as an OmniPage Document, click *Save* to have the file updated. The updating includes changes to page images, zoning, recognition results and settings. Choose *Save As...* to save the latest state of the OmniPage Document under a different name, leaving its state from the previous save under its existing name.

You can also protect work by clicking the *Export* button and saving recognition results to file. If your further work on the document is successful, you can export it again, overwriting the older file.

Your OmniPage Documents can be passed between OmniPage Pro X and OmniPage SE for Mac. In SE training cannot be done.



Moving to other pages

You can move to a different page in a document in the following ways.

- ◆ Click the thumbnail of the page you want to display.
- ◆ Click the forward or backward arrow buttons next to the current page number located bottom left of the Document window.
- ◆ Choose *Go to Page...* in the Edit menu or double-click the current page number to open the Go to Page dialog box. Select *First Page* or *Last Page* or enter a specific number in the *Page* edit box.

Reordering pages

You can reorder pages in a document by dragging their thumbnails to different positions in Thumbnail view. Drag-and-drop pages one after the other.

Deleting a page

You can delete a page from a document that has at least two pages. For example, you may want to delete a page that was poorly scanned.

To delete the current page, choose *Delete Current Page* in the Edit menu. Or, click the thumbnail of the page you want to delete and drag it to *Trash*. Everything is discarded: the thumbnail, page image, and recognition results. Pages are renumbered automatically.

Undoing edits

Choose *Undo* in the Edit menu immediately to reverse an action that produces an unwanted result in Image view or Text view. After you choose *Undo*, it changes to *Redo*. If an action cannot be reversed, the command appears as *Can't Undo*.

Modifying images

You can modify an image when Image view is active. Drag the splitter at the base of the Document window to the right if Image view is not big enough or not visible at all.

Rotating an image

You can rotate a page image when Image view is active. For example, if a page is accidentally scanned upside down, you do not have to scan it again. You can correct the orientation by rotating it. Click the Rotate tools in the Tools palette to turn the entire page 90 degrees left, 180 degrees, or 90 degrees right. If possible, rotate a page *before* you create zones. All zones are deleted during page rotation.

Note You can also specify that images coming from scanner should be flipped around their vertical or horizontal axes. These types of rotation cannot be performed on loaded images; they must be specified in the Scanner panel of the Preferences dialog box before scanning is started.

Erasing areas of an image

You can erase areas of the actual image using the Erase Image tool in the Tools palette. This is useful if you want to get rid of smudges, signatures, or other types of “noise” on the page before OCR.



1. Use the Zoom tool in the Tools palette to enlarge the area of the image you want to erase.



2. Click the Erase Image tool in the Tools palette. The mouse pointer turns into a square box.
3. Click the box over the image area that you want to erase. A piece of the image disappears with each mouse click. You can also hold the mouse button down and drag the mouse pointer over the area you want to erase.

Note

If you do not want to permanently erase parts of the actual image, but want to omit areas of a page from OCR, identify the areas as *Ignore* zone types prior to auto-zoning, or do not include them in zones when you do manual zoning.

Modifying text

You can modify recognized text in Text view before exporting it to another application. Click in Text view to make it active. Move the splitter at the base of the Document window to the left to give more space to Text view. If you drag it far to the left, Image view disappears completely. Select a suitable magnification for Text view. See also *Proofreading OCR results* on page 53.

Selecting all text

To apply formatting, such as a particular font, to all text on a page, you can select the entire page by choosing *Select All* in the Edit menu (or **⌘A**). The entire contents of a recognized page is selected when Text view is active with any style set except True Page. With True Page, only the text within the selected frame is selected. To remove a selection, click anywhere within it.

Selecting a block of text

Click at the start of the desired text and drag the cursor to the desired end point. Release the mouse button. The selected text is highlighted. With the True Page style set, a selection cannot extend beyond a single frame.

Formatting text

Use commands in the Format menu to apply font, font style, and font size formatting to selected text in your recognized document.

Cutting or copying text and graphics

Choose *Cut* in the Edit menu to place selected text or a selected graphic on the Clipboard. Cut items are removed from Text view. Choose *Copy* in the Edit menu to place a copy of selected text or graphics on the Clipboard. Copied items are not removed.

You cannot cut or copy text and graphics at the same time. If both are selected, only the text will be placed on the Clipboard.

Text on the Clipboard can be pasted back into Text view or into another application. Choose *Paste* in the Edit menu to place text at the cursor location in Text view. Graphics cannot be pasted into Text view, but can be pasted into applications that support the PICT format.

Deleting text or graphics

Choose *Clear* in the Edit menu (or press the Delete key) to permanently delete selected text or graphics from Text view.

Printing a document

You can print one or more pages of a document. You can print recognized pages if Text view is active or page images if Image view is active. If you have a color printer, you can choose to print pages in color.

▼ To select options and print pages:

1. Choose *Page Setup...* in the File menu. The options available in the Page Setup dialog box depend on your printer.
2. Select the desired options and then click *OK*.
3. Make the view (Text or Image) you want to print active.
4. Choose *Print Text...* (or *Print Images...*) in the File menu. The choices in the dialog box depend on your printer.
5. Select print options for your document. Choose to print all images or a range of pages.
6. Click *Print* to start the print job.

Listening to a document



English or Spanish text in Text view can be read aloud by the Macintosh Speech Manager software. Choose one of its voices from the Speech Menu. Also select *Speak Selection*, *Speak This Page* or *Speak Document*. The Speech Manager interface appears as the text is read. You can change the reading speed. Select *Pause* to stop the reading. OmniPage SE for Mac does not support this feature.

Closing a document

Choose *Close* in the File menu or ⌘W to close the current document in OmniPage Pro. You can also close the document by closing the Document window. If you have not exported or saved the document or if you have changed it since the last export or save, you will be prompted to save it as an OmniPage Document before closing.

Quitting OmniPage Pro

Choose *Quit* in the Application menu (Mac OS 9: File menu) or ⌘Q to close a document and exit OmniPage Pro. If the current document has not been exported or saved or is changed since the last export or save, you will be prompted to save it as an OmniPage Document before closing.

Exporting documents

You can export original images or recognition results, for use in other applications by:

- ◆ Saving an OmniPage Document
- ◆ Saving images
- ◆ Saving recognition results
- ◆ Saving to Portable Document Format (PDF)
- ◆ Copying a document to the Clipboard
- ◆ Using drag-and-drop functionality

Saving an OmniPage Document

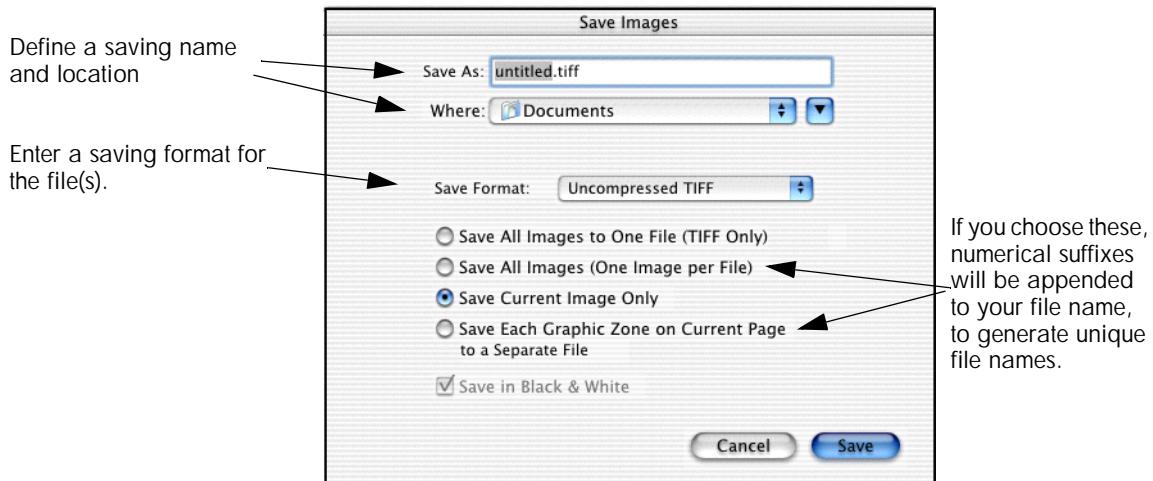
You can save your document as an OmniPage Document file if you want to reopen it in OmniPage Pro again. OmniPage Documents retain all the original images, together with their zones and their properties, some settings and any recognition results. The links between text and image are conserved, so proofing and verifying will still work in another session or at a distant location where OmniPage Pro is located.

Choose *Save* or *Save As...* in the File Menu, or export the document, choosing *OmniPage Document* as the saving format. See *Saving a document as you work* on page 58.

Saving images

You can save images from the current document to one or more image files. Images are stored in the mode they are displayed (black-and-white, grayscale, color). They are stored at their original resolutions, except for high-definition color images, which are reduced to 256 colors.

Make Image view active and choose *Save Images...* from the File menu. The Save Images dialog box appears:



For information on the supported image file formats, see page 114. PDF is not offered for saving images, because it is the recognition results that are saved to PDF, not the original images. See the following two topics.

Saving recognition results

As soon as you have at least one recognized page in a document, you can save recognition results from all the recognized pages to disk in a variety of file formats. See page 113 for information on these formats.

When you do automatic processing, the Export dialog box appears as soon as the last page is recognized or proofed (if requested). Follow the procedure below from point 2 onwards. Point 1 tells you how to start the export manually.

▼ To export recognition results from a document:



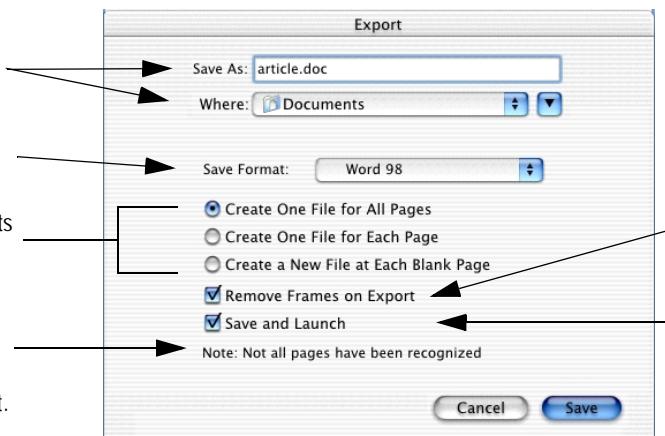
1. Click the *Export* button with *To file...* selected in the Export pop-up menu. The Export dialog box appears.
2. Select the folder where you want your file saved.
3. Type in a file name for your document, using not more than 28 characters.

Type in a name and define a location for your file.

Select a save format.

Select save options when saving to formats other than *OmniPage Document*.

This appears if there are unrecognized pages. They will be skipped during export.



This is available when True Page is set, for some saving formats. Select it to maintain page layout without frames, so text can flow between columns.

Choose this to see your recognition results in their target application immediately after export.

4. Select the appropriate file format for your document in the Save Format pop-up menu.
Formats able to accept True Page output are listed with a **Tp** icon. If your target application cannot handle frames, or you do not want frames to be used, click the check box *Remove Frames on Export*.
5. Select other save options if you are saving the document in a file format other than OmniPage Document.
6. Click *Save*.
The document is saved to disk as specified. If *Retain Graphics* was selected in the OCR panel of the Preferences dialog box, embedded graphics are saved with the file, providing the selected format supports them. The graphics are saved at 75 or 150 dpi, as specified in the Preferences dialog box.
7. If you chose *Save and Launch*, the target application linked to your saving format is activated and the recognition results are loaded. If you chose to save each page to a separate file, only the first file is loaded. OmniPage Pro remains running with the document still available.

Saving to Portable Document Format (PDF)

When saving to PDF, we recommend you choose the True Page style set, because this forms the basis for saving, whatever style set is chosen. Check that all text is visible within the frame borders. You have four choices when saving recognition results to PDF files.

Image only: The PDF file is viewable only and cannot be modified in a PDF editor and text cannot be searched.

Normal: The PDF file can be viewed and searched in a PDF viewer and edited in a PDF editor.

