

# 「MaxReader 中日英文 OCR」

## ヘルプ

中日英文ドキュメント認識システム

5.5 版

S/N:107550-02-01-W-J-040414-03

## 著作権

© 2007–2014, NewSoft Technology Corp. All Rights Reserved.

本書の一部または全部を NewSoft Technology Corp.（以下「NewSoft」と称します）の文書による許可なく、無断でコピーまたは複製することはできません。

プログラム CD に格納されているビデオクリップおよびオーディオクリップはすべて、個人的なデモ、作成、プレゼンテーションにのみ使用することができます。これらを商用を目的とした作品に使用したり、再配布したりすることはできません。

本書の内容については、将来予告なく変更することがあります。

## 商標

IBM PC は International Business Machines Corporation の商標です。

MS Windows XP、Windows Vista、Windows 7、Windows 8 は Microsoft Corporation の商標です。

Pentium、MMX は Intel Corporation の商標です。

その他の商標の所有権は各社に帰属します。

[www.newsoftinc.com](http://www.newsoftinc.com)

[www.newsoft.com.tw](http://www.newsoft.com.tw)

[www.newsoft.co.jp](http://www.newsoft.co.jp)

[www.newsoft.eu.com](http://www.newsoft.eu.com)

[www.newsoft.com.cn](http://www.newsoft.com.cn)

# 目次

<b>第 1 章 はじめに</b>	<b>1</b>
各種機能.....	1
動作条件.....	4
ハードウェアおよびデバイス .....	4
オペレーティングシステム .....	4
プログラムのインストール.....	4
<b>第 2 章 基本コンセプト</b>	<b>5</b>
操作フローチャート.....	6
画面の説明.....	6
画像モード.....	7
メニューバー.....	8
サムネイル表示／ステータスバー エリア .....	9
コマンドバー.....	10
表示ツールバー .....	10
編集ツールバー .....	11
ページコントロール .....	12
文書校正ウィンドウ .....	12
ページモード .....	13
文書校正ウィンドウ .....	14
文字の校正 .....	14
編集ツールバー .....	15
比較モード.....	16
編集モード.....	17
ブロックメニュー.....	18

文書校正ウィンドウ .....	18
候補文字ウィンドウ .....	19
文字画像ウィンドウ .....	19
編集ツールバー .....	19
<b>ツールバーの変更 .....</b>	<b>20</b>
<b>拡大／縮小 .....</b>	<b>20</b>
<b>ドキュメント情報の検索 .....</b>	<b>21</b>
<b>システム設定 .....</b>	<b>21</b>

### **第3章 画像の取り込み** **24**

<b>スキャンまたはファイルを開く .....</b>	<b>24</b>
ドキュメントのスキャン .....	25
画像ファイルを開く .....	25
プロジェクトを開く .....	26
ワードプロセッサから起動 .....	27
<b>画像の処理 .....</b>	<b>28</b>
回転 .....	28
イメージのゆがみ補正 .....	30
ページの方法自動検索 .....	30
消しゴムと塗りつぶし .....	30
抜き取り .....	30
反転 .....	30
<b>スクリーンキャプチャ .....</b>	<b>31</b>
<b>ページの削除 .....</b>	<b>32</b>

### **第4章 ページレイアウトと認識** **33**

認識ライブラリ設定.....	33
ブロックの選択.....	34
ブロックの自動選択.....	34
ブロックの手動選択.....	35
ページレイアウトの設定.....	35
フィールド設定.....	36
配列設定.....	36
テーブル設定.....	36
データ設定.....	37
ブロック設定の変更.....	37
ブロックの順番を変える.....	37
ブロック設定の変更.....	38
ブロックの削除.....	38
ページを開く.....	38
ドキュメントの認識.....	39
校正ライブラリの設定.....	41
校正ライブラリの設定.....	41
校正ライブラリの編集.....	42
自動処理.....	42
作業のキャンセル.....	44

校正と編集 .....	45
部分再認識 .....	47
消しゴム .....	48
ペイント .....	48
文字の分割再認識 .....	49
文字の結合再認識 .....	49
行の分割再認識 .....	49
行の結合再認識 .....	50
ブロックの再認識 .....	50
ブロックの結合再認識 .....	51
ブロックの分割再認識 .....	51
クイックインプット .....	52
登録文字の学習 .....	53
登録文字の削除 .....	54
学習文字ライブラリの設定 .....	54

## 第 6 章 ファイルのエクスポート 56

画像の保存 .....	56
認識結果の保存 .....	57
プロジェクトの保存 .....	60
頻繁に使うページレイアウトの保存 .....	61
印刷 .....	62
他のプログラムに送信 .....	62
自動操作エクスポート設定 .....	65

## 第 7 章 日本語認識サンプル 66

## 第 8 章 英語の認識サンプル 71

## 第 9 章 フォームの認識サンプル 76

フォームの認識.....	76
罫線非表示図表.....	81

<b>第 10 章 自動認識のサンプル</b>	<b>89</b>
-------------------------	-----------

<b>付録A 用語集</b>	<b>92</b>
----------------	-----------

<b>付録B メニューコマンド</b>	<b>94</b>
---------------------	-----------

<b>付録C コマンドツールバー</b>	<b>102</b>
----------------------	------------

<b>付録D 編集ツールバー</b>	<b>106</b>
--------------------	------------

<b>付録E スキャンヒント</b>	<b>109</b>
--------------------	------------

図 1 : スキャン解像度の提案 .....	110
------------------------	-----

図 2 : 標準.....	111
---------------	-----

図 3 : 暗すぎる場合 .....	111
--------------------	-----

図 4 : 明るすぎる場合 .....	112
---------------------	-----

## 第 1 章 はじめに

「MaxReader 中日英文 OCR」は画像ファイルをすばやく、簡単に編集可能なテキストファイルに変換するためのプログラムです。「MaxReader 中日英文 OCR」では TIF (G3、G4、PackBits、複数ページ TIF)、PCX、BMP、JPEG などの画像ファイルをインポートすることができます。このプログラムは中国語 (簡体字、繁体字)、日本語、英語、数字、各種書き込み用紙を識別します。認識したテキストファイルは認識前の画像ファイルよりも少ないメモリしか使用しません。認識したファイルは TXT、RTF、DOC、XLS、SLK、CVS などの形式で保存することができます。また保存したファイルは通常のワープロソフトで開き、編集することができます。また HTML 形式でファイルを保存すると Web ブラウザで開くことができます。

- [各種機能](#)
- [動作条件](#)
- [プログラムのインストール](#)

### 各種機能

- 自動処理：このプログラムはすばやく OCR プロジェクトを完了できるように、4 種類の独立した自動処理機能が備えられています。画像ファイルのインポートからファイルの分析、ドキュメントの認識、OCR 結果をテキストファイルにエクスポートするまでを自動で行います。
- 高速処理：通常の PC (例：Pentium III 667) で 1 秒間に漢字 150 文字を認識することができます。
- 「MaxReader 中日英文 OCR」は複数の OCR エンジンを使い、中国語 (簡体字、繁体字)、日本語、英語、デンマーク語、オランダ語フィンランド語、フランス語、ドイツ語、イタリア語、ノルウェー語(ブークモール)、ノルウェー語 (ニーノシク)、ポルトガル語、スペイン語、スウェーデン語などの書類の認識をおこなえます。
- 複数ページのドキュメントを処理：最高 200 ページまでのドキュメントを処理し、認識します。



- 白黒およびカラーのドキュメントを認識：「MaxReader 中日英文 OCR」では白黒、カラーのいずれのドキュメントでも最高の認識結果が得られます。
- 複数スタイルの認識と再現：「MaxReader 中日英文 OCR」では明朝体、ゴシック体、楷書体、丸ゴシック体、など多数の印刷フォントを識別することができます。認識後も元のドキュメントと同じフォントで再現されます。
- オリジナルのフォーマットに再現：認識結果はオリジナルファイルと同じページフォーマットで表示されますので、簡単に編集できるだけでなく、フォーマットを調整する手間を省くことができます。
- 画像とテキストを自動分割：認識結果から画像とテキストを自動的に分割し、MS Word などの通常のワープロソフトで編集できるように結果を TXT、DOC、RTF などの形式で保存します。
- フォーム認識：あらゆるフォーム画像を認識し、MS Excel で処理できるようにその結果を XLS、SLK、CSV などの形式で保存することができます。またファイルを RTF 形式で保存すると、MS Word などのワープロソフトを使ってオリジナルのフォーマットでフォームを表示することができます。
- 複数のファイル形式：認識結果は TXT、RTF、DOC、XLS、SLK、CSV、HTML、PDF など異なるファイル形式で保存することができます。
- 直接エクスポート：認識結果をあらかじめ指定したアプリケーションに直接エクスポートすることができます。たとえば、結果を HTML 形式に変換して Web ブラウザに転送したり、PDF 形式に変換して Acrobat Reader に送信したりすることが可能です。
- 画面のキャプチャ：フィールドキャプチャ、または Window のキャプチャなどが可能です。
- 傾斜角度の自動検出：画像の傾斜角度を自動的に検出し、画像を回転させます。
- ゆがんだイメージの自動補正：MaxReader 5 はゆがんだイメージの自動補正をおこなうことにより、認識度を向上させます。
- MaxReader をワードから使用可能。ツールバーにワードプログラ

ムのボタンが追加されました。

- 精度の自己確認：「MaxReader 中日英文 OCR」が認識結果を内蔵されている辞書に照らし合わせてチェックを行います。これによって「疑わしい」文字を表示しますので、校正がしやすくなります。
- 候補文字：似たような文字、同じような意味の文字、または前後の文字に関連した文字を候補文字として表示しますので、簡単に誤認識された文字を修正することができます。
- 学習文字：間違えやすい文字を新しく学習した文字として記憶させ、認識精度を高めます。
- 情報交換：日本語版 Windows や他の Windows アプリケーションのクリップボードを使用し、画像をインポートしたり、テキストをエクスポートしたりすることができます。
- 横書き／縦書きおよびシングル／複数カラムの認識：縦書き、横書き、シングル／複数カラムが自動的に分析されます。インポートした画像ページの設定が複雑すぎる場合は、システムが正しく分析、認識できるようにテキスト／画像フォーマットを設定することもできます。
- テンプレートの保存：特定のページ設定をテンプレートとして保存することができます。テンプレートは「MaxReader 中日英文 OCR」ページ設定ファイル (\*.tpl) として保存されます。保存したテンプレートを使って新しくファイルをインポートすると、分析時間を短縮化することができます。
- ドキュメント校正：「MaxReader 中日英文 OCR」ではキーボード、候補文字、ユーザー辞書を使ってオリジナル画像／テキストと認識結果を並べて比較し、誤認識された文字を校正することができます。
- 再認識：文字、列、ブロックを結合／分離させ、ブロックのプロパティを変更することができます。これによって、誤認識された文字を校正したり、オリジナルのドキュメントと比較したりすることができます。

## 動作条件

### ハードウェアおよびデバイス

- CPU およびメモリは、ご使用の OS の最低動作条件を参照。詳細は、Microsoft Web サイトの関連項目をご参照ください。
- Windows XP: Pentium III 800 MHz CPU またはそれ以上  
Windows Vista, Windows 7, Windows 8: Pentium 4, 3.0GHz CPU  
またはそれ以上
- RAM : 512MB またはそれ以上
- モニター : 800x600 ハイカラー解像度
- ハードディスクに 400MB の空き領域
- CR-ROM ドライブ
- TWAIN 対応スキャナ（詳細はスキャナの取り扱い説明書をご参照ください。）