With Image on text: The PDF file is viewable only and cannot be modified in a PDF editor. There is a text file behind each image, so text can be searched. A found word is highlighted in the image.

With image substitutes: Words with reject and suspect characters have image overlays, so uncertain characters display as they were in the original document. The PDF file can be viewed, edited and searched.

PDF saving is not included in OmniPage SE for Mac.



Copying a document to the Clipboard

You can choose to send a copy of the recognition results from all recognized pages in the document to Clipboard. This can then be pasted into another application. You can also copy the image block from a zone in Image view to the Clipboard.

▼ To copy an entire document to Clipboard:

1. Select *To Clipboard* in the Export button's pop-up menu.
2. Click the *Start* button for automatic processing or the *Export* button to export pages manually. The results from every recognized page are copied to the Clipboard. With manual processing this happens immediately. With automatic processing it happens when the last page is recognized or proofed.
3. Paste the Clipboard contents to a target application.
Text formatting, such as bold and italics, is retained if you paste it into an application that supports RTF information. Otherwise,



only plain text is pasted. Graphics are retained if you selected *Retain Graphics* and the target application supports them. The graphics have the resolution chosen in the OCR panel of the Preferences dialog box.

▼ **To copy the image from a zone to Clipboard:**

1. Make Image view active.
2. Click the Draw>Select Zones tool in the Tool palette.
3. Select the zone you want to copy by clicking it.
4. Choose *Copy* in the Edit menu. A copy of the image from the zone area is placed on the Clipboard. It can be pasted into any target application capable of handling PICT images. It retains its original resolution and color depth value (up to 256 colors).

Note Copying through Clipboard (and Direct OCR) work best for processing just a few pages, especially under Mac OS 9 if an application's partition is almost full. Save larger documents to a file format compatible with your application.

Using drag-and-drop functionality

Drag-and-drop can be used for import (see page 40) and export.

Dragging a thumbnail for whole page export

You can drag a thumbnail from Thumbnail view to the Desktop, to a folder or to another application that supports drag-and-drop functionality. The image of the thumbnail's page is placed as a PICT image with the same resolution and mode (black-and-white, grayscale or color) as the original image. If it is dragged to the Desktop or a folder, it is named *Picture clipping*, with a numerical suffix if necessary.

Dragging a zone from Image view

You can drag a single selected zone from Image view to the same locations. A copy of the zone contents is placed as a PICT image, with the same behavior as for a whole page.

Dragging from Text view

You can drag a block of selected recognized text from Text view to the Desktop or another application that supports drag-and-drop functionality. Text formatting will be transferred if possible. The result appears on the Desktop as a picture clipping icon, and double-clicking on it allows you to view the text only. But if you drag the icon into a text editing application, it is inserted as editable text. An embedded graphic can be exported by drag-and-drop from Text view. However, you cannot drag-and-drop text and graphics together.

Direct OCR



The Direct OCR™ feature allows you to activate OmniPage Pro from the Dock (Mac OS 9: Apple menu), perform OCR on one or more images, and have the recognized text placed at the insertion point in a target application.

Direct OCR works with virtually any Macintosh application that supports pasting text from the Clipboard. Your Macintosh must have enough memory to run both OmniPage Pro and the application.

OmniPage Pro does not have to be running when you start Direct OCR. If it is running with no document, it will remain open afterwards. If it is running with a document open, you will be prompted to close it first. Before starting Direct OCR, be sure the Clipboard does not contain something you still want to paste.

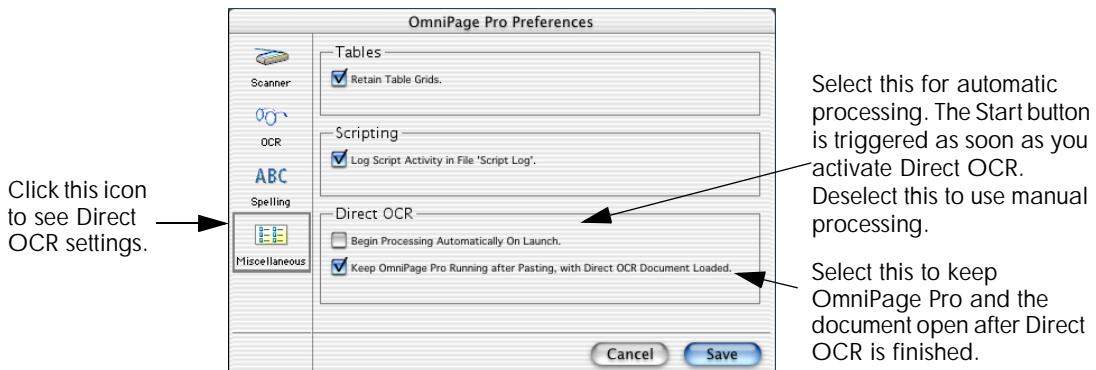
Text formatting, such as bold and italics, is retained if you are pasting into an application that supports RTF information. Otherwise, only plain text will be pasted. Graphics are transferred if *Retain Graphics* was selected and the target application supports them.

Note If the Direct OCR icon does not appear automatically in the Dock, you should drag the icon from the OmniPage Pro: OmniPage Extras folder and drop it into the Dock.

Using Direct OCR

You can run Direct OCR using automatic or manual processing. For automatic processing, all settings should be selected suitably in OmniPage Pro before using Direct OCR. If you are uncertain whether settings are suitable or not, or if you want to exclude parts of the pages, use manual processing instead. This allows you to check and change settings and also do manual zoning.

Choose Direct OCR settings (including the choice of automatic or manual processing) in the Miscellaneous panel on the Preferences dialog box before you use Direct OCR.



▼ To use Direct OCR with automatic processing:

1. Align a page in your scanner or a stack of pages in its automatic document feeder (ADF) if you plan to scan. Be sure *Scan Until Empty* is enabled if you want to scan multiple pages from the ADF.
2. Open or switch to the application and place the insertion point where you want recognized text to be placed. You do not need to open OmniPage Pro itself.
3. Click the *OmniPage Direct OCR* icon on the Dock. OmniPage Pro opens in Direct OCR mode. Either scanning starts or the Load Images dialog box appears so you can select image files.
4. Pages are processed automatically. This includes auto-zoning, unless you apply a template and choose *Use Only Current Zones*. The *Export* button displays *To Application*, blocking other export

until the Direct OCR operation is finished. Proofing starts as soon as the last page is recognized, if *OCR & Proof* was selected.

5. When recognition or proofing is finished, the recognition results appear at the insertion point in the target application.

▼ **To use Direct OCR with manual processing:**

1. Follow points 1 to 3 as for automatic processing.
2. The OCR Toolbar appears. Scanning starts or the Load Images dialog box lets you name image files.
3. Do manual zoning on the resulting page images if you wish. Modify settings as necessary.
4. Select an OCR method and click the *OCR* button for each page, or click the *Start* button and then choose *Recognize All Unrecognized Pages*.
5. Proof each page if you asked it to start automatically. Verify and edit text as desired. Start proofing manually if you wish.
6. The *Export* button displays *To Application*. If you clicked *Start*, export follows automatically. If not, click the *Export* button. All recognized pages are placed at the insertion point in the target application.

▼ **What happens after Direct OCR**

If you selected *Keep OmniPage Pro Running after Pasting, with Direct OCR Document Loaded* in the Miscellaneous panel of the Preferences dialog box, OmniPage Pro remains open with the images and recognition results, allowing you to verify, edit and save the document to file.

If you deselected this option, the recognition results are available only in the target application and on the Clipboard. If OmniPage Pro was closed when you started Direct OCR, it will be closed down. If it was open when you started Direct OCR, it will remain open, without a document.

Chapter 4



Settings

This chapter provides more detailed information on the options available in the pop-up menus on the OCR Toolbar and settings you can select in the Preferences dialog box.

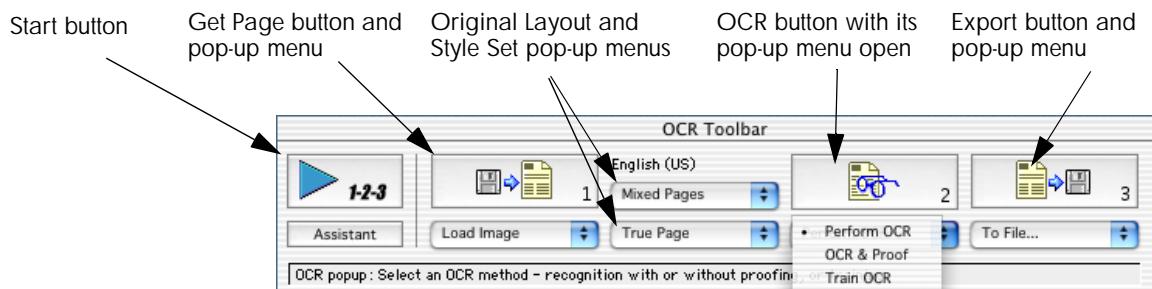
Make sure that settings are appropriate for your document *before* you start processing it. You may have to experiment with different settings to get the results you want.

Please continue reading this chapter for information on these topics:

- ◆ OCR Toolbar options
 - ◆ Get Page options
 - ◆ Original Layout options
 - ◆ Style Set options
 - ◆ OCR options
 - ◆ Export options
- ◆ Preference settings
 - ◆ Scanner settings
 - ◆ OCR settings
 - ◆ Spelling settings
 - ◆ Miscellaneous settings

OCR Toolbar options

The three numbered OCR Toolbar buttons allow you to take a document through each step of the OCR process. The *Start* button begins automatic processing. You can select options in the five pop-up menus as described below.



Pictures on the three buttons change as you select different options, to indicate what will happen when the button is clicked or when automatic processing is run. The pictures on the left show the button's appearance when each option is selected.

Get Page options

You can select from the following options in the Get Page pop-up menu. The selection is activated at the start of automatic processing (images are acquired and recognized) or by clicking the *Get Page* button (images are acquired without recognition).

Scan in B&W



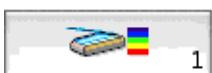
Select *this* to scan paper documents from your scanner with black-and-white scanning. Choose this if you wish to retain diagrams or line-art in your output document. For best OCR accuracy, choose this for good quality pages with crisp black text on a white background.

Scan in Gray



Select *this* to scan paper documents from your scanner with grayscale scanning. Choose this if you wish to retain pictures or photos in your output document. For best OCR accuracy, choose this for lower quality pages, for example with low or varying contrast, or with text on shaded or colored backgrounds.

Scan in Color



Select *this* to scan paper documents from your scanner in color. Choose this only if you wish to retain color graphics in your recognized document. Handling color documents needs extra memory and time. It yields no accuracy benefits for OCR compared to grayscale scanning (at a given resolution). It is available only when a color scanner is installed.

Note The scanner options in the Get Page pop-up menu may vary depending on your scanner configuration. Scanning modes not supported by your scanner will be grayed. If you see only one item *Scan Image*, you should select the scanning mode (black-and-white, grayscale or color) on the scanner interface.

Load Image

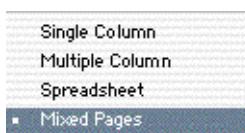


Select *Load Image* to load one or more existing image files. Multi-page image files (TIFF and PDF formats in OmniPage Pro and TIFF in OmniPage SE for Mac) can be handled; you can specify which page images to open. You cannot modify the brightness, contrast, resolution or mode (black-and-white, gray or color) of image files when you load them. They are opened as they were saved. Images are automatically straightened, if necessary.

For step-by-step guidance on scanning, see *Scanning pages* on page 38. For similar guidance on opening images, see *Loading image files* on page 38 and *Supported image file formats* on page 113 and 114.

Original Layout options

You can select from the following options in the Original Layout pop-up menu. These let you describe the incoming pages, to assist the program in auto-zoning. Auto-zoning always runs when you perform automatic processing (unless you load a zone template), and sometimes runs during manual processing.



Single Column

Select *this* to have OmniPage Pro automatically draw and order zones on single-column page images, such as letters, memos or book pages. Select it to deter the program from searching for columns.

Multiple Column

Select *this* to have OmniPage Pro automatically draw and order zones on multiple column page images such as from magazines or newspapers. The program will try to find columns.

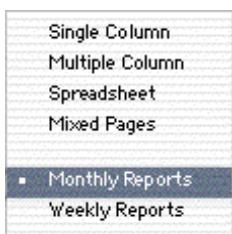
Spreadsheet

Select *this* for pages containing spreadsheets or where you want the whole contents of the page treated as a table. Do not select it for pages containing tables along with text or other non-table elements. Use the Miscellaneous panel of the Preferences dialog box to determine whether the table data will be placed in a grid or in tab-separated columns.

Mixed Pages

Select *this* for complex pages or if you are unsure. Select it also for a multiple-page document with a variety of page layouts. This gives OmniPage Pro full control in drawing and ordering zones on each page.

For more information, see *Creating zones automatically* on page 42.



[Zone Templates]

Select the name of a zone template file that you want to use to place zones on new incoming pages. Any zone templates you have created appear at the bottom of the pop-up menu. The example comes from a user who has created two templates to process standardized form-like printed reports – one type arrives each week, the other each month.

To place template zones on an existing page, select the template here, then click the **Apply Template** tool in the Tools palette. For more information, see *Zone templates* on page 98.

OmniPage SE for Mac does not support this feature.

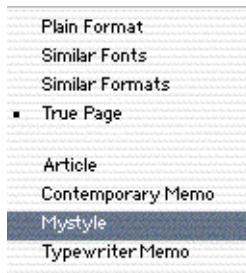
Style Set options

You can select a page-level style set option from the Style Set pop-up menu. The choice made here determines the appearance or formatting level to be applied to the recognition results coming from new incoming pages.

The selected OCR Toolbar option has no influence on existing pages, even if you re-recognize them. Use the Zone Info palette to change the style set for an existing page.

Tables and graphics can be handled by all style sets. With True Page, these are retained at their original location on the page. With all other style sets, tables are placed at their location in the decolumnized text and graphics are placed at the end of the text from the page.

The first four style sets define basic formatting levels. The remaining style sets are fully editable. Choose from the following options:



Plain Format

Select this to have plain text in one font and size. Text is left aligned, decolumnized and wrapped to the whole page width.

Similar Fonts

Select this to have text with font formatting retained. Fonts are mapped. Font sizes and bold, italic and underlined texts are detected and maintained. Text is left aligned, decolumnized and wrapped.

Similar Formats

Select this to have results similar to *Similar Fonts*, but with column widths maintained when multi-column pages are decolumnized.

True Page

Select this to have the original page layout maintained as closely as possible. Text blocks, headings, tables, graphics and other elements are placed in frames. This is recommended when exporting to PDF format (see page 66). It is suitable only for saving formats marked **Tp** in the Export dialog box.

Article

This is an editable sample style set. Select it to have the *Similar Formats* layout, but with additional zone styles. You can change the properties of these zone styles and add new styles.

Contemporary Memo

This is an editable sample style set. Select it to have the *Similar Formats* layout, but with additional editable zone styles. Use this for memos or similar documents you want exported with proportionally spaced fonts.

Typewriter Memo

This is an editable sample style set. Select it to have the *Similar Formats* layout, but with additional editable zone styles. Use this for memos or similar documents you want exported with monospaced fonts, so they appear to be typewritten.

[Custom styles]

If you have created your own style sets, these appear in the alphabetical order of the lower part of this pop-up menu. Choose a custom style to impose your own formatting wishes on incoming pages. See *Creating style sets* on page 92.

OCR options

You can select the following OCR options in the OCR pop-up menu. The selected option is activated during manual processing by clicking the *OCR* button. This performs recognition or training on the current page only. The option is also activated during automatic processing, in which case it may be applied to a series of pages.

Perform OCR



Select *Perform OCR* to recognize text on pages. During OCR, OmniPage Pro analyzes the image and interprets character shapes to produce editable text. It may also transfer image areas from graphics zones into the recognition results. Proofing will not start automatically.

For more information, see *Performing OCR* on page 52.

OCR & Proof



Select *OCR & Proof* to recognize text and then automatically start the OCR Proofreader, allowing you to check for errors.

For more information, see *Proofreading OCR results* on page 53.

Train OCR



Select *Train OCR* to teach OmniPage Pro how to recognize special or stylized characters taken from the current page. Automatic processing is not available when this option is selected. See *Training OCR* on page 99. OmniPage SE for Mac does not support *Train OCR*.

Export options

You can select from two of the following export options in the Export pop-up menu. Your choice is activated at the end of automatic processing or whenever you click the *Export* button.

To File



Select *this* to save your recognition results to a document you will name in a specified file format.

For more information, see *Saving a document as you work* on page 58, *Exporting documents* (page 63) and *Supported file types* in online Help.

To Clipboard



Select *To Clipboard* to place a copy of a document's recognition results (text and embedded graphics) on the Clipboard.

See *Copying a document to the Clipboard* on page 66.

To Application



This option cannot be selected. It appears when Direct OCR is in use. Other export options are not available at that time. When the Direct OCR recognition (and optionally proofing) is finished, the recognition results are sent through the Clipboard to the cursor position in the target application. See *Direct OCR* on page 68.

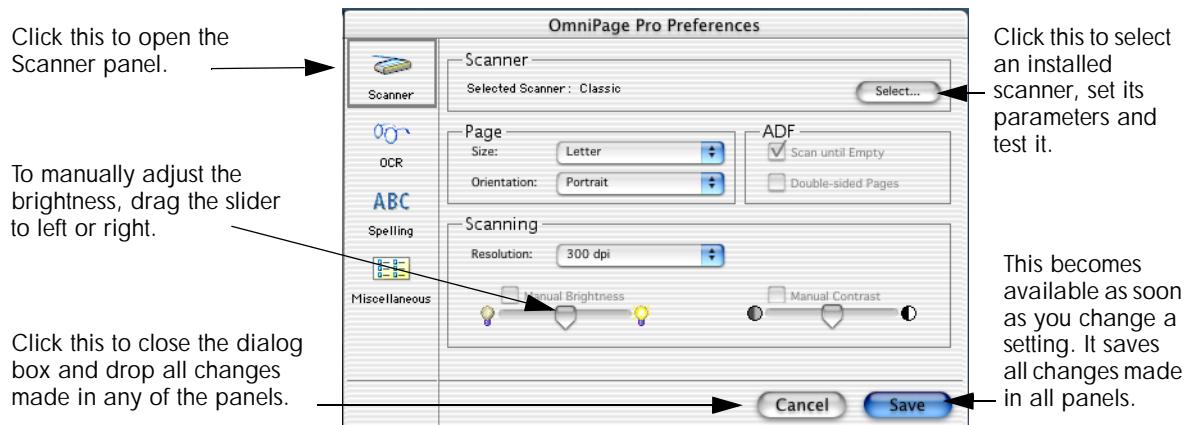
Preference settings

The Preferences dialog box is the central location of OmniPage Pro settings. To open it, click *Preferences...* in the Application menu (Mac OS 9: Edit menu). The dialog box has four panels. Each panel can be displayed by clicking its icon on the left. When the dialog box is reopened, it displays the last selected panel.

See the online Help topic *Settings Guidelines* for recommendations in choosing settings and options for various types of documents and tasks.

Scanner settings

Click the *Scanner* icon on the left of the Preferences dialog box to display this panel. It allows you to select a scanner and the settings that control the way it will scan pages.



Scanner

This displays the currently selected scanner. Click *Select...* to select a different scanner. Only scanners already installed on your system can be selected. For guidance on selecting or changing scanners and drivers, see chapter 1. The controls offered in this Scanner panel depend on the facilities supported by your scanner.

Page Size

Select the dimensions of the pages you plan to scan in the *Size* pop-up menu.

- Select *Letter* for 8.5 by 11 inch pages.
- Select *A4* for 21 by 29.7 cm pages (8.27 x 11.7 inches).
- Select *Legal* for 8.5 by 14 inch pages.

Page Orientation

Select the orientation of the pages you plan to scan in the *Orientation* pop-up menu. Be sure to also load pages correctly in your scanner.

- Select *Portrait* for vertically-oriented pages (the shorter page edge is parallel to the scanning head).
- Select *Landscape* for horizontally-oriented pages (the longer page edge is parallel to the scanning head).
- Select *Flipped* to have portrait images rotated by 180 degrees.

- Select *Flipscape* to have landscape images rotated by 180 degrees.

Tip

Flipped and *Flipscape* options are useful if you are scanning pages in a book and have trouble positioning the book correctly in the scanner. You can also rotate a page image after it is loaded into OmniPage Pro. For more information, see *Rotating an image* on page 59.

ADF settings

If you use a scanner with an *automatic document feeder* (ADF), you can use the following settings.

- Select *Scan until Empty* to scan every page in your scanner's ADF.

This setting is useful when you want to scan a stack of pages at once. If it is not selected, OmniPage Pro only scans the first page in your ADF and you must click the *Get Page* or *Start* button to scan each subsequent page.
- Select *Double-sided Pages* to scan pages that have text printed on both sides.

OmniPage Pro scans pages and then prompts you to turn them over so it can scan the reverse sides. If you have a stack of double-sided pages, also select *Scan Until Empty*. After scanning, page images are displayed in Image view in the correct order. If you have a duplex scanner, do not set this; the scanner's own software can handle the double-sided scanning.

Scanning Resolution

Use this to select a scanning resolution in dots per inch (dpi). The values offered are scanner dependent. For non-color scanning they may range from 200 to 600 dpi, and from 200 to 300 for color scanning. In general, 300 dpi is best for OCR accuracy. 400 dpi may be better for very small print. Higher resolutions may be desirable for saving higher-quality images to file or to OmniPage Documents, at the expense of increased file size, processing time and maybe OCR accuracy.

Brightness

The brightness setting for scanning a page works like that on a photocopier. This setting can compensate for variations in paper and print quality, so it can have a big influence on OCR accuracy.

Click the *Manual Brightness* check box and move the slider to lighten or darken the brightness for your scanning.

The following illustrates optimum and unsuitable brightness.



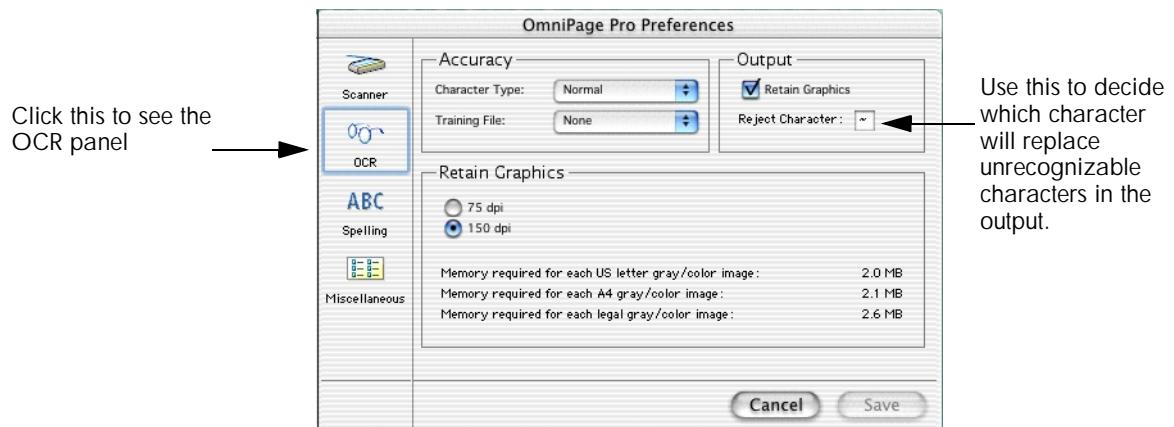
Contrast

The contrast setting for scanning a page works like that on a television set. This setting is only activated if you have Grayscale or Color selected in the Scanner settings. It lets you increase or decrease the difference between light and dark areas on the image. Click the *Manual Contrast* check box and move the slider to make a contrast setting.

Note Some scanners offer only automatic detection for brightness and contrast. Some require a manual setting. Others offer both methods. In this case, automatic detection may be better; some scanners do this dynamically, varying the setting for different parts of the page. If results are disappointing, try using manual adjustment.