### オペレーティングシステム

- Microsoft Windows XP, Windows Vista, Windows 7, Windows 8

## プログラムのインストール

1. 「MaxReader 中日英文 OCR」インストール CD を CD-ROM ドライブに挿入します。
2. 画面の指示にしたがってインストールを行ってください。
3. インストールが完了したら、[スタート]—[プログラム]—「MaxReader 中日英文 OCR」—「MaxReader 中日英文 OCR」を選択してプログラムを起動します。またデスクトップ上で「MaxReader 中日英文 OCR」アイコンをダブルクリックしてもプログラムを起動することができます。



インストール時には、以下の点にご注意ください。

- (1) インストールと操作にはパワーユーザーまたは管理者の権限が必要です。
- (2) インストール時のエラーを防止するために、インストールを実行する前に OS をデフォルトの言語に戻してください。

## 第2章 基本コンセプト

この章では「MaxReader 中日英文 OCR」の操作手順とインターフェース、ディスプレイレイアウトの変更、画像表示倍率の変更、ファイル情報の検索、別のページへの移動、システム設定の変更、オンラインヘルプの使い方など、「MaxReader 中日英文 OCR」の基本コンセプトと、「MaxReader 中日英文 OCR」を操作する手順について説明します。

- [操作フローチャート](#)
- [画面の説明](#)
- [ツールバーの変更](#)
- [拡大／縮小](#)
- [ドキュメント情報の検索](#)
- [システム設定](#)

## 操作フローチャート

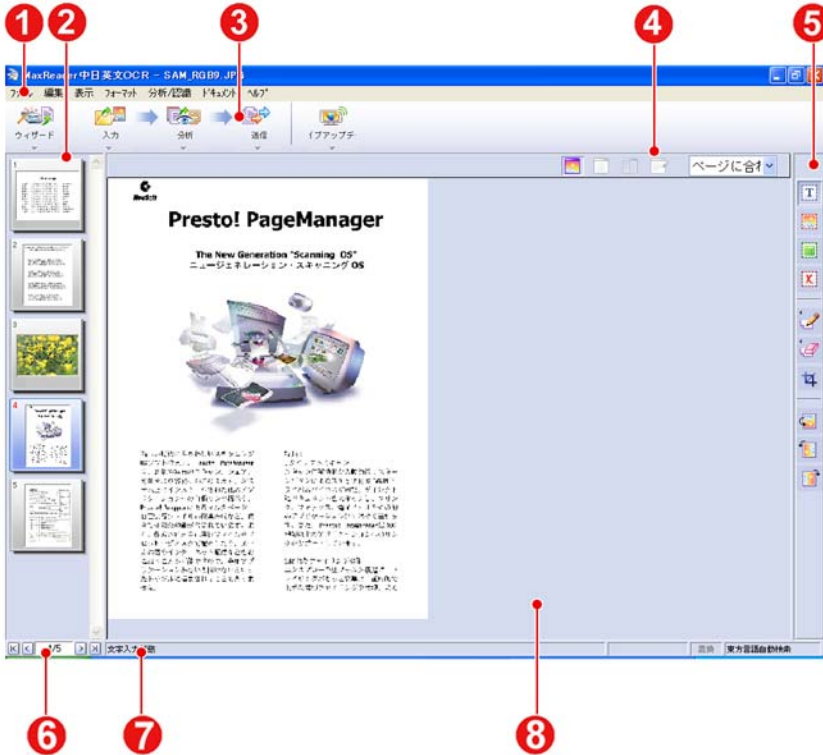


## 画面の説明

「MaxReader中日英文OCR」は次の4種類のモードで作業を行うことができます。[画像モード](#)、[ページモード](#)、[比較モード](#)、[編集モード](#)。

## 画像モード

画像モードでは認識前に必要なすべての作業を行うことができます。たとえばスキャナや CD-ROM ドライブから画像ファイルをインポートした後は、編集ツールで画像を調整し、認識設定を行うことができます（認識言語の設定、テキスト配置、テーブルの挿入など）。



### 画像モード

- |                                |                               |
|--------------------------------|-------------------------------|
| 1. <a href="#">メニューバー</a>      | 2. <a href="#">サムネイル表示エリア</a> |
| 3. <a href="#">コマンドバー</a>      | 4. <a href="#">表示ツールバー</a>    |
| 5. <a href="#">編集ツールバー</a>     | 6. <a href="#">ページコントロール</a>  |
| 7. <a href="#">ステータスバー エリア</a> | 8. <a href="#">文書校正ウィンドウ</a>  |

## メニューバー

メニューバーにはファイルを開く、保存、編集、認識などのすべてのコマンドがリストされます。



ファイル 編集 表示 フォーマット 分析/認識 ドキュメント ヘルプ\*

特定のコマンドメニューを開くには、そのコマンドをクリックしてください。

特定のコマンドを実行するには、次のいずれかの操作を行ってください。

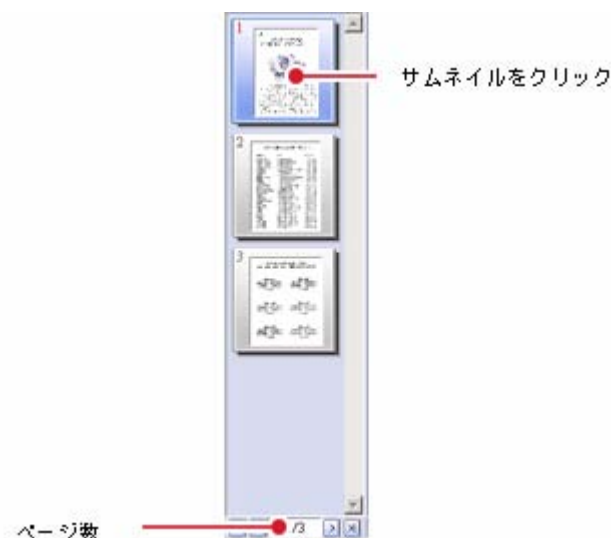
- コマンドメニューを開き、カーソルをコマンドに合わせてクリックします。
- コマンドの中にはショートカットを持つものがありますので、コマンドメニューを開かずにそのコマンドを実行することができます。ショートカットとは各コマンドの右側に表示される複数のキーを組み合わせたものです。

開いているコマンドメニューを閉じるには、次のいずれかの操作を行ってください。

- コマンドメニューの外側をどこでもクリックします。
- Esc キーを押します。

## サムネイル表示／ステータスバー エリア

サムネイル表示ではインポートしたファイルをサムネイルで表示しますので、サムネイルをクリックすると直接そのページに移動することができます。また下のステータスバー エリアをクリックすると、任意のページに移動することができます。




キー	機能
	先頭ページ
	最終ページ
	前のページ
	次のページ

またサムネイル表示でドキュメントを右クリックすると、[現在開いているページを削除する]や[現在開いているページの OCR の結果をすべて破棄する]など、頻繁に使用されるコマンドのいくつかを選択することができます。



## コマンドバー

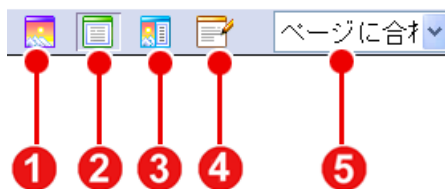
コマンドバー上のアイコンをクリックすると、多くのコマンドをすばやく簡単に実行することができます。各コマンドアイコンの下にはプルダウンメニュー  があります。通常プルダウンメニューからコマンドを選択すると、コマンドアイコンが変わります。中には選択してもアイコンが変わらないものもあります。コマンドのアイコンをクリックすると、相当するコマンドが実行されます。



クリックするとプルダウンメニューが表示されます

## 表示ツールバー

ツールバーを表示しておくことで、画像表示倍率を拡大／縮小したり、画像モード、ページモード、編集モードに切り替えたりすることができます。

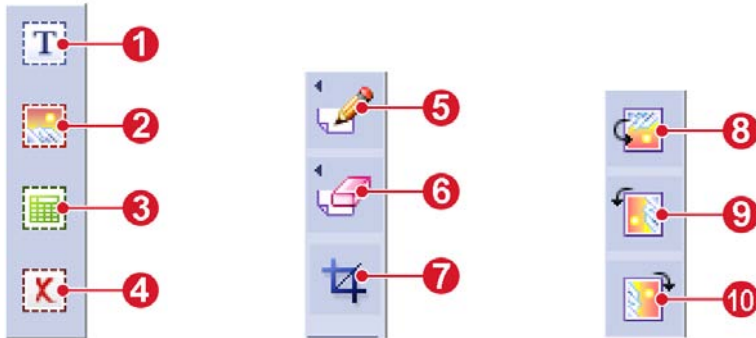


1. [画像モード](#)
2. [ページモード](#)
3. [比較モード](#)
4. [編集モード](#)
5. [拡大／縮小](#)

## 編集ツールバー

編集ツールバーからは、インポートしたスキャン済み画像ファイルを編集したり、ドキュメント内のブロックを変更したりといった、認識段階が異なるドキュメントを処理することができます。編集ツールバー上のアイテムは認識段階によって変わります。詳細は[付録D – 編集ツールバー](#)を参照してください。

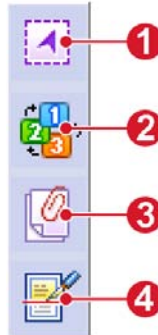
## 認識前ツールバー



画像モードの編集ツールバー（OCR 前）

- |                                |                                |
|--------------------------------|--------------------------------|
| 1. <a href="#">テキストの選択</a>     | 2. <a href="#">画像の選択</a>       |
| 3. <a href="#">表の選択</a>        | 4. <a href="#">削除</a>          |
| 5. <a href="#">ペン</a>          | 6. <a href="#">消しゴム</a>        |
| 7. <a href="#">抜き取り</a>        | 8. <a href="#">180度回転</a>      |
| 9. <a href="#">反時計廻りに90度回転</a> | 10. <a href="#">時計廻りに90度回転</a> |

## 認識後ツールバー



画像モードの編集ツールバー（認識後）

1. [ブロックをマーク](#)
2. [ブロックの順番](#)
3. [ブロックの結合](#)
4. [ブロックの分割](#)

## ページコントロール

ページコントロールには次のような情報が表示されます（左から右へ）：  
 オブジェクトの情報、カーソルの X 軸と Y 軸、現在選択されているテキスト  
 入力モード、認識ライブラリ。ページコントロールの右にある[認識ライ  
 ブラリ]をクリックすると、リストから文字を選択することができます。