OCR settings

Click the *OCR* icon in the Preferences dialog box to select accuracy and output options.



Character Type

Select a setting to characterize the printed text on your pages in the *Character Type* pop-up menu.

- Select *Normal* for conventionally printed text characters. Select it also for dot-matrix texts printed in fine mode or with 24-pins. Select it also for fax files, but ask your senders to use Fine Mode.
- Select *Dot Matrix* for text characters printed in draft mode with a 9-pin, monospaced dot-matrix printer.

Training File

A *training file* is a set of up to 256 pre-recognized character shapes linked to OCR solutions, that OmniPage Pro can use to compare with shapes it is trying to recognize. For most recognition tasks, a training file is not necessary. If you have a training file you wish to use, select it in the *Training File* pop-up menu. *None* is the only option if you have not created any training files.

OmniPage SE for Mac does not support training files.



Training files are useful for recognizing characters that prove difficult to recognize or are being regularly misrecognized. To create a training file, see *Training OCR* on page 99.

Retain Graphics switch

Select *Retain Graphics* if you want OmniPage Pro to retain original graphics, such as photographs or drawings, in the recognized document. They will be displayed in Text view and exported to file, provided the selected file format supports graphics. Graphics can be exported by drag-and-drop, copying to Clipboard and Direct OCR.

Make sure that all the pictures you want retained are correctly enclosed in zones with the zone type *Graphic*. These have black borders and display a graphic icon. See *Specifying zone types* on page 43.

If you deselect this, the contents of graphics zones are ignored. Pictures will neither appear in Text view nor be available for export.

In the lower part of the panel you specify the resolution for graphics exported in grayscale or color. Exported graphics appear as they do in Text view (black-and-white, grayscale or color).

Reject Character

Words containing unrecognizable characters appear in red in the Proofread OCR dialog box and optionally in Text view. Unrecognized characters are replaced by a red reject character. The default character is a tilde (~). Type the character you want to use in the *Reject Character* edit box.

For example, if OmniPage Pro could not recognize the J in REJECT, and the tilde (~) was the reject character, the string **RE~ECT** would appear in your recognized document.

Retain Graphics settings

Choose a resolution setting (75 or 150 dpi) to be used for the export of grayscale or color image areas embedded in Text view. The settings are applied when you save recognition results from the whole document to file, send them to Clipboard or use Direct OCR.

The settings have no effect on recognition accuracy, nor on the display of the embedded images in Text view. They are not used when saving to OmniPage Documents, nor when saving page images, nor when exporting single graphics zones or areas by drag-and-drop or through the Clipboard.

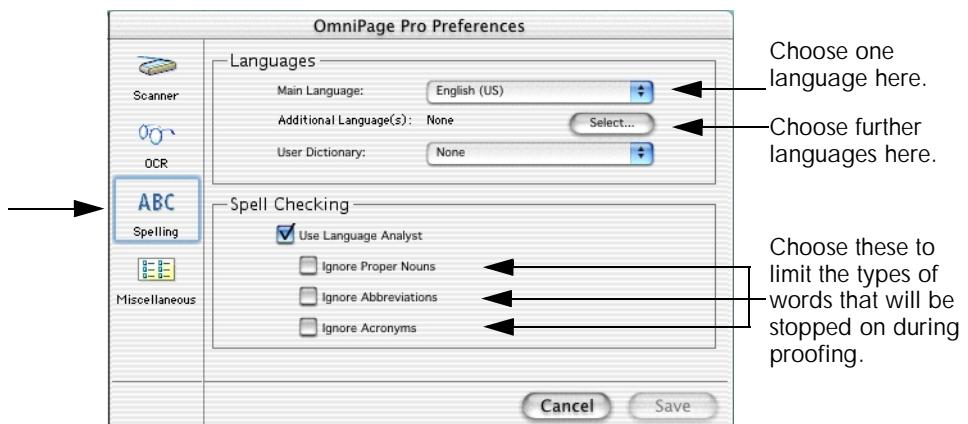
The 150 dpi setting yields higher quality pictures, but consumes more disk space when the file is saved. You can use the 75 dpi setting to save disk space, with a corresponding loss of image quality.

The memory requirements for a typical exported page of a given size, stored at the selected resolution are displayed below the options. This is for a typical page with about 70% text and 30% embedded image.

Spelling settings

Click the *Spelling* icon on the left of the Preferences dialog box to select recognition languages, user dictionaries and spell checking settings. These settings are used by the Language Analyst during OCR and for proofreading after OCR.

Click this to see the Spelling panel



Main Language

The Main Language pop-up menu enables you to choose the main language for the page(s) you intend to recognize. Your choice determines which characters are validated for recognition and which main dictionary will be used.

The languages available are Danish, Dutch, English (UK and US), Finnish, French, German, Italian, Norwegian, Portuguese (Standard and Brazilian), Spanish and Swedish.

Additional Language(s)

In addition to the Main Language for recognition, you may select one or more secondary languages. Specifying additional languages broadens the range of accented letters validated for recognition. It also enables more than one dictionary. Then the program monitors text as it is recognized to determine its language and which dictionary to apply. This lengthens the processing time, so you should only activate additional languages if your pages really contain more than one language.

The Main Recognition Language is displayed on the OCR Toolbar. It is followed by three dots if any additional languages are selected.

▼ To select secondary languages and dictionaries:

1. Click the *Select...* button to the right of the Additional Language(s) display. The Select Secondary Languages dialog box appears displaying all the available languages, except the current main language.
In this example, the main language is US English and the secondary language will be Spanish.
2. Click a language name to select it. Command-click to select more than one language.
3. Command-click a selected language to remove its selection.
4. Click *OK* to save your selected language(s).



Note It is possible to read more languages than those offered as main and secondary languages, providing you disable the Language Analyst and make a suitable language selection. See *Supported languages* on page 112 for advice.

User Dictionary

Select a user (personal) dictionary in the *User Dictionary* pop-up menu. For information on creating and editing user dictionaries, see *User dictionaries* on page 103.

Use Language Analyst

Select *Use Language Analyst* to have dictionaries and other linguistic aids used during recognition. Proofing will then stop on all doubted words, and the Language Analyst may suggest replacement words. This is similar to the automatic spell-checking feature in many word processors. If this is selected, marking is available in Text view for all doubted words – those with rejected or questionable characters and those not found in a dictionary.

If you deselect *Use Language Analyst*, proofing will stop only on words containing unrecognizable characters, and only these words will be available for marking (in red) in Text view. OmniPage Pro can handle almost sixty more languages than those directly selectable (see the list in *Supported languages* on page 112). To read these languages, you must deselect *Use Language Analyst*.

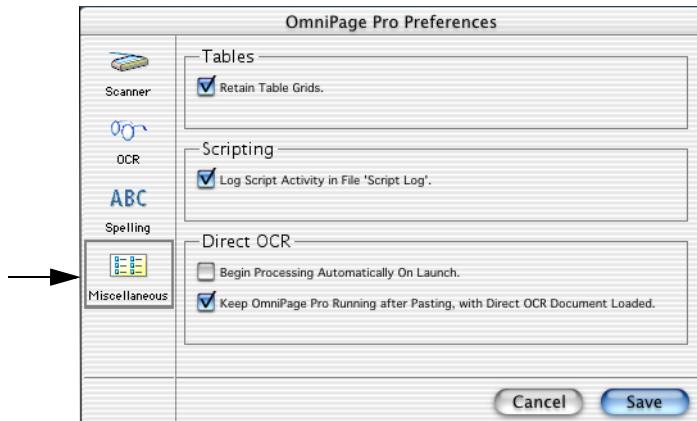
Choose other options to decrease the number of words the Language Analyst will stop on:

- Select *Ignore Proper Nouns* to ignore any word not beginning a sentence with a capitalized first letter followed by three or more lowercase letters (for example, He saw Jane throw...).
- Select *Ignore Abbreviations* to ignore a capitalized letter followed by three or fewer lowercase letters and a period (for example, Mrs., Dr., and so on).
- Select *Ignore Acronyms* to ignore any word with a capitalized letter followed by three or fewer letters of which at least one is capitalized (for example, TIFF, NASA, DoT, and so on).

Miscellaneous settings

Click the *Miscellaneous* icon on the left of the Preferences dialog box to select options for table handling, scripting and the Direct OCR feature.

Click this to see the Miscellaneous panel



Tables

Select *Retain Table Grids* to have gridded tables in the original document placed in grids in Text view after they are recognized. They will also be exported in grids if the target application supports grids.

Deselect this to have the data from all tables detected in the original document placed in tab-separated columns. Grids will not be used for export.

Scripting

Select *Log Script Activity...* to have a record of events placed in a file named 'Script Log'. This applies when OmniPage Pro X is run from the Macintosh system by AppleScript commands driving Apple Events. See the topic *Using AppleScript commands* in online Help.

Direct OCR

Direct OCR allows you to initiate OCR from the Mac OS X Dock and paste recognized text directly into another open application. (In Mac OS 9 Direct OCR is started from the Apple menu). See Direct OCR on page 68 for more information.

Direct OCR settings should be selected *before* you use the Direct OCR feature because they influence what happens as soon as you use it.

- Select *Begin Processing Automatically on Launch* if you want OmniPage Pro to trigger the *Start* button as soon as you activate Direct OCR. Text will be recognized automatically: images will be scanned or loaded, auto-zoned, recognized and (if requested) presented for proofing. Recognition results will be placed at the insertion point in the target application. Deselect *Begin Processing Automatically on Launch* if you want to control when to start scanning, loading, recognition, and pasting. This is recommended if you want to check settings, change settings from page to page, draw zones manually or verify and edit the recognized text inside OmniPage Pro.
- Select *Keep OmniPage Pro Running after Pasting, with Direct OCR Document Loaded* if you want the recognized document to be retained in OmniPage Pro. This allows you to work further with it, adding or re-recognizing pages and saving the results to file. You can save it in more than one format, including the OmniPage Document format. Deselect this setting if you do not want the recognized document to be available in OmniPage Pro after the text is pasted into your application. OmniPage Pro will also close if it was not open before you activated Direct OCR.

Note You can save all the current settings from the Preferences dialog box (except which scanner is selected) to a *settings file*. You can then load this file anytime you want to restore the preselected values. See page 104 for more information.

Chapter 5



Customizing OCR

OmniPage Pro X has many features that allow you to customize the way your documents are handled during OCR and how they appear after recognition. This chapter describes how to use these facilities.

Please continue reading for information on the following topics:

- ◆ Specifying the style set
- ◆ Applying and editing zone styles
- ◆ Zone templates
- ◆ Training OCR
- ◆ User dictionaries
- ◆ Settings files

Specifying the style set

A *style set* determines the appearance of the recognition results for each recognized page. The program is supplied with seven built-in style sets and users can create their own custom style sets.

Each *style set* contains one or more *zone styles*. A zone style defines formatting elements such as fonts, text flow, alignment and indentation to be used for text within any zone the zone style is applied to.

The following tables give an overview of the built-in style sets and the zone styles offered by each of them.

Four of these style sets define basic formatting levels. These cannot be deleted and allow only limited editing. They are useful mainly for processing documents automatically or for applying standard formatting during manual processing.

The remaining three built-in style sets can be considered samples. They can be edited and deleted. These style sets can accept new zone styles and allow the zone style values to be changed. These are useful for reformatting documents, mainly during manual processing.

Basic built-in style sets

| Style sets | Formatting | Zone style |
|-----------------|---|-------------|
| Plain Format | The whole text appears in one definable font and font size (by default 10pt. Geneva). There is no font mapping. Text is left aligned and wrapped. Multi-column text is decolumnized. | Plain |
| Similar Fonts | Font formatting is maintained. Fonts are mapped as specified, font sizes and bold, italic and underlined text are detected and maintained. Text is left aligned and wrapped. Multi-column text is decolumnized and displayed at page width. | Auto Fonts |
| Similar Formats | Font formatting, paragraph alignment and indenting are maintained. Multi-column text is decolumnized, and column widths are maintained. | Auto Detect |
| True Page | Font and paragraph formatting are maintained. Page layout is conserved by placing page elements (text blocks, headings, graphics, tables and so on) in frames. Select this only for saving formats marked with TP in the Export dialog box. | Auto Detect |

Each of these basic style sets has only one zone style. They cannot be deleted and new zone styles cannot be added. The Zone style *Plain* allows you to specify one font and font size, but cannot be edited beyond that. The zone styles *Auto Fonts* and *Auto Detect* allow only the font mapping settings to be modified.

Whichever style set is chosen, you can still apply font formatting to selected blocks of recognized text in Text view after recognition.

All four styles can transmit graphics. For the first three, the graphics are placed at the end of the recognized text. In True Page the graphic is placed in a frame in its location on the original page.

All four styles can accept tables. For the first three, tables are placed at their locations in the decolumnized text. In True Page the table is placed in a frame at its location on the original page. Tables appear either in grids or tabbed columns.

Editable built-in style sets

The following style sets are all based on the basic style set *Similar Formats*. These style sets can all be freely edited.

| Style sets | Useful for | Zone styles |
|-------------------|---|--|
| Article | Pages from magazines or newspapers you want to reformat using manual processing. Poetry or texts where the original line breaks should be conserved. | Author, Auto Detect, Body, Date of Publication, Poetry, Publication, Subject |
| Contemporary Memo | Memos or similar documents to be displayed and exported with proportionally spaced text. | Auto Detect, Body, cc, Date, From, Subject, To |
| Typewriter Memo | Memos or similar documents to be displayed and exported as monospaced text, so it appears typewritten. | Auto Detect, Body, cc, Date, From, Subject, To |

You can modify the styling of all provided zone styles except *Auto Detect*. You can add new zone styles. *Auto Detect* is set as default, but you can change the default zone style. All zone styles except *Auto Detect* can be deleted. If you try to delete the zone style selected as default, you will be warned. If you do delete it, the default reverts to *Auto Detect*.

Specifying a global style set

Select a style set from the Style Set pop-up menu in the OCR Toolbar. The selected style set is applied to all incoming pages until you change the setting. A new setting here has no effect on existing pages, even if you re-recognize them.

▼ To modify the style set for a page:

1. Make Image view active. The Zone Info palette appears.
2. Select the desired style set in its *Style Set for Page* pop-up menu. The zone styles available for the page may change.
3. If the page has already been recognized, you will have to recognize it again for the new style set to take effect.

Creating style sets

You can create and use custom style sets. This is useful for imposing consistent formatting on particular types of documents.

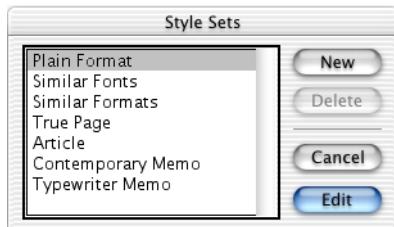
For example, if you often recognize recipes, you can design your own style set that contains a zone style for the recipe title, a style for the list of ingredients, and a style for the directions. You can then use this style set for all the recipes you recognize, even if the original pages have different layouts and formatting.

Note OmniPage Pro X is shipped with three sample style sets, for instance *Article*. You can use this as a guide when you create zone styles for your new style set. See page 97 for instructions on editing style sets.

▼ To create a style set:

1. Choose *Style Sets...* in the Edit menu.

A dialog box appears displaying all available style sets.



2. Click *New*. The New Style Set dialog box appears.

3. Enter a name for your style set.

For example, you could enter *Bibliography* as the name if you are creating a style set for handling bibliographies.

4. Click *New*.

The Edit Style Set dialog box appears. Your new style set will inherit its behavior from the style set *Similar Formats*. That means text is decolumnized, but original column widths can be maintained and frames are not used. *Auto Detect* is the only zone style automatically created.

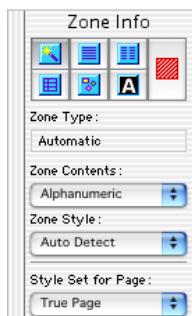
5. Add zone styles and define their properties as described in the following section.

Applying and editing zone styles

Much like applying styles to paragraphs in your word processor, OmniPage Pro allows you to apply zone styles to individual zones. The zone styles specify how text from each zone should be formatted.

Style sets and zone styles can be selected in the Zone Info palette. You can use only one style set for each page in a document. However, different style sets can be used for different pages in the same document.

▼ To apply styles to existing zones:



1. Make Image view active. The Zone Info palette appears.
2. Check that the style set for the page is suitable. Change it if desired.
3. Click the Draw/Select Zones tool in the Tools palette if it is not already selected.
4. Select the zone you want to specify by clicking it.
 - Shift-click to select additional zones.
 - Double-click the Draw/Select Zones tool or choose *Select All* in the Edit menu to select all zones on the current page.
5. Select the desired zone style in the *Zone Style* pop-up menu.
6. Select other zone properties as desired. Selecting zone type and zone contents were described on page 43.

Note **Shortcut for applying zone styles**

Hold the mouse button down while the mouse pointer is over a zone. A menu of all the zone styles in the current style set is displayed. Select the style you want to use for that zone. If the style set for the page only contains one style, no menu will appear.

▼ To apply styles to new zones:

There are two ways of doing this. Decide which you prefer:

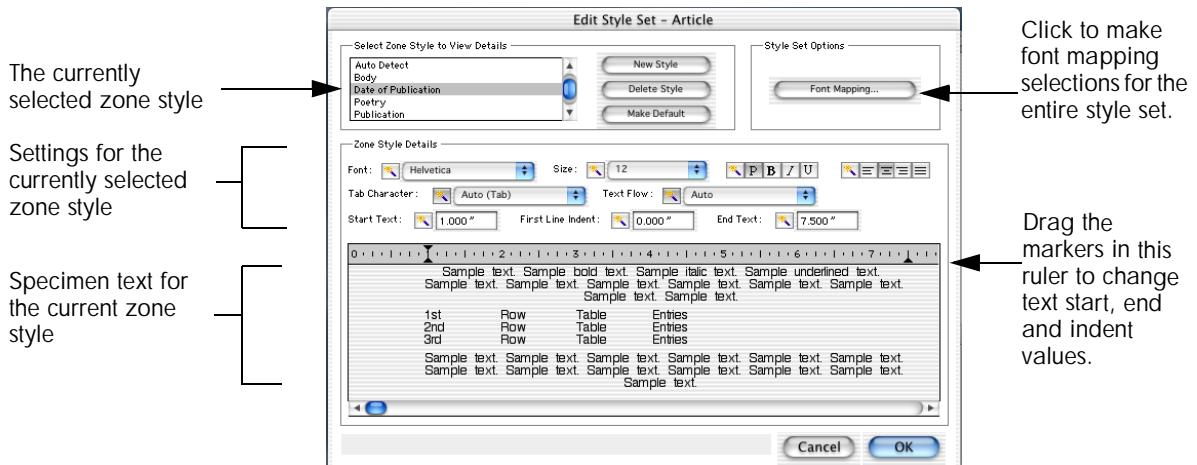
- Draw a zone. It will inherit the zone style and other properties of the last selected zone. If more than one zone is selected, the zone style is taken from the first zone in the selection.
- Make sure no zones are selected. Select the desired zone style and other properties in the Zone Info palette. Draw the zone.

▼ To edit zone styles in a style set:

The basic style sets allow very little editing. You will normally edit the built-in sample style sets or ones you have created yourself.

1. Choose *Style Sets...* in the Edit menu.
2. Double-click the style set you want to edit, or click *Edit*.

The Edit Style Set dialog box lists the zone styles in the set.



3. Click the name of the zone style you want to edit. The formatting attributes for the selected zone style are displayed.
4. Change these formatting attributes as detailed in steps 5 to 11 (described from left to right and top to bottom). Whenever the auto button to the left of an attribute is selected (pressed in), OmniPage Pro will detect and transmit the formatting for you.
5. Choose *Auto* for *Font* to have automatic character mapping (see below). Choose a font name to have it applied to all texts inside zones with this zone style instead of mapping.
6. Choose *Auto* to have the original character sizing detected and retained, or choose one fixed point size for all text in the zones.
7. Choose *Auto* to have attributes (bold, italic, underline) detected and retained from the original, or choose a value.
8. Choose *Auto* to have paragraph alignment detected and retained, or choose an alignment for all text in the zones.
9. Choose *Auto* to have tabs detected and retained. Or choose replacement character(s) to be placed instead of tabs.
10. Choose *Auto* to let the program decide whether to flow text or not. Choose *Word Wrap* to make all text flow within the text areas. Choose *Hard Line Returns* to keep all line endings as they were in the original document.

11. The last three settings define the left and right limits of the text area and first-line indenting. Choose *Auto* to let OmniPage Pro decide the values. Enter numerical values or drag the markers in the ruler to change settings.
The panel below the ruler displays the effects of your settings.
12. Repeat the above steps to edit other zone styles. Click *Delete Style* to delete a selected zone style from the style set. Click *Make Default* to make a selected zone style the default style applied to all zones when a style set is first selected for a page.

▼ **To add new zone styles to the current style set:**

1. Open the Edit Style Set dialog box and click *New Style*.
2. Enter a name for the zone style you want to add and click *OK*. For example, you could enter *Heading* as the name if you are creating a style for heading-type paragraphs.
3. Modify the desired formatting attributes for the new style, as described in the previous procedure.
Repeat steps 2-4 to continue adding new styles to the style set.
4. Click *OK* when you are finished editing the style set.
5. Click *Done* in the Style Sets dialog box if you do not want to edit any other style sets.

Font mapping

If *Auto* is selected as the font setting for a zone style, OmniPage Pro analyses the text styling inside the zone and assigns it to one of four categories. More than one text category may be detected within a single zone. Each category is *mapped* to a font which you can specify.

◆ **Proportional Serif**

Character widths vary and short lines finish off letter strokes. This text is an example of this font type. The default font is Times.

◆ **Proportional Sans-Serif**

Character widths vary; letter strokes do not have finishing lines. The default font is Helvetica.

◆ **Monospaced Serif**

Character width is the same for each character; short lines finish off the letter strokes. The default font is *Courier*.

◆ **Monospaced Sans-Serif**

Character width is the same for each character; letter strokes do not have finishing lines. The default font is *Monaco*.

Note Font mapping is not applicable to the Plain Format style set. It is always performed with the style sets Similar Fonts, Similar Formats or True Page. It is available but not compulsory for editable style sets.

Note To avoid font mapping during manual processing, specify a font name for a zone style in place of *Auto*. This font will be applied to all text in all zones with this zone style. To avoid font mapping in automatic processing, select an editable style set, define a zone style with a specific font name instead of *Auto*, make this the default zone style and then choose the style set in the OCR Toolbar before starting the automatic processing.

▼ **To change font mapping for a style set:**

1. Choose *Style Sets...* in the Edit menu.
2. Double-click the style set for which you want to change font mapping selections.
3. Click *Font Mapping...* in the Edit Style Set dialog box. The Automatic Font Mapping dialog box appears.



4. Select the font you want used for each category. You can select any fonts available on your system.



Zone templates

You can use a *zone template* to quickly and efficiently create zones on documents that have the same zoning requirements. For example, if you frequently process documents with layouts and content that require the same type of zoning, you can create and save a zone template and apply it to all such pages or documents.

A zone template can have up to 64 zones. It remembers the size, position, order, type, style and contents of zones.

OmniPage SE for Mac does not support Zone Templates.

▼ To save a zone template:

1. Create the desired zones on a page image, manually or automatically with checking and modification as required. See *Creating zones automatically* on page 42.
2. Choose *Save Zone Template...* in the File menu. The Save Zone Template dialog box appears.
3. Type a name for your file and click *Save*. The zone template file is saved in the *Zone Templates* folder within your installation folder.

▼ To apply a zone template to future pages:

- Select the zone template you want to use in the Original Layout pop-up menu on the OCR Toolbar. OmniPage Pro places template zones on all incoming page images while the template remains in effect.

▼ To apply a zone template to an existing page:

1. Make sure the desired template is selected in the Original Layout pop-up menu on the OCR Toolbar.
2. Make Image view active, with the desired page displayed.
3. Click the Apply Template tool in the Zone Info palette.