マウスを動かして認識部分をドラッグして選択する

X:848, Y:2882

置換

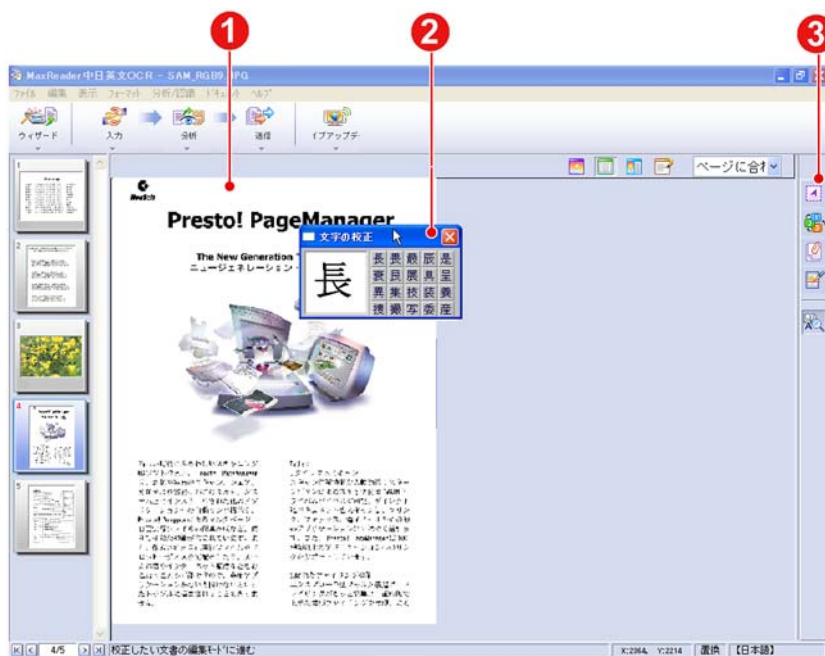
【日本語】

## 文書校正ウィンドウ

文書校正ウィンドウでは画像を回転させたり、不明瞭な部分を除去したり、  
 認識不要な部分を切り取ったりするなど、画像を処理して OCR 結果を高  
 めることができます。また文書校正ウィンドウでドキュメントを右クリッ  
 クすると、現在開いているページを認識する、拡大／縮小、180度回転な  
 ど、いくつかの頻繁に使用するコマンドを選択することができます。

## ページモード



ページモードでは認識後のドキュメント全体を表示し、ブロックの結合／分離、ブロックのプロパティの変更、ブロックの順番変更などのページ設定を調整することができます。またこのモードでは、間違って認識された単語を直接修正することができます。




ページモード

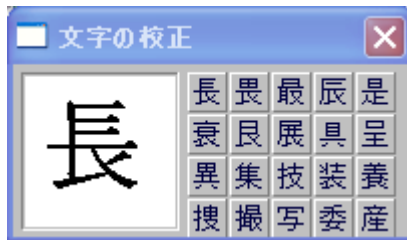
1. [文書校正ウィンドウ](#)
2. [文字の校正](#)
3. [編集ツールバー](#)

## 文書校正ウィンドウ


文字構成ウィンドウには OCR 後のドキュメント全体が表示されます。[文字の校正]ツール  をクリックすると、青い文字構成ウィンドウの背景にいくつかの文字が黄色で表示されます。これらは認識後にシステムが認識した疑わしい文字です。[ブロックの順番]ツール  をクリックすると、すべての認識部分がグレイの線で区切られ、それぞれ左上隅に番号が付けられます。

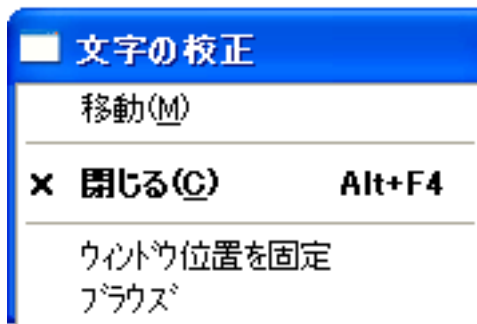
## 文字の校正

間違って認識された文字が見つかった場合は、[文字の校正]ツール  を使ってその文字をクリックすると、文書校正ウィンドウにウィンドウが開きます。



文字の校正ウィンドウ

そのウィンドウの左上隅にある  アイコンをクリックすると、文書の校正ウィンドウのプロパティを設定することができます。

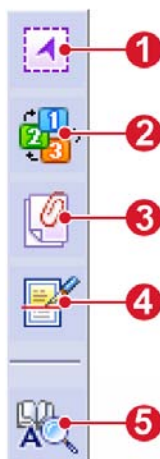


文字の校正ウィンドウのメニュー

- 移動** キーボードの矢印キー（↑↓←→）を使ってウィンドウの位置を移動することができます。
- 閉じる** 現在開いているウィンドウを閉じます。
- ウィンドウ位置を固定** ウィンドウを任意の場所にドラッグした後は、ウィンドウが動かないように、この機能を使ってウィンドウの位置をロックすることができます。
- ブラウズ** この機能を選択すると、ウィンドウ内の文字がカーソルの動きに伴って変わります。この機能を選択しなければ、ウィンドウは特定の文字の候補文字が表示されるまで変わりません。

## 編集ツールバー

編集ツールバーにはページモードで必要な校正／編集ツールがあります。詳細は[付録D－編集ツールバー](#)を参照してください。

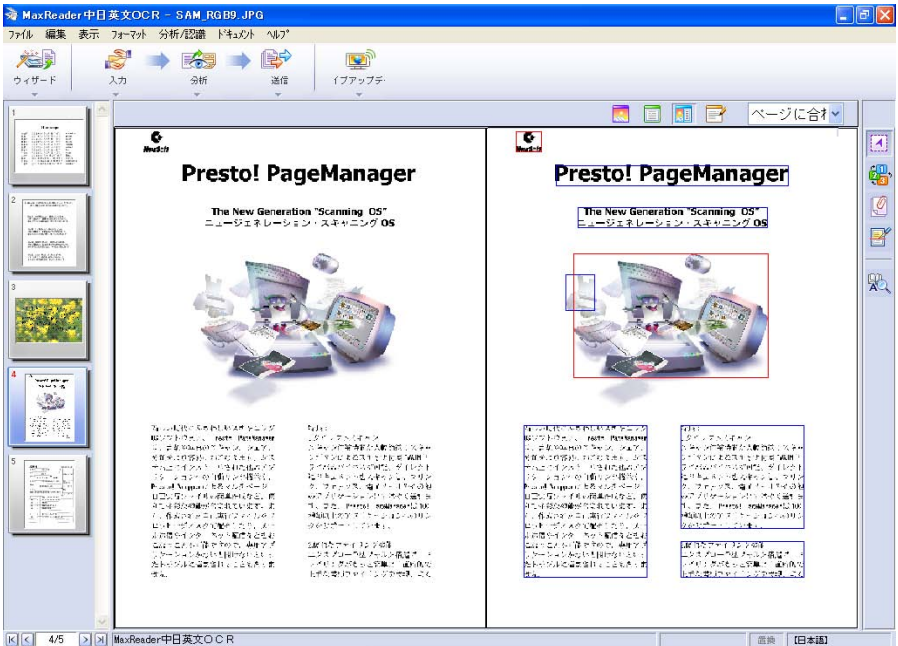


ページモードの編集ツールバー

1. [ブロックをマーク](#)
2. [ブロックの順番](#)
3. [ブロックの結合](#)
4. [ブロックの分割](#)
5. [文字の校正](#)

## 比較モード

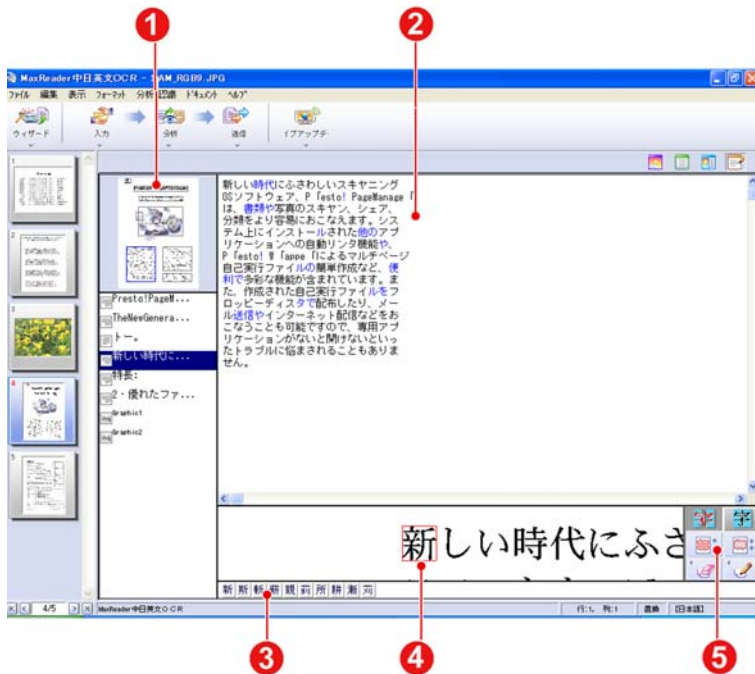
**表示ツールバー**から比較モードをクリックし、OCR処理後のイメージをオリジナルと比較します。



オリジナルイメージがイメージビューアに表示されます。イメージ内でもとのテキストを参照しながら、OCRの結果を確認できます。編集ツールを使い、誤字の訂正をおこないます。

## 編集モード

編集モードで最も重要な機能は、OCR後のドキュメントを校正する機能です。編集モードではOCR後のテキストを表示し、テキストを編集することができます。「MaxReader 中日英文 OCR」には文字の提案、テキストの分割／結合、文字の認識、編集など、さまざまなテキスト編集機能が用意されています。また正しい文字を直接入力して、間違っていて認識された文字を修正することもできます。



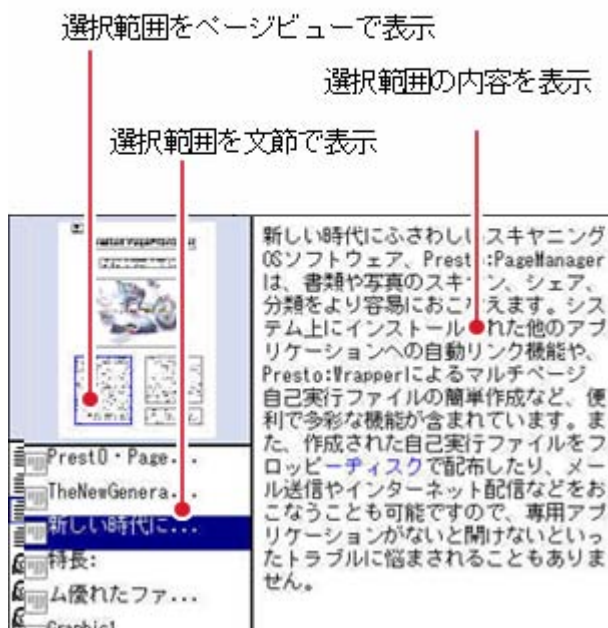
編集モード

1. [ブロックメニュー](#)
2. [文書校正ウィンドウ](#)
3. [候補文字ウィンドウ](#)
4. [文字画像ウィンドウ](#)
5. [編集ツールバー](#)



## ブロックメニュー

カーソルをブラウズしたいブロックに合わせてクリックするか、ドキュメントブラウズバーから直接任意のブロックを選択してください。ブロックの内容は文書校正ウィンドウに表示されます。



ブロックメニュー

## 文書校正ウィンドウ

文書校正ウィンドウには OCR 後のテキストが表示されます。青いテキストは認識中に見つかった疑わしい文字です。

## 候補文字ウィンドウ

候補文字ウィンドウから正しい文字を選択し、間違って認識された文字に置き換えることができます。

新 斯 新 親 莉 耕 漸 朝 執 所 初 許 祝 射 動 荒 離 禍 若 崩 流 難 苦 施 細 寂 押 卸 刺 制 濟 致 疏 汚 如 話 挿 揮 芳 純 乾

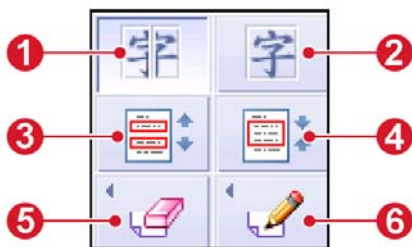
## 文字画像ウィンドウ

文字画像ウィンドウでは、ドキュメントを編集しやすいようにカーソルを合わせた文字画像を拡大することができます。

新 しい時代にふさわしいスキ

## 編集ツールバー

編集ツールバーには編集モードで必要な校正／編集ツールがあります。詳細は[付録D - 編集ツールバー](#)を参照してください。



編集モードの編集ツールバー

1. [文字の分割再認識](#)
2. [文字の結合再認識](#)
3. [行の分割再認識](#)
4. [行の結合再認識](#)
5. [消しゴム](#)
6. [ペイント](#)

## ツールバーの変更

スクリーン上のツールバーは、必要に応じて表示したり、隠したりすることができます。[表示]メニューでコマンドバー、サムネイル表示、編集ツールバーを有効または無効にして、ツールバーを変更してください。

- ✓ コマンドバーを表示
- ✓ サムネイル表示
- ✓ 編集ツールバーの表示

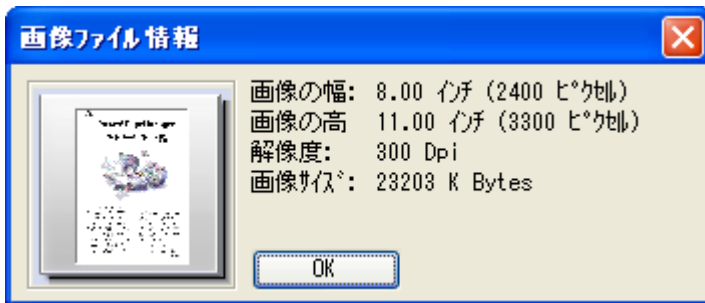
## 拡大／縮小

ウィンドウに表示されている画像は、編集しやすいように必要に応じて表示倍率を変えることができます。[表示]メニューから画面に合わせる、ページに合わせる、実サイズ、縮小表示、拡大表示、縮小、拡大を選択して、画像サイズを変更してください。また次のホットキーを使用することもできます。

ホットキー	機能
Ctrl + F5	拡大
Ctrl + F6	縮小
Ctrl + F4	実サイズ(100%)
Ctrl + F3	ページに合わせる
Ctrl + F2	画面に合わせる

## ドキュメント情報の検索

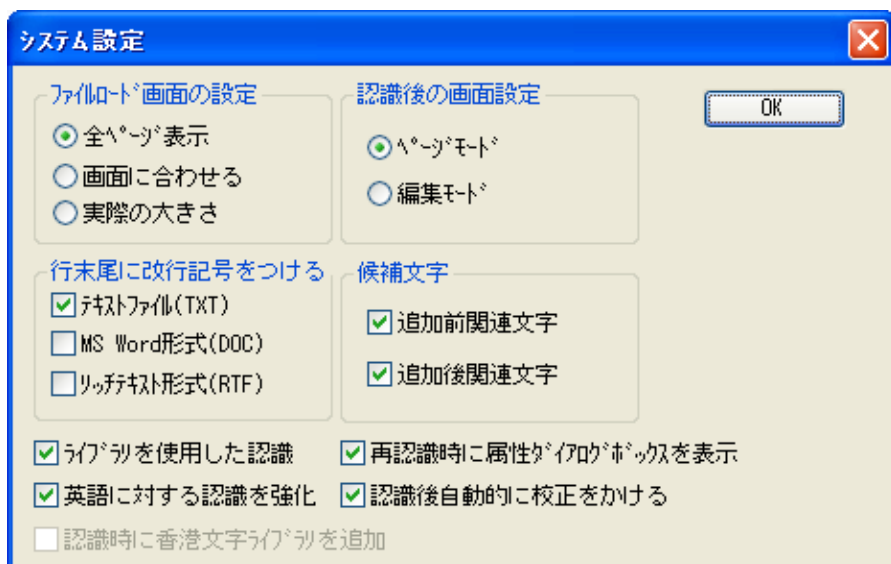
インポートした画像については、サイズや解像度などの情報を見ることができます。複数のページを編集するときには、[ドキュメント]メニューから特定のページを選択し、[ページ情報]を選択してください。または画像モードで画像を右クリックして[画像内容]を選択すると、関連する情報を見ることができます。



[ページ情報]のダイアログボックス

## システム設定

[ファイル]→[システム設定]を選択すると、システムのデフォルト設定を行うことができます。[システム設定]のダイアログボックスには次のような設定があります。



[システム設定]のダイアログボックス

### ファイルロード 画面の設定

画像を表示する倍率を[ページに合わせる]、[画面に合わせる]、[実サイズ]の中から選択してください。

### 認識後の画面設定

認識後のスクリーンの表示モードをページモードまたは編集モードに切り替えます。

### 行末尾に改行記号をつける

認識後に結果をエクスポートまたは保存するために、各行の末尾に改行記号を付けるかどうかを選択します。

### 候補文字

前後の文字を考慮して、候補文字が表示されます。

**追加前関連文字**：選択した文字の1つ前の文字を考慮して、関連する文字を表示します。

**追加後関連文字**：選択した文字の1つ後の文字を考慮して、関連する文字を表示します。

### ライブラリを使用した認識

認識精度を高めるために、ライブラリを使って認識を行います。

**英語に対する認識を強化**

中国語／英語のドキュメントを認識するときにスペルチェックを行います。

**認識時に香港文字ライブラリを追加**

ライブラリに香港文字を追加します。

**再認識時に属性ダイアログボックスを表示**

再認識を行う前に、関連する設定を変更するために[属性]ダイアログボックスを表示します。

**認識後自動的に校正をかける**

認識時にシステムに内蔵されている辞書を使って校正を行います。

## 第3章 画像の取り込み

認識が必要なすべての画像ファイルは、まず「MaxReader 中日英文 OCR」にインポートする必要があります。したがって、まず画像ファイルを「MaxReader 中日英文 OCR」にインポートする必要があります。また未完成のファイルを開いて編集したり、認識したりすることもできます。「MaxReader 中日英文 OCR」では一度に200ページまでインポートし、認識することができます。

- [スキャンまたはファイルを開く](#)
- [画像の処理](#)
- [スクリーンキャプチャ](#)
- [ページの削除](#)

### スキャンまたはファイルを開く

認識が必要な画像ファイルは、そのほとんどがスキャナでスキャンしたものです。「MaxReader 中日英文 OCR」システムで使用するスキャナをあらかじめ設定しておき、画像ファイルとしてインポートするドキュメントをスキャンすることができます。また既存の画像ファイルや未完成のプロジェクトを開いて編集したり、認識したりすることもできます。

「MaxReader 中日英文 OCR」では TIF (G3、G4、PackBits、複数ページ TIF)、PCX、BMP、JPEG などの画像ファイルを認識することができます。



- [ドキュメントのスキャン](#)
- [画像ファイルを開く](#)
- [プロジェクトを開く](#)
- [ワードプロセッサから起動](#)

## ドキュメントのスキャン



### ステップ1：スキャナの設定

1. [ファイル]-[ソースの選択]を選択します。
2. [ソースの選択]のダイアログボックスでスキャナドライバを選択し、[選択]をクリックします。

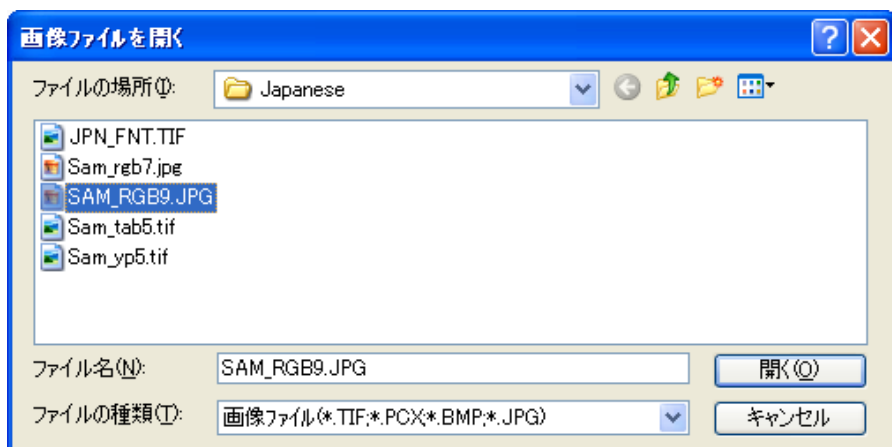
### ステップ2：スキャン

1. 次のいずれの方法でドキュメントをスキャンしてください。
  - コマンドバー上の2つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[スキャン]を選択します。
  - メニューバーから[ファイル]-[スキャン]を選択します。
2. スキャン画面が開きますので、[スキャン]をクリックして画像をインポートします（スキャン画面についての詳細は、スキャナの取り扱い説明書を参照してください）。

## 画像ファイルを開く

1. 次のいずれかの方法で画像ファイルを開いてください。
  - コマンドバー上の2つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[画像を開く]を選択します。
  - メニューバーから[ファイル]-[画像を開く]を選択します。







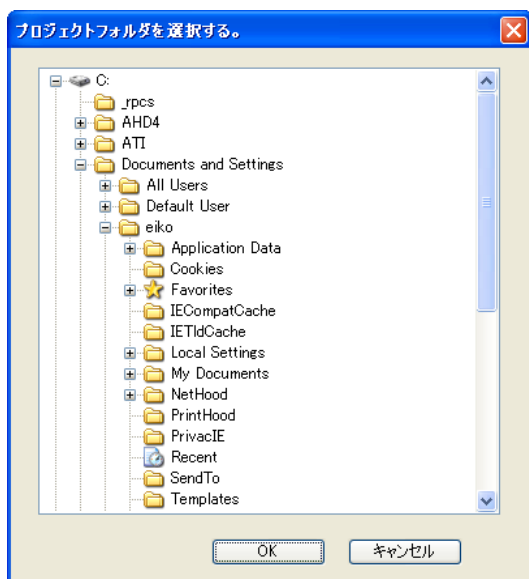
[画像を開く]のダイアログボックス

2. [画像を開く]のダイアログボックスで画像ファイルを選択し、[開く]をクリックすると選択した画像がインポートされます。また Shift キーや Ctrl キーを使うと、一度に複数のファイルを選択することができます。

## プロジェクトを開く

すでに保存されているプロジェクトがある場合は、次のいずれかの方法でそれを開くことができます。プロジェクトを保存する方法については、第6章- [プロジェクトの保存](#) を参照してください。

- コマンドバー上の2つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[プロジェクトを開く]を選択します。すると[プロジェクトフォルダを選択する]のダイアログボックスが開きます。
- メニューバーから[ファイル]-[プロジェクトを開く]を選択します。



[プロジェクトフォルダを選択する]のダイアログボックス

プロジェクトが保存されているフォルダを選択したら、[OK]をクリックしてください。ファイルがインポートされます。

## ワードプロセッサから起動

MaxReaderのインストールすると、ご使用のワードプロセッサにアイコンとして追加されます。ツールバーからこのボタンをクリックすることにより、直接ワードからMaxReaderを起動できます。

1. ワードプログラムを起動します。
2. MaxReaderのアイコンをクリックし、プログラムを起動します。
3. ファイルメニューをクリックし、認識するイメージファイルを開きます。ダイアログボックスで認識に使用する言語を選択、またはシステムのデフォルト言語を使用します。OKをクリックすると、テキストの認識がおこなわれ、ワードプロセッサで開かれます。
4. **MaxReader**のダイアログボックスを終了します。