▼ **To remove a zone template:**

- Select a non-template setting in the Original Layout pop-up menu on the OCR Toolbar.

OmniPage Pro will no longer place template zones on incoming page images. This does not remove template zones from existing zoned pages. Just delete or modify them or choose *Discard Current Zones and Find New Zones* in the Zoning Instructions dialog box.



Training OCR

You can create a *training file* to handle characters that are being consistently misrecognized. A training file is a set of up to 256 pre-recognized character shapes each linked to an OCR solution.

OmniPage Pro compares the stored shapes with those encountered on incoming documents. OmniPage SE does not support this feature.

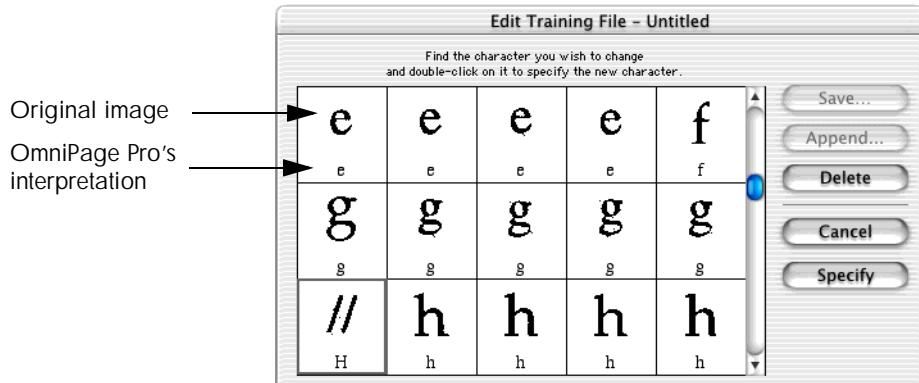
OmniPage Pro X is a powerful, pre-trained OCR product. For recognizing ordinary characters in everyday fonts, training files should not be needed. Training is useful mainly for long documents (or a set of documents) in which a few character shapes are being repeatedly misrecognized in the same way. Training is not useful for poorly formed characters unlikely to occur again in the document. For instance, a character shape damaged by spots on the image is a poor candidate for training. Do not attempt to create a training file for an unsupported language or alphabet.

▼ **To create a training file:**

1. Open an image file or scan a page that includes the characters you want to train or use a page you have already recognized.
If you select a recognized page, its recognition results are deleted. Accept the invitation that appears when you finish, to re-recognize the page with the new training file.
2. Create or modify zones on the page image if you want to train characters from only part of the page.
3. Select *Train OCR* as the option in the OCR pop-up menu.



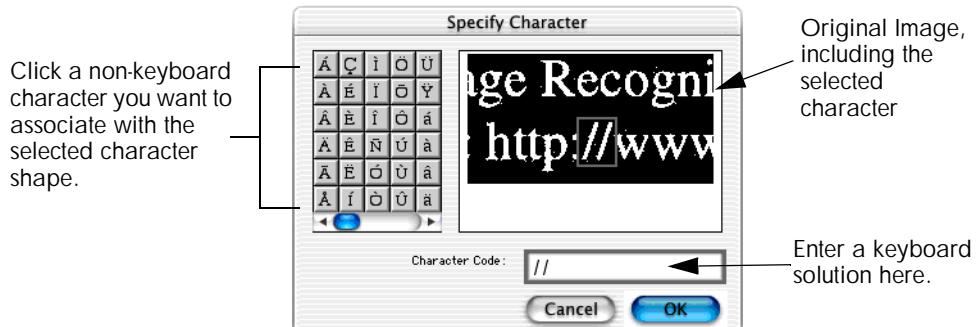
- Click the *OCR* button. OmniPage Pro analyzes the page and opens the *Edit Training File* dialog box. Original character images are displayed along with OmniPage Pro's interpretation of each character. Characters appear in the alphabetical order of their interpretations.



Most characters do not need to be trained. Look for uncommon and run-together characters. Look for characters whose interpretation is incorrect. An example in the picture above is the bottom left square.

- Double-click a character you want to train. Or select it and click *Specify*.

The *Specify Character* dialog box displays the selected character as it appears in the original page image.



- Specify how you want OmniPage Pro to interpret the character shape during OCR. Type the desired character(s) in the *Character*

Code edit box, or click a non-keyboard character in the scrolling display to add it to the edit box.

In our example, the 'H' has been cleared and '//' entered.

7. Click *OK* to accept the character specification.
The Training File dialog box reappears.
8. Repeat steps 5–7 to continue specifying characters.
The *Delete* button is not needed when you create a new training file. Any untouched character is excluded from the training file.
9. Click *Save...* to save the characters whose solutions you changed to a new training file which you will name.
Or, click *Append...* to add these characters to an existing training file which you select. In this case, no new training file is created.
After saving or appending a file, you are asked if you want to make this the current training file. Click *OK* to (re-)recognize the current page using the training file you have just created. Click *Cancel* to return to the image without recognizing it.

▼ **To load a training file:**

1. Choose *Preferences...* from the Application menu (OS 9: Edit).
2. Click the OCR icon to display the OCR panel.
3. Select a training file in the Training File pop-up menu.
This file remains loaded until you unload it or replace it with another training file.

▼ **To unload a training file:**

1. Choose *Preferences...* from the Application menu (OS9: Edit).
2. Click the OCR icon to display the OCR panel.
3. Select *None* in the Training File pop-up menu.

Note It is important to unload a training file when you finish processing pages for which it was prepared. A training file is likely to lower accuracy if it remains loaded for pages with different typestyles.

▼ **To edit a training file:**

1. Choose *Training Files...* in the Edit menu. The Training Files dialog box lists all training files in the Training Files folder.
2. Double-click the training file you want to edit, or select it and click *Open*.
The Training File dialog box displays the characters in the training file you specified.
3. Double-click a character you want to edit.
The Specify Character dialog box appears.
4. Edit the interpretations associated with the selected character shape, as described under *Creating a training file*. Type one or more characters into the *Character Code* edit box or select non-keyboard characters from the scrolling display.
5. Click *OK* to accept each character specification and repeat steps 3 and 4 to continue editing specified characters.
Click *Delete* to discard a selected character from the training file. Untypically misformed character shapes are bad candidates for training and should be deleted.
6. Click *Save...* to save the edited training file under its existing name. Or, click *Append...* to add the trained characters to an existing training file. The file you selected to edit will not be modified.

▼ **To delete a training file:**

1. Choose *Training Files...* in the Edit menu.
2. Select a training file to be deleted.
3. Click *Delete* and then *OK* in the warning box. Click *Done*.

User dictionaries

Dictionaries are used to assist recognition and provide suggestions during proofing. A *user dictionary* is a personal dictionary that you build and customize, to supplement a built-in main dictionary. Entries for a user dictionary must consist of 2 to 32 characters, without spaces or control characters, such as tabs. The program is supplied with one empty user dictionary, named *User Dictionary*.

▼ To create or edit a user dictionary:

1. Choose *User Dictionaries...* in the Edit menu. The User Dictionaries dialog box lists all user dictionary files.
2. Do one of the following:
 - Select a file and click *Open* to edit an existing user dictionary.
 - Click *New* to create a new user dictionary. Enter a name in the dialog box that appears and click *New*.
 The Edit User Dictionary dialog box appears.

The words in an existing user dictionary appear in the list box. No words are listed for a new dictionary.



3. Add or delete words as desired:
 - Type a word in the *New Word* edit box and click *Add* to add it.
 - Select a word in the list box and click *Delete* to delete it.
 - Click *Delete All* to remove all words from the dictionary.
 - Click *Import...* to add all words from a specified plain text file, with each word on a separate line.

4. Optionally, click *Export...* to save your user dictionary as a plain text file, for protection or use outside the program.
5. Click *Done* to save the changed state of your user dictionary within the program and exit.

User dictionaries are saved in the *User Dictionaries* folder within your installation folder. Select one for use in the *Spelling* panel of the Preferences dialog box. Select *None* to unload a user dictionary.

Words can also be added to the loaded user dictionary during proofing (see page 53).

Settings files

You can save customized settings to a *settings file*. This is useful for quickly restoring OmniPage Pro to settings required by particular documents. A settings file contains all settings made in all panels of the Preferences dialog box, except your current scanner selection (and the training file name in OmniPage SE for Mac). To change this, use the Scanner panel of the Preferences dialog box.

▼ To save settings:

1. Check the Preferences dialog box to be sure all its settings are suitable for saving to file.
2. Choose *Save Settings...* in the File menu.
The Save Settings File dialog box appears.
3. Type a name for your settings file.
4. Click *Save* to save the settings file in the *Settings* folder, located within your installation folder (under *Components*).

▼ To load settings:

1. Choose *Load Settings...* in the File menu.
2. Double-click the settings file you want to load, or select it and click *Load*.

You cannot unload a settings file. Just change settings as required.

Chapter 6



Technical information

This chapter provides troubleshooting and other technical information to help you use OmniPage Pro X.

Please also consult the PDF Readme file and other online help topics, or visit the Support section in the ScanSoft web pages. This answers Frequently Asked Questions (FAQ) and provides other useful guidance. The web site includes a Scanner Guide with regularly updated information about supported scanners. Access to ScanSoft's web pages is provided from the online Help topic *Getting Help*.

This chapter contains the following information:

- ◆ Troubleshooting
 - ◆ Solutions to try first
 - ◆ Low memory situations
 - ◆ Low disk space situations
 - ◆ Improving accuracy
 - ◆ Improving fax recognition
 - ◆ Interface problems and solutions
 - ◆ System failure during OCR
- ◆ Supported languages
- ◆ Supported saving formats
- ◆ Supported image file formats

Troubleshooting

Solutions to try first

Try these solutions if you experience problems starting the program:

- ◆ Ensure that your system meets all requirements listed under *System requirements* in chapter 1.
- ◆ Make sure that your scanner is plugged in and that all cable connections are secure.
- ◆ Turn off your computer and your scanner, turn your scanner back on, and then restart your computer. Make sure other applications are functioning properly.
- ◆ Use the software that came with your scanner to verify that it is working properly before using it with OmniPage Pro.
- ◆ Make sure you have the correct and up-to-date drivers for your scanner, printer and video card. See the Scanner Guide on ScanSoft's web site for more information.
- ◆ Delete the file 'OmniPage Pro X Prefs' if unsuitable settings are generating error messages or problems. The program will create a new preference settings file with default values.
- ◆ Run Disk First Aid to check your hard disk for errors. See Macintosh Help for more information.

Low memory situations

OmniPage Pro may run slowly or poorly under low-memory conditions. Try these solutions if you get low memory warnings:

- ◆ Restart your computer.
- ◆ Close other open applications to release memory.
- ◆ Increase the amount of free hard disk space.
- ◆ Increase your computer's physical memory (RAM).

- ◆ Do not scan in color unless you need colored graphics in your output files. Prefer *Web color* or *256 colors* (8-bit pixel depth) rather than *True color* (16-bit depth) or similar choices.
- ▼ **To adjust preferred memory size for an application under OS 9.X:**
 1. Make sure OmniPage Pro X is closed.
 2. Select *OmniPage Pro X* under *Components* in the program folder.
 3. Select *Get Info* then *Memory* from the File menu of the Finder.
 4. Adjust *Preferred Size* under *Memory Requirements*.

Low disk space situations

Problems may occur if your system runs low on free disk space. Try these solutions for low disk space situations:

- ◆ Empty the Trash.
- ◆ Close all open applications that are not immediately needed.
- ◆ List your OPD files. Delete any you no longer need. Open OPD files and save their recognition results as desired, then delete them. OPD files tend to be large, especially if they contain color pages. To keep OPD files as a document archiving system, consider transferring them to a ZIP drive or another mass storage device.
- ◆ Delete files that are no longer needed. Transfer large but seldom used files to backup storage.
- ◆ Run Disk First Aid (in the Utilities Folder) to check for errors that may be using disk space. See Macintosh online Help.

Improving accuracy

Try the following solutions if accuracy is lower than you expected.

Acquire high-quality images

- ◆ In general, try to use original pages when scanning documents. High-quality typeset pages yield the best OCR accuracy.