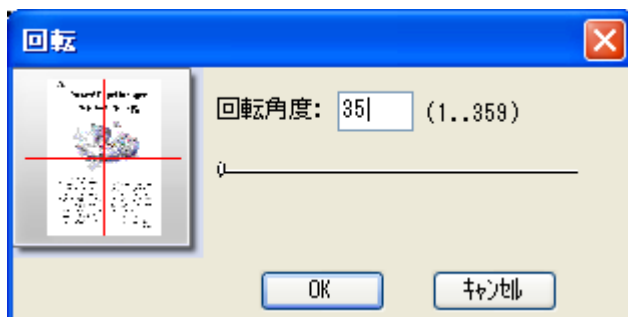
## 画像の処理

最高の認識結果を得るためには、明瞭な画像ファイルを用意することが鍵になります。したがって認識を行う前に、インポートした画像をチェックしてから処理を行ってください。たとえば、歪んだドキュメントを回転させたり、黒い点を消したり、白いスポットを消したり、認識する必要がない部分を切り抜いたり、画像を反転させたりして、正しく認識されるように準備を行ってください。

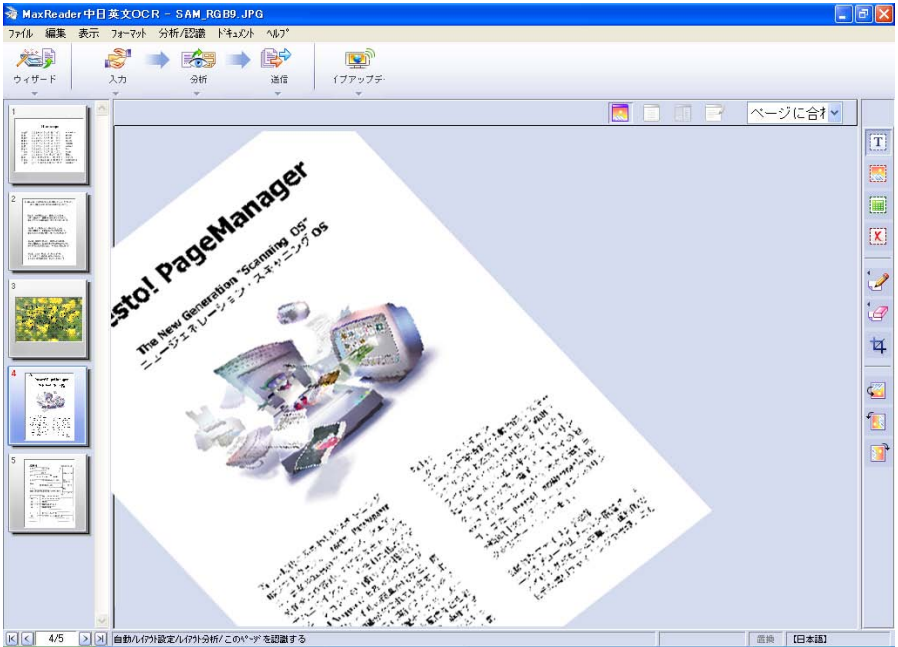
- [回転](#)
- [イメージのゆがみ補正](#)
- [ページの方向自動検索](#)
- [消しゴムと塗りつぶし](#)
- [抜き取り](#)
- [反転](#)

## 回転



- 画像の傾き具合が3度以内の場合：回転させる必要はありません。
- 画像の傾き具合が3度以上10度以内の場合：[編集]-[回転]-[角度指定で回転]を選択するか、画像モードに切り替えてドキュメントを右クリックし、[角度指定で回転]を選択します。すると次のダイアログボックスが開きますので、自動的に角度が調整されます。



- 画像の傾き具合が10度以上90度以内の場合：スキャンした画像が下の図のようになっている場合は、もう一度画像をスキャンしてください。



大きく傾いた画像

- 画像が90度傾いている場合：メニューバーから[編集]-[回転]-[(反)時計回りに90度回転]を選択するか、ツールボックスで  をクリックして画像を回転させます。また画像モードに切り替えてドキュメントを右クリックし、[(反)時計回りに90度回転]を選択する方法もあります。
- 画像が180度傾いている場合：メニューバーから[編集]-[回転]-[180度回転]を選択するか、ツールボックスで  をクリックして画像を回転させます。また画像モードに切り替えてドキュメントを右クリックし、[180度回転]を選択する方法もあります。



## イメージのゆがみ補正

スキャンしたイメージが斜めになっている場合、OCRの精度に影響します。このような場合、[編集]—[全ファイルのゆがみを自動補正]を選択すると、MaxReaderで自動的に補正をおこなえます。このオプションを選択すると、毎回OCRをおこなう前に、自動補正がおこなわれます。


## ページの方向自動検索

ページの向きをただしく補正することにより、OCRの精度を向上させることができます。[編集]—[全ファイルの方向を自動補正]を選択すると、MaxReaderで自動的に補正をおこなえます。このオプションは、ツールバーのステップにそって作業をおこなうことにより、スキップ可能です。

## 消しゴムと塗りつぶし

認識精度を高めるには、編集ツールバーの[消しゴム]  を使用してブロックの近くにある黒い点を消すか、[ペン]  を使って白いスポットを塗りつぶしてください。

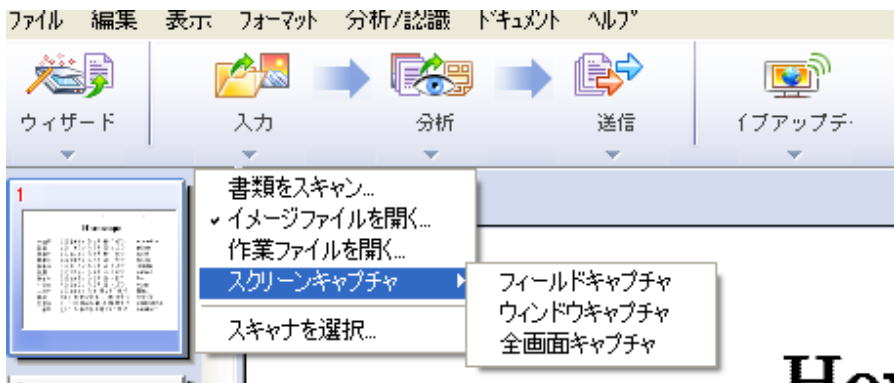
## 抜き取り



インポートした画像をすべて認識する必要がないときには、編集ツールバーから  を選択して必要な部分を選択してください。次に[編集]—[切り取り]を選択して不要な部分を切り取ります。

## 反転

「MaxReader 中日英文 OCR」では黒い背景に白い文字で書かれた画像を認識することはできません。このような画像をインポートした場合は、認識する前に[編集]—[反転]を選択して背景を白に、文字を黒に変換してください。

## スクリーンキャプチャ



MaxReaderでは画面から直接イメージをキャプチャし、認識をおこなうことができます。ツールバーで  の下の  をクリックし、リストから[スクリーンキャプチャ]を選択します。さらにサブメニューから、[フィールドキャプチャ]、[ウインドウキャプチャ]、[全画面キャプチャ]のいずれかを選択します。各オプションの詳細は、以下を参照ください。

**フィールドキャプチャ:** 画面の一部の認識をおこなう場合、このオプションを選択します。画面上でアイコンをドラッグして、選択した部分をキャプチャし、キャプチャされたイメージを MaxReader で開きます。

**ウインドウキャプチャ:** ウィンドウの一部をキャプチャする場合、カーソルをキャプチャしたい部分に移動します。選択された部分に外枠が表示されます。マウスをクリックして、キャプチャをおこない、キャプチャしたイメージを MaxReader で開きます。

**全画面キャプチャ:** マウスをクリックして、画面全体をキャプチャし、キャプチャしたイメージを MaxReader で開きます。

## ページの削除

「MaxReader 中日英文 OCR」から不要なページを削除するには、次の手順にしたがってください。

1. サムネイル表示から削除したいページを選択します。
2. 次のいずれかの方法でページを削除してください。
  - [ファイル]–[ページの削除]を選択します。
  - サムネイル表示でページを右クリックし、[ページの削除]を選択します。

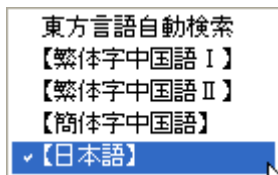
## 第4章 ページレイアウトと認識

OCR機能を実行する前に、効率よく、正確に認識できるように認識ライブラリの設定、ブロックの選択、ページレイアウトの設定、ページ分析、校正ライブラリの選択などを行い、認識の準備をしてください。またインポートから認識までを自動で行うこともできます。

- [認識ライブラリ設定](#)
- [ドキュメントの認識](#)
- [ブロックの選択](#)
- [校正ライブラリの設定](#)
- [ページレイアウトの設定](#)
- [自動処理](#)
- [ブロック設定の変更](#)
- [作業のキャンセル](#)

### 認識ライブラリ設定

ドキュメントに中国漢字が含まれている場合は、[フォーマット]-[認識ライブラリ設定]を選択して適切なライブラリを設定してください。



**繁体字中国語Ⅰ** 中国語の新聞や雑誌などを認識するのに適しています。

**繁体字中国語Ⅱ** 漢詩や古文書など、古い中国文字が含まれたドキュメントを認識するのに適しています。この辞書を使用すると、繁体字常用文字を使用する場合を比較して認識速度が遅くなります。



簡体字中国語	中国語簡体字が含まれたドキュメントを認識するのに適しています。
日本語	日本語文字が含まれたドキュメントを認識するのに適しています。



## ブロックの選択

ページ全体を認識するときにはブロックを選択する必要はありませんが、ページの一部だけを認識したい場合はその部分を手動で選択しなければなりません。ページレイアウトや認識などの機能を実行すると、選択したブロックだけが処理されます。

- [ブロックの自動選択](#)
- [ブロックの手動選択](#)




## ブロックの自動選択

[ページ分析]を実行すると画像とテキストが分割されますので、システムが効率よく作業を行えるように認識したいブロックを選択し、認識する順番を指定することができます。次のいずれかの方法で自動的にページ分析を行ってください。

- [分析/認識]–[ページ分析]を選択します。
- コマンドバー上の3つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[ページ分析]を選択します。

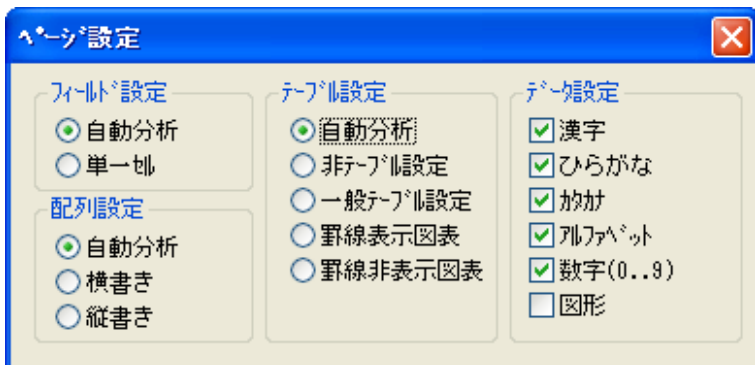
ページ分析を行った後は、システムが選択したブロックが表示されず。

## ブロックの手動選択

1. 編集ツールバーで[テキストの選択]  をクリックします。
2. マウスをドラッグして認識したいテキスト（フォーム以外）を選択します。選択したテキストは青い線で囲まれます。
3. ドキュメント内の画像を維持したい場合は、編集ツールバーの[画像の選択]  をクリックしてください。そのエリアは画像として処理され、赤い線で囲まれます。
4. ドキュメント内のフォームが含まれている場合は、編集ツールバーで[表の選択]  をクリックしてフォームを選択してください。フォームは緑の線で囲まれます。
5. 処理後のブロックを1つの完全なブロックとして維持するには、[フォーマット]—[段落を分割しない]を選択してください。ブロックを囲む枠が単線から二重線に変わります。

## ページレイアウトの設定

[ページフォーマット]では、縦書き／横書き、単一／複数セル、言語、フォーム関連の設定など、OCR ドキュメントのプロパティを設定します。[フォーマット]—[ページフォーマット]を選択し、それぞれの設定を行います。



[ページ設定]のダイアログボックス

- [フィールド設定](#)
- [配列設定](#)
- [テーブル設定](#)
- [データ設定](#)

## フィールド設定

[フィールド設定]はデフォルトで[自動分析]に設定されています。この設定では、画像ファイルからテキストカラムの設定が自動的に検出されます。次のような場合は、ドキュメントを強制的に単一セルとして設定することができます。

- 箇条書きされたドキュメントの場合は、認識前に[単一セル]を選択してください。
- 画像ファイルから複数セルのテキストを単一セルテキストとして保存するには、認識前に[単一セル]を選択してください。

## 配列設定

[配列設定]はデフォルトで[自動分析]に設定されています。この設定では、画像ファイルから配列設定が自動的に検出されます。特別な配列設定が必要な場合は、正しくページ分析されるように[縦書き]か[横書き]を選択することができます。[縦書き]を選択すると、縦書きテキストとして認識されます。同様に[横書き]を選択すると、横書きテキストとして認識されます。つまり選択した配列設定によって、認識結果が変わるということです。

## テーブル設定

[テーブル設定]はデフォルトで[自動分析]に設定されています。つまり画像ファイルにフォームまたはフォーム関連の属性が含まれているかどうか自動的に検出されます。ドキュメントにフォームが含まれていない場合は、[非テーブル書式]に設定することができます。ドキュメントにフォームが含まれている場合は、フォームの属性に応じて[一般テーブル書式]、[罫線表示図表]、[罫線非表示図表]を選択することができます。

## データ設定


日本語、英語、または日本語、英語まじりのドキュメントなど、画像ドキュメントの属性を設定することができます。また数字や画像が含まれているかどうかを設定することが可能です。

## ブロック設定の変更


ブロックを選択した後、必要であればブロックの順番や属性を変更することができます。また不要なブロックがあれば、それを削除することも可能です。

- [ブロックの順番を変える](#)
- [ブロック設定の変更](#)
- [ブロックの削除](#)
- [ページを開く](#)

## ブロックの順番を変える

[ページ分析]を行った後のブロックには、数字が付けられます。システムにページを分析させる場合は、テキストが縦書きか横書きかによってブロックの順番が自動的に決定されます。手動でブロックを選択すると、ブロックを選択した順番で認識が行われます。編集ツールバーで[ブロックの順番]  を選択すると、各ブロックの左上隅に番号が付けられます。ブロックの順番は、認識後のテキストの出力内容に影響します。

ブロックの順番を変えるには：

1. 編集ツールバーで[ブロックの順番]  を選択します。
2. ブロックをクリックした後、任意の場所にドラッグして順番を変えます。

す。たとえばブロック3を2に変えたい場合は、ブロック3をクリックしてブロック2までドラッグし、マウスを離すと左上隅の順番が変わります。


3. また必要に応じて各ブロックの順番を設定することもできます。


## ブロック設定の変更

手動でブロックを選択した後、テキストブロックを画像ブロックに変更するなど、ブロックの属性を変更したい場合は、そのブロックを右クリックして[段落の属性を変更する]を選択し、次の中から任意の属性を選択してください。イメージ、テキスト、属性の応用設定。[属性の応用設定]を選択すると、[ページレイアウト]のダイアログボックスが表示されます。詳しくは、第4章 - [ページレイアウトの設定](#)を参照してください。

## ブロックの削除

不要なブロックは削除することができます。編集ツールバーの状態によって、次の2つのソリューションから選択することができます。

ソリューション1：[削除]  を選択し、削除したいブロックを選択します。

ソリューション2：画像モードに切り替えて[ブロックをマーク]  を選択し、削除するブロックを選択します。次に右クリックして[ブロックを削除]を選択してください。

## ページを開く

頻繁に使用するページレイアウトを保存しておき、ドキュメントに直接適用すると、ページ分析処理を短縮化することができます。ページレイアウトの保存については、第6章 - [頻繁に使用するページレイアウトの保存](#)を参照してください。

既存のページレイアウトを使用するには、[フォーマット]-[ページを開く]を選択します。

## ドキュメントの認識

画像ファイルのインポート、認識ライブラリの設定、ページレイアウトの設定、ブロックの並べ替えを行った後は、設定に基づいてシステムがドキュメントの認識を開始します。



認識が完了すると、システムはページモードか編集モードに入りますので認識処理したテキストドキュメントを校正してください。認識後の表示モードを設定するには、[ファイル]–[システム設定]を選択します。

定期的に特定の種類のドキュメントを認識する場合は、認識設定を保存しておくで毎回設定する手間を省くことができます。

選択したページだけを認識するには、次のいずれかの手順にしたがってください。

- メニューバーから[分析/認識]–[現在開いているページを認識]を選択します。





- コマンドバー上の3つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから現在開いているページを認識]を選択します。
- 画像モードに入りサムネイル表示か文書校正ウィンドウでドキュメントを右クリックし、[現在開いているページを認識]を選択します。

すべてのページを認識するには、次のいずれかの手順にしたがってください。

- メニューバーから[分析/認識]–[すべてのページを認識]を選択します。




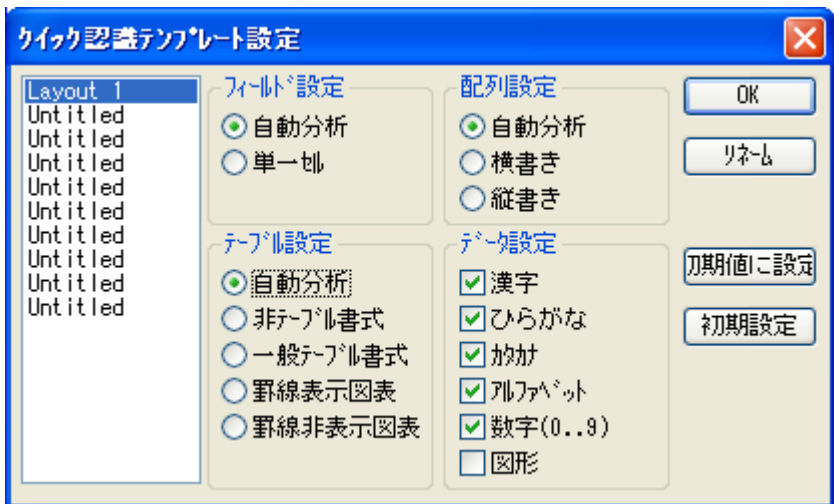
- コマンドバー上の3つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[すべてのページを認識]を選択します。

既存のブロックだけを認識するには：

まずブロックを選択し、次に右クリックして[ブロックを分析]を選択します。

認識環境の設定：

- 次のいずれかの方法で認識環境を設定してください。
  - [分析/認識]—[認識環境設定]を選択します。
  - コマンドバー上の3つ目のアイコンの下にある  をクリックし、リストから[認識環境設定]を選択します。
- ダイアログボックスから任意のテンプレート名を選択し、ページレイアウトを設定します。[初期値に設定]をクリックすると、現在のテンプレートがデフォルトの認識設定になります。元のデフォルト設定に戻すには、[初期設定]をクリックしてください。
- また[リネーム]をクリックすると、現在のテンプレートの名前を変更することができます。




[認識環境設定]のダイアログボックス

- [OK]をクリックします。

## 校正ライブラリの設定

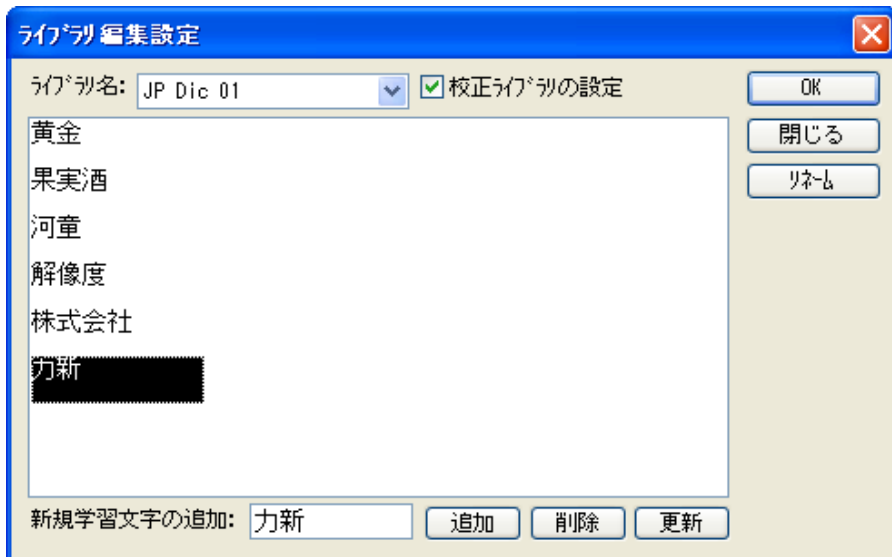
校正ライブラリには、認識中によく使用する単語が保存されています。したがってドキュメントの種類によって異なる校正ライブラリを設定することによって、認識性能が高まり、校正が容易になります。

 これらのライブラリ画像はサンプル用に表示するだけであり、実際の校正ライブラリではありません。

- [校正ライブラリの設定](#)
- [校正ライブラリの編集](#)

## 校正ライブラリの設定

1. [分析/認識]—[辞書編集設定]を選択します。
2. 必要に応じて次の設定を行ってください。



[辞書編集設定]のダイアログボックス



<b>リネーム</b>	現在使用している辞書の名前を変更します。
<b>追加</b>	[新規学習文字の追加]に表示されているテキストブロックの記号や単語を校正ライブラリに追加します。
<b>削除</b>	辞書から単語を削除します。
<b>更新</b>	辞書の単語を更新します。
<b>校正ライブラリの設定</b>	選択した辞書を校正ライブラリとして設定します(複数選択可)。


## 校正ライブラリの編集

校正ライブラリを切り替えて、校正を再度実行することができます。校正ライブラリを編集するには：

1. [分析/認識]ー[辞書編集設定]を選択し、任意の校正ライブラリを設定した後で[OK]をクリックします。
2. [分析/認識]ー[校正ライブラリの編集]を選択して校正を開始します。

## 自動処理

「MaxReader 中日英文 OCR」では自動的に認識を行い、認識結果を出力できるように、自動処理機能を使用することができます。自動処理の手順は次の通りです。

1. コマンドバー上で最初のアイコンの下にある  をクリックして操作を選択するか、メニューバーから[ファイル]ー[自動操作]を選択します。
2. 選択したコマンドによって、手順が変わります。