- ◆ With low-quality originals, sometimes a good-quality photocopy can yield better OCR results. This may be true for documents with low contrast or printed on thin paper. On the other hand, poor-quality photocopies with stripes, blotches or uneven brightness will usually give worse results.
- ◆ Page images should be free of notes, lines, doodles or spots. Anything in a text zone that is not a printed character slows recognition. Exclude such marks from text zones, or enclose them in Ignore-type zones or use the eraser in the Tools palette to delete them from the image.
- ◆ Check the glass, mirrors, and lenses in your scanner for dust, smudges, or scratches. Clean if necessary.
- ◆ If your only criteria is OCR accuracy, prefer black-and-white scanning for good quality documents with crisp black text on a white background. Choose grayscale scanning if you are scanning pages with text on colored or shaded backgrounds, or for degraded documents with low or varied contrast.
- ◆ Adjust the brightness and contrast sliders in the Scanner panel of the Preferences dialog box, or on the scanner's own interface. Or choose Auto-brightness, if available. Experiment with different settings combinations to get the desired results. See how to optimize brightness on page 81.
- ◆ Text in page images should be reasonably clean and crisp. Characters should be separated from each other and not blotched together or overlapping. Characters distorted by marks or smudges may be unrecognizable.
- ◆ If you have influence over the styling used in documents you want to recognize, avoid having underlines used. It is difficult to recognize underlined text because the underline changes the shape of descenders on the letters q, g, y, p, and j.
- ◆ Check the image resolution by selecting *Show Page Info* from the File menu. The ideal resolution for OCR is 300 dpi. Images with 200 to 250 dpi or more than 400 dpi are liable to yield lower

accuracy. The program will not open image files with resolutions below 200 dpi. If this happens and you have the documents on paper, scan them again with better settings.

Ensure zones are suitable

- ◆ Look at the original page images and ensure that all required text areas are enclosed by text zones. If an area is not enclosed by a zone, it is generally ignored during OCR.
- ◆ Make sure zone borders do not cut through text and the graphics are correctly zoned. Resize zones as necessary.
- ◆ Make sure text zones are specified correctly. Change zone types, zone contents, or zone styles as necessary, and perform OCR on the document again. See *Specifying zone types* on page 43.
- ◆ Be sure you do not have an unsuitable zone template loaded by mistake. If zone borders cut through text, recognition is impaired.
- ◆ Be sure the original layout option you selected best describes your incoming pages because this influences auto-zoning.
- ◆ To retain handwritten text, such as a signature, specify it as a graphic zone and be sure *Retain Graphics* is specified.

Use suitable recognition settings

- ◆ Make sure the correct main recognition language is selected in the Spelling panel of the Preferences dialog box. Select secondary languages only if the document really contains them. A flood of blue words in Text view suggests an incorrect language choice.
- ◆ Check in the OCR panel that *Dot Matrix* is not selected for Character Type, unless the document really contains draft-mode 9-point dot-matrix text.
- ◆ Use the Train OCR mode, or use a suitable existing training file. This is most likely to help with stylized fonts or uniformly degraded documents. See *Training OCR* on page 99.

Hints relating to Train OCR, training files or Zone Templates do not apply to OmniPage SE for Mac.



- ◆ If you are getting poor results with a training file loaded, check its contents by clicking *Training Files...* from the Edit menu. Make sure the training file is appropriate for the current document. If it is not, either unload it or edit its contents to remove training from poorly formed character shapes. Unsuitable training can yield worse results than no training at all.
- ◆ If proofing is skipping too many unsuitable words, be sure *Use Language Analyst* is enabled. If you have a user dictionary loaded, check its contents by choosing *User Dictionaries...* from the Edit menu. Delete entries added in error, especially misspelled words.
- ◆ If the recognition results do not appear in Text view as you expected, consult the Zone Info palette to check that your style set selection is appropriate. Check that the zone styles are suitably assigned and defined. See chapter 5.
- ◆ With the True Page style set, recognized text is put into frames (formatting boxes). Some text may be hidden if a frame is too small. You can see a plus (+) sign in the bottom right corner of the frame in this case. To view the text, place the cursor in the text frame and use the arrow keys on your keyboard to scroll to the top or bottom of the frame. Reduce the point size of framed text to make the whole text visible, or resize the frames in your target application, or choose *Remove Frames on Export* in the Export dialog box.

Improving fax recognition

Try these solutions to improve OCR accuracy on fax images:

- ◆ Ask senders to use clean, original documents if possible. Sans serif fonts are easier to recognize than serif fonts.
- ◆ Ask senders to select Fine or Best Mode when they send you a fax. This produces a resolution of 200 x 200 dpi.
- ◆ Ask senders to transmit files directly to your computer via fax modem if you both have one. You can save fax images as image files and then load them into OmniPage Pro. See *Loading image files* on page 38 for more information.

Interface problems and solutions

The *Start* button is disabled.

Be sure *Train OCR* is not selected in the OCR pop-up menu. Training can only be done on a single page at a time.

The *Save* button in the Preferences dialog box is grayed.

Change a setting in one of the panels, then it will become available.

The Verify window refuses to appear.

Keep this window open or close it; do not minimize it. If it remains minimized, it cannot jump to a new location.

Image view has disappeared completely.

Drag the splitter between the two views to the right.

The table editing tools in the Tools palette are grayed.

These become available only if the current page contains a table zone.

The Export pop-up menu offers no choices.

You are probably using Direct OCR, which places the value *To Application*. The pop-up menu becomes available again only when recognition results have been placed in the target application.

System failure during OCR

Try these solutions if a system failure occurs during OCR or if processing takes a very long time:

- ◆ Resolve low memory or low disk space problems.
- ◆ Check the quality of the images you are recognizing.
- ◆ Consult your scanner documentation on ways to improve the quality of scanned images.
- ◆ Break complex page images (lots of text blocks and graphics or elaborate formatting) into smaller jobs. Draw zones manually or modify automatically created zones and perform OCR on one page area at a time.

Supported languages

The program supports thirteen languages with a main dictionary and Language Analyst. The program can recognize other languages, but without these facilities. To read text in these languages, select the language(s) indicated and deselect *Use Language Analyst* in the Preferences dialog box. Proofing suggestions will not be available, but in most cases recognition accuracy should remain acceptable.

| To read: | Select: |
|---------------|--------------------------|
| Afrikaans | Dutch |
| Albanian | French |
| Aymara | Spanish |
| Bemba | English |
| Blackfoot | English |
| Breton | French and Spanish |
| Bugotu | English |
| Catalan | French and Spanish |
| Chamorro | Finnish |
| Crow | English |
| Estonian | Finnish and Portuguese |
| Flemish | This is Dutch in Belgium |
| Frisian | French |
| Friulian | French and Italian |
| Gaelic | Italian |
| Guarani | Spanish and Finnish |
| Hani | English ** |
| Hawaiian | English |
| Ido | English |
| Indonesian | English |
| Interlingua | English |
| Kawa | English** |
| Kongo | English |
| Kpelle | English |
| Latin | English* |
| Luxembourgian | French |
| Malagasy*** | French and Portuguese |
| Malinke | French |
| Maori | English |

| To read: | Select: |
|---------------|--------------------------|
| Mayan | Finnish |
| Miao | English ** |
| Mohawk | English |
| Nahuatl | English |
| Nyanja | English |
| Occidental | English |
| Papiamento | Spanish and French |
| Pigin English | English |
| Provencal | French and Dutch |
| Quechua | Spanish |
| Rhaetic | French, German, Italian. |
| Ruanda | English |
| Rundi | English |
| Shona | English |
| Sioux | English |
| Somali | English |
| Sotho | English |
| Sundanese | English |
| Swahili | English |
| Tagalog | English |
| Tahitian | French and Italian |
| Tanna | English |
| Tinpo | English** |
| Tongan | English |
| Tun | English** |
| Visayan | English |
| Welsh | French**** |
| Wolof | French |
| Xhosa | English |
| Zapotec | English |

* Latin with diacritical marks cannot be handled.

** Supported only when written in the Latin alphabet.

*** Some dialects do not use accents; then select English.

**** The rare letters w- and y-circumflex cannot be handled.

The accented letters used in less spoken languages may vary with dialects, variants, changes over time and transcription norms. Therefore, this table can serve only as a general guide.

Supported saving formats

Recognition results can be saved to a wide range of target applications and saving formats. The following table provides information on this:

| Save format | File extension | True Page support | Graphics supported |
|--------------------------------------|----------------|-------------------|--------------------|
| OmniPage Document | opd | Yes | Yes |
| ASCII Text | txt | No | No |
| ASCII Text with line breaks | txt | No | No |
| ClarisWorks/AppleWorks (RTF/MacLink) | rtf | No | Yes |
| Excel 98, 2001, X | xls | Yes | Yes |
| FrameMaker 4.0, 5.0, 5.5 | fm | No | Yes, from 5.0 |
| HTML 2.0 | html | No | No |
| HTML 3.2 | html | No | Yes* |
| HTML 4.0 | html | Yes | Yes* |
| MacWrite Pro | | Yes | Yes |
| PDF, image only | pdf | ** | Yes |
| PDF, normal | pdf | ** | Yes |
| PDF, with image on text | pdf | ** | Yes |
| PDF, with image substitutes | pdf | ** | Yes |
| RTF 1.0 and 2.0 | rtf | Yes | Yes |
| Word 98, 2001, X | doc | Yes | Yes |

* Each graphic area is saved to a separate JPEG file within a separate folder, saved to the same location as the HTML file. When the HTML file is loaded into an HTML viewer or editor, the JPEG images are embedded, provided you have not moved, deleted or edited them.

** The PDF pages take their appearance from a True Page representation of each page, regardless of the style set used during processing. See page 76.



OmniPage SE for Mac does not support HTML 4.0 and the PDF formats.

Supported image file formats

Page images can be acquired from image files. Scanned images can be saved to file: current page only, all document pages (one file per page or one multipage file), or each graphic zone on a page to a separate file. The following table details the program's image file support.

| Formats | Multipage | Open/Save | Black-and-white, Grayscale, Color |
|-----------------------|-----------|---------------|--------------------------------------|
| BMP (Windows Bitmap) | No | Open and Save | All |
| GIF | No | Open | All |
| JPEG | No | Open and Save | Grayscale, Color |
| PDF | Yes | Open * | All |
| Photoshop PSD | No | Open and Save | All |
| PICT | No | Open and Save | All |
| PNG | No | Open and Save | All |
| TIFF Compressed G3/G4 | Yes | Open and Save | B/W only |
| TIFF Packbits | Yes | Open and Save | All |
| TIFF Uncompressed | Yes | Open and Save | All |

* The program offers PDF as a saving format, but it is the recognition results that are saved, not the original images. The PDF saving option 'images only' means recognized pages are saved to a PDF file that can be viewed but not edited.



OmniPage SE for Mac does not support Windows BMP, GIF, Photoshop PSD, PNG and PDF formats.

We trust this User's Guide and the online Help will assist you in getting the most out of ScanSoft's OmniPage Pro X or SE and that it will make your work more productive and satisfying.