コマンド	手順
[スキャン>>認識]	スクリーンにスキャナドライバの画面が表示されます。[スキャン]をクリックするとドキュメントがスキャンされ、自動的にスキャンした画像が認識されます。
[スキャン>>認識>>エクスポート]	スクリーンにスキャナドライバの画面が表示されます。[スキャン]をクリックするとドキュメントがスキャンされ、自動的にスキャンした画像が認識されます。その後、認識結果が[自動操作エクスポート設定]のダイアログボックスで選択したアプリケーションにエクスポートされます。
[開く>>認識]	[画像を開く]のダイアログボックスが開きますので、任意のファイルを選択してください。Shift キーか Ctrl キーを押しながら選択すると、一度に複数のファイルを選択できます。[開く]をクリックすると、自動的にドキュメントが認識されます。
[開く>>認識>>エクスポート]	[画像を開く]のダイアログボックスが開きますので、任意のファイルを選択してください。Shift キーか Ctrl キーを押しながら選択すると、一度に複数のファイルを選択できます。[開く]をクリックすると、自動的にドキュメントが認識されます。その後、認識結果が[自動操作エクスポート設定]のダイアログボックスで選択したアプリケーションにエクスポートされます。

## 作業のキャンセル

認識を行ってリセットした後で認識結果をキャンセルするには、次のいずれかの手順にしたがってください。

現在のページの認識結果をキャンセルするには：

- [分析/認識]—[このページのみキャンセル]を選択します。
- 画像モードに切り替えて、ドキュメントを右クリックして[このページのみキャンセル]を選択します。
- サムネイル表示でドキュメントを右クリックして[このページのみキャンセル]を選択します。

すべての結果をキャンセルするには、[分析/認識]—[すべてキャンセル]を選択してください。

## 第5章 **校正と編集**


システムが認識を完了したら、ページモードか編集モードで認識結果を校正することができます。認識中に疑わしい文字が表示されます。疑わしい文字は、画像が不鮮明であったり、レイアウト設定や認識の段階でエラーが発生した場合に生じることがあります。必要であればドキュメントを校正してください。以下に認識したファイルを校正、編集する手順を説明します。

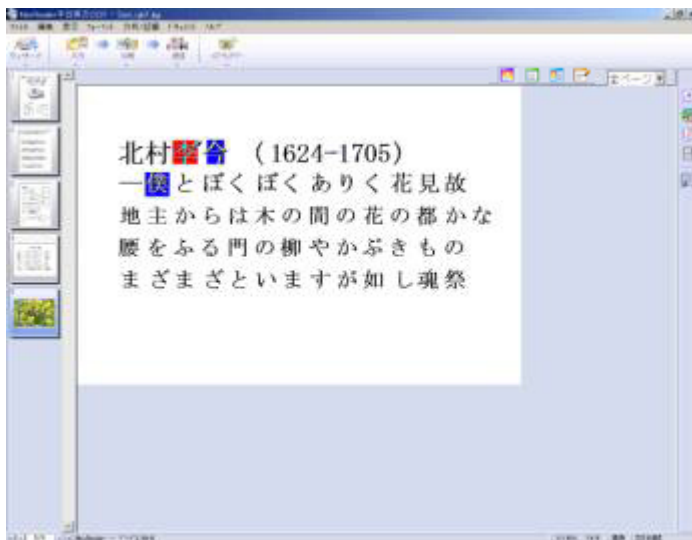
- [校正と編集](#)
- [部分再認識](#)
- [クイックインプット](#)
- [登録文字の学習](#)

### 校正と編集

認識が完了すると、デフォルトの設定に応じてシステムはページモードか編集モードに入ります。ページモードでは認識結果を一度にすべて見ることができます。また編集モードでは認識結果をブロック毎に見ることができます。これら2つのモードを切り替えるには、表示ツールバー上のツールをお使いください。

**ページモードでの校正：**

1. 編集ツールバーで[文字の校正]  をクリックすると、青い背景に黄色い文字で疑わしい文字が表示されます。



青い背景に黄色い文字で示された疑わしい文字

2. [文字の校正]ツールで最初の疑わしい文字をクリックし、文字の校正ウィンドウで正しい文字を選択してください。選択した文字が疑わしい文字に置き換えられます。



文字の校正ウィンドウ

文字の校正ウィンドウに正しい文字が表示されない場合は、キーボードを使って正しい文字を入力してください。

3. カーソルを次の疑わしい文字に合わせるには、[文字の校正]ツールで誤

認識された文字を選択し、校正を続行してください。

### 編集モードでの校正：

1. ドキュメントブラウザバーかブロックメニューから校正したいブロックを選択すると、文書校正ウィンドウに疑わしい文字を青で示したブロックの認識結果が表示されます。
2. 文書校正ウィンドウで誤認識された文字をクリックすると、文字画像ウィンドウにその文字が赤い枠で囲まれます。
3. 文字画像ウィンドウの下に表示された[候補文字]から正しい文字を選択すると、その文字が文字校正ウィンドウでカーソルが示す文字に置き換わります。

候補文字ウィンドウに正しい文字がリストされない場合は、カーソルを疑わしい文字の左側に合わせて、キーボードを使って正しい文字を入力してください。

4. 次の疑わしい文字をクリックすると、その文字が選択されます。



## 部分再認識

画像ドキュメントによっては、システムが間違った認識結果をエクスポートしてしまう場合があります。そのような場合は、黒い点を消したり、文字、行、ブロックを分割／結合したり、誤認識されたエリアを認識し直したりしてください。

- [消しゴム](#)
- [行の結合再認識](#)
- [ペイント](#)
- [ブロックの再認識](#)
- [文字の分割再認識](#)
- [ブロックの結合再認識](#)
- [文字の結合再認識](#)
- [ブロックの分割再認識](#)
- [行の分割再認識](#)



## 消しゴム

画像から黒い点を除去すると認識精度を高めることができます。編集モードに切り替えると、[消しゴム]ツールを使って文字画像ウィンドウの赤い枠にある黒い点を消すことができます。

1. 文書校正ウィンドウで誤認識された文字を選択すると、文字画像ウィンドウの赤い枠にその文字が表示されます。
2. 編集ツールバーから[消しゴム]  を選択します。消しゴムのサイズは、左上隅の  をクリックすると変更できます。
3. カーソルをドラッグして赤い枠内の黒い点を消してください。誤って消してしまった場合は、Esc キーを押してください。
4. 赤い枠の外側をどこでもクリックするか、Enter キーを押すと、その文字が直ちに再認識されます。


## ペイント

画像の中に現れた白いスポットをペイントすると、認識の精度を高めることができます。編集モードに切り替えると、[ペイント]ツールを使って文字画像ウィンドウの赤い枠にある文字をペイントすることができます。

1. 文書校正ウィンドウで誤認識された文字を選択すると、文字画像ウィンドウの赤い枠にその文字が表示されます。
2. 編集ツールバーから[ペン]  を選択します。ペンのサイズは、左上隅の  をクリックすると変更できます。
3. カーソルを赤い枠内の文字のペンにドラッグします。誤ってペイントしてしまった場合は、Esc キーを押してください。
4. 赤い枠の外側をどこでもクリックするか、Enter キーを押すと、その文字が直ちに再認識されます。


## 文字の分割再認識

隣り合う、またはシステムが誤って結合した2個以上の文字を分割し、再認識することができます。

1. 文書校正ウィンドウで誤認識された文字を選択すると、文字画像ウィンドウの赤い枠にその文字が表示されます。
2. 編集ツールバーから[文字の分割再認識]  を選択します。
3. マウスを押したままにすると、赤い枠の中に赤い線が表われ文字を2つの部分に分割します。赤い線の位置を調整して、マウスを離してください。分割した文字が再認識され、文字校正ウィンドウの認識結果が更新されます。


## 文字の結合再認識

システムが誤って2つ以上に分割した文字を結合し、再認識することができます。

1. 文書校正ウィンドウで誤認識された文字を選択すると、文字画像ウィンドウの赤い枠にその文字が表示されます。
2. 編集ツールバーから[文字の結合再認識]  を選択します。
3. 赤い枠内でマウスを押したまま結合したい文字にドラッグした後、マウスを離します。接続した文字が再認識され、文字校正ウィンドウの認識結果が更新されます。

## 行の分割再認識

システムが誤って結合した2つ以上の行を分割し、再認識することができます。


1. 文書校正ウィンドウで誤認識された文字を選択すると、文字画像ウィンドウの赤い枠にその文字が表示されます。
2. 編集ツールバーから[行の分割再認識]  を選択します。



3. 赤い枠内をクリックすると、結合した行が分割され再認識されます。それに伴い、文字校正ウィンドウの結果も更新されます。



## 行の結合再認識

システムが誤って2つ以上に分割した行を結合し、再認識することができます。

1. 文書校正ウィンドウで誤認識された文字を選択すると、文字画像ウィンドウの赤い枠にその文字が表示されます。
2. 編集ツールバーから[行の結合再認識]  を選択します。
3. 赤い枠内でマウスを押したまま結合したい行にドラッグした後、マウスを離します。

## ブロックの再認識

特定のブロック内で認識結果にエラーが見つかった場合は（配列設定や文字種の設定など）、そのブロックだけを再認識することができます。

1. 表示ツールバーの[ページモード]アイコン  をクリックします。
2. 編集ツールバーから[ブロックをマーク]  を選択し、任意のブロックをクリックします。するとそのブロックが黄色で表示されます。
3. [システム設定]のダイアログボックスで[再認識時に属性ダイアログボックスを表示]機能を無効にしてある場合は、次のいずれかの方法でブロックの属性をリセットする必要があります。
  - [フォーマット]–[ページレイアウト]を選択します。
  - 画像モードに切り替えてドキュメントを右クリックし、[段落の属性を変更する]–[属性の応用設定]を選択します。



[再認識時に属性ダイアログボックスを表示]機能が有効になっている場合はこのステップは省略し、そのままステップ4に進んでください。

4. [分析/認識]–[ブロック再認識]を選択するか、F8 キーを押します。[システム設定]のダイアログボックスで[再認識時に属性ダイアログボックスを表示]機能を有効にしてある場合は、画面に[ブロック設定]のダ

イアログボックスが表示されます。ここでブロックの属性をリセットすることができます。[OK]をクリックすると、システムが選択したブロックを再認識し、新しい結果を表示します。



## ブロックの結合再認識

誤って分割されたブロックを結合し、そのブロックを再認識します。

1. 表示ツールバーの[ページモード]アイコン  をクリックします。
2. 編集ツールバーで[ブロックの結合]  を選択します。
3. マウスを押したまま結合したいブロックを囲んでください。
4. マウスを離すと、結合したブロックが再認識されます。[システム設定]のダイアログボックスで[再認識時に属性ダイアログボックスを表示]機能を有効にしてある場合は、画面に[ブロック設定]のダイアログボックスが表示されます。ここでブロックの属性をリセットすることができます。[OK]をクリックすると、システムが選択したブロックを再認識し、新しい結果を表示します。

## ブロックの分割再認識

誤って結合されたブロックを分割し、そのブロックを再認識します。

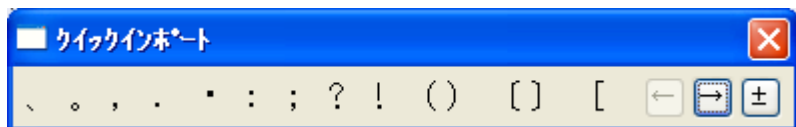
1. 表示ツールバーの[ページモード]アイコン  をクリックします。
2. 編集ツールバーから[ブロックの分割再認識]  を選択します。
3. 任意のブロック内をクリックし、赤い線の位置を調整するか、マウスをドラッグして赤い線を移動すると、ブロックが分割されます。
4. Enter キーを押すか、ブロックの外側をどこでもクリックすると、分割したブロックが再認識されます。[システム設定]のダイアログボックスで[再認識時に属性ダイアログボックスを表示]機能を有効にしてある場合は、画面に[ブロック設定]のダイアログボックスが表示されます。ここでブロックの属性をリセットすることができます。[OK]をクリックすると、システムが選択したブロックを再認識し、新しい結果を表示します。