I N D E X

A

Abbreviations, ignoring, 86
Accuracy
 best resolution for, 80
 brightness options for, 81
 improving, 107
Acquiring images, 23, 31, 34, 38
Acronyms, ignoring, 86
Adding
 areas to zones, 49
 new styles to a style set, 96
 pages to a document, 33, 40
 trained characters to training files, 101
 words to a user dictionary, 54
ADF (automatic document feeder)
 settings, 80
 when to use, 38
Alphanumeric zone, 46
Applying zone styles, 94
 shortcut for, 94
Article style set, 76
ASCII text output, 113
Auto Detect zone style, 90, 91, 93
Automatic
 font mapping, 95, 96
 processing, 30
 proofreading, 53
 spell checking (Language Analyst), 86
 zones, 42
Automatic document feeder
 see **ADF**
Automatic zone type, 43
Auto-zoning, 30, 42, 74

B

Basic formatting levels, 90
Basic processing steps, 23, 30
Black-and-white scanning, 72
Books, scanning, 80
Brightness, 81
Built-in style sets
 Article, 76
 Contemporary Memo, 76
Plain Format, 75, 90

Similar Fonts, 75, 90
Similar Formats, 76, 90
True Page, 66, 76, 90, 113
Typewriter Memo, 76

C

Chapter outline, 7
Character mapping, 95, 96
Character sets, selecting, 85
Character type, selecting, 82
Characters
 checking for errors, 53
 deleting from training file, 102
 specifying for training, 100
 unrecognizable, 53, 83
 verifying against image, 55
 when to train, 99
Checking OCR results, 53, 55
Clipboard
 copying a document to, 66
 copying selection to, 61
 copying zones to, 67
Closing a document, 62
Color markers, 53, 56
Color scanning, 73
Column dividers
 inserting in tables, 51
Command key symbol, 8, 50
Comparing text with images, 55
Connecting zones, 50
Contemporary Memo style set, 76
Contents of OmniPage Documents, 40
Contrast, 81
Control over processing, 34
Conventions in this Guide, 8
Conversion of image files, 114
Copying
 document to the Clipboard, 66
 selections to the Clipboard, 61
 zones to the Clipboard, 67
Creating
 style sets, 92
 training files, 99
 user dictionaries, 103
 zone styles, 96
 zone templates, 98
Custom style sets, 76, 92

Cutting text or graphics, 61

D

Deleting
 characters from training file, 102
 current page, 59
 graphics, 61
 style sets, 93
 text, 61
 training files, 102
 zone styles from a style set, 96
 zones, 50
Describing document layout, 31, 74
Deselecting a selected page, 60
Direct OCR
 about, 68
 settings, 70, 87
 supported applications, 68
 using, 69
Dividing zones, 50
Document
 checking for errors in, 53
 closing, 62
 copying, 66
 exporting, 63
 printing, 61
 processing automatically, 30
 processing manually, 34
 with varied layout, 74
 working with, 57
Document window, 25, 26
Dot matrix texts, 82
Double-sided pages
 ADF settings, 80
Drag-and-drop functionality, 40, 67
Drawing zones
 one side at a time, 47
 rectangular, 46

E

Editing
 PDF output, 66
 recognized text, 60
 style sets, 94
 training files, 102
 user dictionaries, 103

zones styles, 95
English texts read aloud, 62
Erasing image areas, 60
Export
 To Application, 68, 78
 To Clipboard, 78
 To File, 77
Export button, 24, 34, 35, 77
Exporting documents
 copying to the Clipboard, 66
 saving recognition results, 64

F
Fax recognition, 110
File types, supported, 113
Finding suspect words, 53
Font attributes, 95
Font formatting, 61
Font mapping, 95, 96
Font size, changing, 61
Formatting text, 61
Frames, supported, 76

G
Get Page button, 24, 34, 72
Getting online Help, 9
Going to a particular page, 58
Graphic zone type, 44
Graphics
 copying graphic zones, 67
 cutting or copying from Text view, 61
 deleting, 61
 retaining during OCR, 83
Grayscale scanning, 73

H
Hearing text read aloud, 62
Help, online, 9

I
Ignore zone type, 45
Ignoring
 abbreviations, 86
 acronyms, 86
 page areas during OCR, 41, 45
 proper nouns, 86
Image files
 conversion of, 114
 formats, supported, 114
 loading, 38, 73
Image substitutes in PDF, 66
Image view, 26

I Images
 acquiring, 23, 31, 34, 38
 bringing into OmniPage Pro, 38
 defined, 22
 erasing areas of, 60
 loading, 38
 modifying, 59
 printing, 61
 reordering pages, 58
 rotating, 59
 saving, 63, 114
 scanning pages, 38
 substitutes in PDF, 66
 thumbnails of, 26
 zooming in and out, 57

Input
 from image file, 31, 73
 from scanner, 31, 72
Inserting
 column dividers in tables, 51
 row dividers in tables, 51
Installing
 OmniPage Pro X, 14
 selecting a scanner for OmniPage Pro X, 17
Interface problems, 111
Irregular zones, 47

L
Language Analyst, 53, 56, 86
Languages
 for reading aloud, 62
 for recognition, 85
 supported, 112
Loading
 a settings file, 104
 a training file, 101
 a user dictionary, 86
 image files, 38, 73
Low disk space problems, 14, 107
Low memory problems, 14, 106

M
Manual
 processing, 34
 zoning, 46
Markers, 53, 56
Memory requirements, 14
Miscellaneous settings, 87
Modifying
 images, 59
 text, 60
 zones, 48
Moving
 table dividers, 51
 to pages, 58

zones, 48
Multi Column Text zone type, 44
Multi-page image files, 39
Multiple column pages, 74
Multiple-page document
 using an ADF with, 80

N
New features, 11
Numeric zone, 46

O
OCR
 automatic processing, 30
 basic steps of, 23
 defined, 22
 manual processing, 34
 performing, 52
 system failure during, 111
 training, 99
OCR Assistant, 36
OCR button, 24, 34, 52
OCR settings, 82
 character type, 82
 Language Analyst, 86
 retaining graphics, 83
 training files, 82
OCR Toolbar
 buttons, 24, 30, 72
 language display on, 24
 pop-up menus, 24, 30, 72
 selecting options on, 72

OmniPage Documents
 description of, 40
 opening, 40
 saving, 63
 why to save, 58, 63
OmniPage Pro X
 basic processing steps, 30
 installing, 14
 new features, 11
 quitting, 62
 running under Mac OS 9, 15
 settings, 28
 starting, 16
 system requirements, 14
 user interface, 25

OmniPage SE for Mac, 7, 12
Online HTML Help, 9
Open document
 adding images to, 40
 creating zones on, 41
 exporting, 63
 performing OCR on, 52
 proofreading, 53

Opening
 OmniPage Documents, 40
Optical character recognition
 see **OCR**
Optimizing image quality, 81
Ordering zones, 48
Orientation
 rotating an image manually, 59
 selecting for scanning, 79

P

Pages
 adding to a document, 40
 deleting current, 59
 going to, 58
 loading images files, 38
 processing all unrecognized, 33
 reordering, 58
 reprocessing, 33
 re-recognizing, 48
 resizing view of, 57
 rotating in Image view, 59
 scanning, 38
 size and orientation, 79
Paragraph attributes, 95
PDF input, 38, 114
PDF output, 66, 113
Performing OCR, 34, 52, 77
Photoshop plug-in, 18
Plain Format style set, 75, 90
Preferences dialog box, 28, 78
 Miscellaneous panel, 87
 OCR panel, 82
 Scanner panel, 78
 Spelling panel, 84
Printing
 images, 61
 setup options, 62
 text, 61
Problems
 during OCR, 111
 with fax recognition, 110
 with interface, 111
Procedures, processing, 31, 34
Processing documents
 automatically, 30
 automatically and manually, 35
 from other applications, 68
 in future sessions, 58, 63
 manually, 34
Processing overview, 30
Proofreading, 53
Proper nouns, ignoring, 86
Purpose of OPD files, 58, 63

Q
Quitting OmniPage Pro X, 62

R

Reading text aloud, 62
Recognizing text, 52, 77
Rectangular zones, 46
Redetecting table dividers, 51
Reject character
 default value, 53
 specifying, 83
 stop on, 56
Remove Frames on Export, 65
Removing table dividers, 51
Reordering
 pages, 58
 zones, 48
Reprocessing pages, 33
Resizing
 a page display, 57
 zones, 48
Resolution, 63, 80
Restricted shapes for zones, 47
Retain Graphics setting, 44, 52, 65, 66, 68, 83
Retain Table Grids, 87
Reverse Text zone type, 44
Rotating images, 59
Row dividers
 inserting in tables, 51
Running OmniPage Pro X, 16

S

Sample style sets
 Article, 76, 91
 Contemporary Memo, 76, 91
 Typewriter Memo, 76, 91
Save and Launch, 32, 65
Saving
 current document, 58, 63
 formats, 113
 images, 63
 recognition results, 64
 settings files, 104
 text, 64
 to OPD format, 58, 63
 to PDF, 66
 training files, 101
 user dictionary as text file, 104
 zone templates, 98
Scanner
 ADF options, 80
 selecting, 17
 settings, 78

Scanning
 black-and-white, 72
 books, 80
 color, 73
 double-sided pages, 80
 grayscale, 73
 pages, 38
 resolution, 80

Script Log file, 87
Searching PDF output, 66

Selecting
 all text, 60
 languages, 85, 112
 options, 72
 scanners, 17
 settings, 78
 style sets, 92, 93
 training files, 82
 user dictionaries, 86
 zone styles, 93
 zone templates, 75
 zones, 48

Settings
 Direct OCR, 69, 88
 Miscellaneous, 87
 OCR, 82
 Scanner, 78
 Spelling, 84

Settings files
 loading, 104
 saving, 104

Showing or hiding markers, 53, 56
Showing page info, 56
Similar Fonts style set, 75, 90
Similar Formats style set, 76, 90
Single column pages, 74
Single Column Text zone type, 44
Solutions for poor performance, 106
Spanish texts read aloud, 62

Specifying
 reject character, 83
 zone contents, 45
 zone types, 43

Spelling
 checking for errors, 53
 settings for, 84
 using the Language Analyst, 86

Spreadsheet pages, 74
Start button, 24, 30, 31, 32
Starting OmniPage Pro X, 16
Step-by-step processing, 34
Stop button, 32
Stopping automatic processing, 32

Style sets
 creating, 92
 Custom style sets, 76
 defined, 89
 deleting, 93

editing, 94
selecting, 92, 93
Style sets, built-in
Article, 76, 91
Contemporary Memo, 76, 91
Plain Format, 75, 90
Similar Fonts, 75, 90
Similar Formats, 76, 90
True Page, 66, 76, 90
Typewriter Memo, 76, 91
Subtracting from zones, 49
Suggestion from dictionaries for proofing, 86
Supported file formats, 113
Suspect words, 53, 56
System requirements, 14

T

Table dividers, 51
Table zone type, 44, 47, 51
Tables
column dividers in, 51
retain grids, 87
row dividers in, 51
Technical information, 105
Templates, 75
Text
checking for errors, 77
copying, 61
cutting, 61
deleting, 61
drag-and-drop, 68
editing, 60
formatting, 61
PDF output, 66
printing, 61
recognizing, 77
saving, 64
saving formats, supported, 113
selecting, 60
verifying, 55
Text recognition
creating zones for, 34, 41
Text view, 26
Text zones, 45
Thumbnail view, 26, 58
Thumbnail window, 26
reordering pages in, 58

Toolbar
see **OCR Toolbar**
Tools palette, 27, 46
Train OCR, 77
Trained characters
appending to another file, 101
deleting, 102
Training files
creating, 99
deleting, 102
editing, 102
loading, 101
saving, 101
selecting for OCR, 82
unloading, 101
Troubleshooting, 106
True Page style set, 66, 76, 90
True Page support, 113
TWAIN driver, 18
Typewriter Memo style set, 76

U

Undoing edits, 59
Unrecognizable characters, 83
User dictionary
creating or editing, 103
loading, 86
saving as text file, 104
selecting, 86
Using drag-and-drop, 40, 67
Using the Language Analyst, 86

V

Verification window, zooming in or out, 55
Verifying text, 55
Viewing PDF output, 66

W

Word processor formats, supported, 113
Word wrap, 95

Z

Zone contents
copying with drag-and-drop, 67
specifying, 45
Zone Info palette
applying zone styles, 94
applying zone types, 43
described, 27
Zone styles
applying, 94
Auto Detect, 90, 91
creating, 96
defined, 89
deleting, 96
editing, 95
make as default, 96
selecting, 93
Zone template
applying, 75, 98
creating, 98
removing, 99
saving, 98
selecting, 75
Zone types, specifying, 43
Zones
adding to, 49
applying styles to, 94
connecting, 50
creating automatically, 42
deleting, 50
dividing, 50
drawing manually, 34, 46
irregular, 47
maximum allowed, 47
moving, 48
rectangular, 46
reordering, 48
reshaped, 49
resizing, 48
restricted shapes, 47
selecting, 48
specifying types, 43
subtracting from, 49
using templates, 98
Zones styles
Auto Detect, 93
Zoom tool, 57
Zooming in or out, 57