## クイックインプット

編集モードで校正や編集を行うときには、特別な記号を入力する場合があります。そのような場合は、クイックインプット機能を使うとこれらの記号を簡単に入力することができます。


クイックインプットを行うには：

1. 文字校正ウィンドウで記号を挿入したい場所をクリックすると、カーソルが表示されます。
2. [表示]ー[クイックインプット]を選択します。すると[クイックインプット]ツールバーが表示されます。


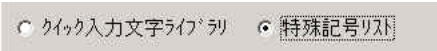


3. 挿入したい記号が見つからない場合は、かボタンをクリックして前後に移動し、記号を探してください。
4. 記号をクリックすると、カーソルがある場所にその記号が挿入されます。

記号を[クイックインプット]ツールバーに追加するには：

1. [表示]ー[クイックインプット]を選択します。
2.  をクリックすると、[クイック入力設定]のダイアログボックスが開きます。
3. [前ページ]か[次ページ]をクリックして記号リスト全体をブラウズし、[記号リスト]に追加したい記号を検索します。追加したい記号をクリックしてください。または[新規登録文字の追加]に手動で記号を入力することもできます。
4. [追加]をクリックします。
5. [終了]をクリックするとダイアログボックスが閉じます。


記号を[クイックインプット]ツールバーから削除するには：

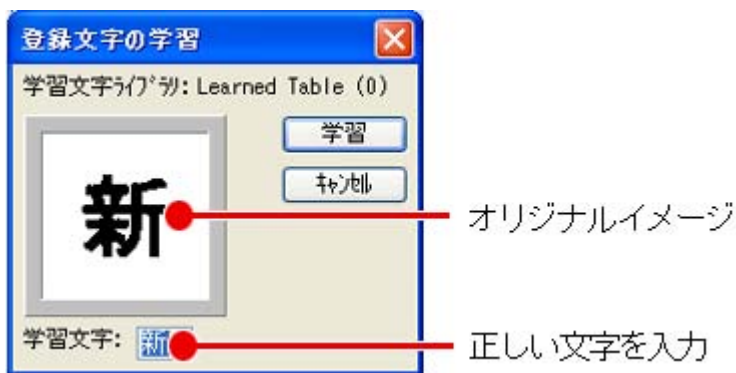
1. [表示]—[クイックインプット]を選択します。
2.  をクリックすると、[クイック入力設定]のダイアログボックスが開きます。
3. [特殊記号リスト]を選択します。  

4. [特殊記号リスト]から削除したい記号を選択し、[削除]をクリックします。
5. [終了]をクリックするとダイアログボックスが閉じます。

## 登録文字の学習

システムが何度も同じ文字を誤認識する場合は、登録文字の学習機能を使ってそれらの文字をライブラリに登録することができます。ドキュメントの文字の種類によって、ライブラリを使い分けることも可能です。

### ユーザー登録文字

1. 表示ツールバーの[編集モード]アイコン  をクリックします。
2. 文字校正ウィンドウで誤って認識された文字を選択します。
3. [分析/認識]—[登録文字の学習]を選択すると、そのダイアログボックスが開きます。



選択した文字がダイアログボックスの上部に表示されます（これは文字画像ウィンドウの赤い枠内に表示された文字と同じです）。

4. [学習文字]に正しい文字を入力します。
6. [登録文字の学習]キーをクリックすると新しい文字がライブラリに登録され、文字校正ウィンドウに新しい文字が表示されます。

登録文字の学習についての詳細は、次のトピックを参照してください。

- ▶ [登録文字の削除](#)
- ▶ [学習文字ライブラリ設定](#)

## 登録文字の削除

ライブラリから文字を削除するには、[分析/認識]－[登録文字の削除]を選択します。ライブラリから削除したい文字を選択し、[削除]をクリックします。

## 学習文字ライブラリの設定

学習文字ライブラリの設定機能を使うと、特定のライブラリをデフォルトのライブラリとして設定したり、ライブラリに別名を付けたりすることができます。

1. [分析/認識]－[学習文字ライブラリ設定]を選択します。



2. デフォルトのライブラリとして設定したいライブラリを選択し、[OK]をクリックします。ライブラリに別名を付けるには、[リネーム]をクリックしてください。

## 第6章 ファイルのエクスポート

「MaxReader 中日英文 OCR」では将来いつでも編集できるように、認識前の画像ファイル、頻繁に使用するページレイアウト、認識後のテキスト／画像／フォームを個別に保存することができます。

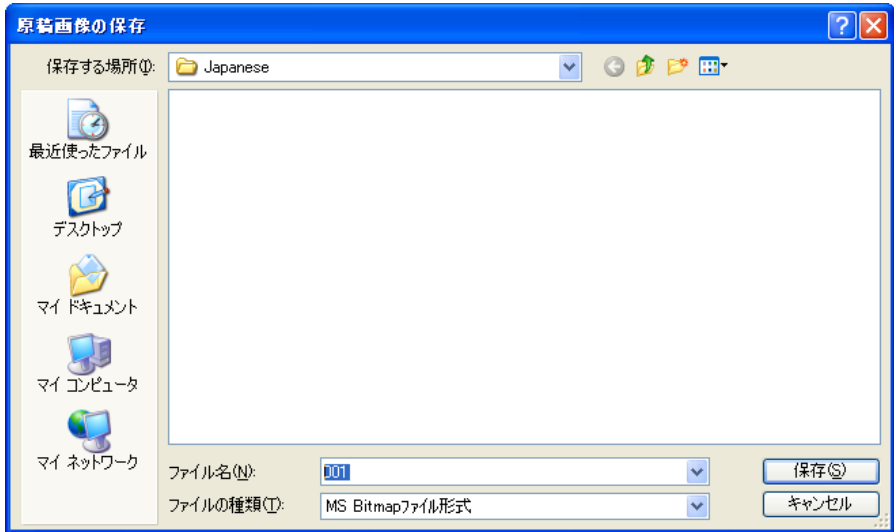
- [画像の保存](#)
- [認識結果の保存](#)
- [プロジェクトの保存](#)
- [頻繁に使うページレイアウトの保存](#)
- [印刷](#)
- [他のプログラムに送信](#)
- [自動操作エクスポート設定](#)

### 画像の保存

スキャナやその他のインポート元から取り込んだ画像を保存しておきたい場合は、それらの画像を BMP、TIFF、PCX、JPEG などの形式で保存することができます。

ドキュメントの画像を保存するには：

1. [ファイル]－[原稿画像の保存]を選択します。
2. [原稿画像の保存]のダイアログボックスでファイルパス、ファイル名、ファイル形式を選択し、[保存]をクリックします。



[原稿画像の保存]のダイアログボックス



## 認識結果の保存

「MaxReader 中日英文 OCR」では認識したドキュメントをさまざまな形式で保存することができます。現在開いているページや、必要であればすべてのページを TXT、DOC、RTF、XLS、SLK、CSV などの形式で保存し、後から WordPad、Word、Excel などのプログラムで編集してください。また HTML 形式で保存すると、Web ブラウザ（Internet Explorer や Netscape Navigator など）で開くことができます。

認識結果を保存するには：



1. 現在開いているページだけを保存するには、次のいずれかの手順にしたがってください。

- [ファイル]—[認識結果の保存]を選択します。

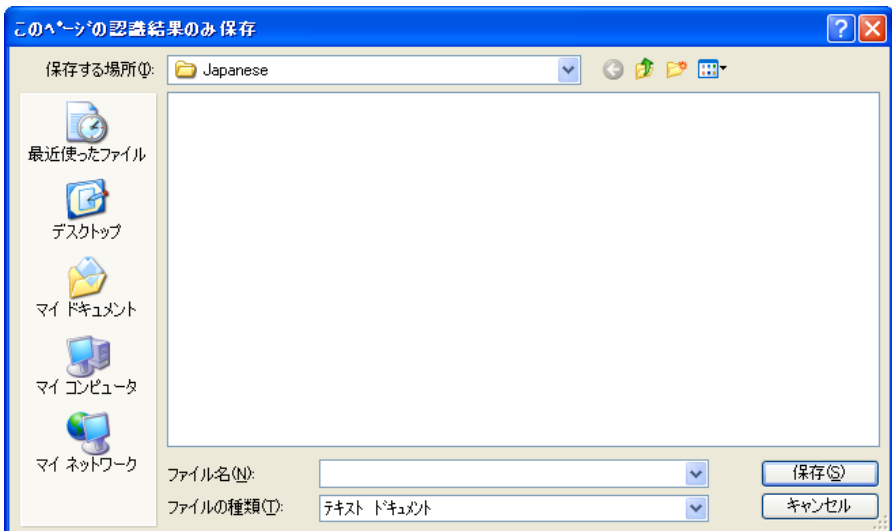
- コマンドバー上の4つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[認識結果の保存]を選択します。



すべてのページを保存するには、次のいずれかの手順にしたがってください。

- [ファイル]—[OCR 後のデータをすべて保存する]を選択します。
- コマンドバー上の4つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[OCR の結果をすべて保存する]を選択します。

すると[このページの認識結果のみ保存]か[OCR 後のデータをすべて保存する]のダイアログボックスが開きます。



2. ファイルパス、ファイル名、ファイル形式を選択してください。



ファイル形式	説明
テキスト ドキュメント	書式なし文書です。これを選択すると、テキストだけが保存され画像は保存されません。
MS_Word	Microsoft Word ファイルです。これを選択すると、ファイルを *.doc ファイルとして保存します。ファイルに画像が含まれている場合は、それらも保存されます。複数のページを持つドキュメントを保存する場合は、自動的にページの区切りが挿入されます。
リッチ テキスト形式	このファイル形式は文字サイズ、ページレイアウト、画像、フォームなどを維持します。複数のページを持つドキュメントを保存する場合は、自動的にページの区切りが挿入されます。
Excel	Excel スプレッドシートとして開くことができます。
SLK	Excel スプレッドシートとして開くことができます。
CSV	Excel スプレッドシートとして開くことができます。
HTML または HTML for IE4.0	Web ブラウザで開くことができます。ドキュメントに画像が含まれている場合は、画像を JPEG ファイルとして保存し、それらの順番を付けます。複数のページを持つドキュメントを保存する場合は、ページとページの間に自動的に横線が配置されます。

### 3. [保存]をクリックします。

## プロジェクトの保存

認識したドキュメントを編集してる途中で作業を中断する必要がある場合は、画像と認識済みファイルを保存してください。

編集が完了していないファイルを保存するには：

1. 次のいずれかの方法でプロジェクトを保存してください。
  - [ファイル]-[プロジェクトを保存する]を選択します。
  - コマンドバー上の4つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[プロジェクトを保存する]を選択します。

すると次ようなダイアログボックスが開きます。



2. フォルダを選択するか、[新しいフォルダ]をクリックして新しくフォルダを作成してください。

3. [OK]をクリックして現在の編集状態と画像ファイルを選択したフォルダに保存します。

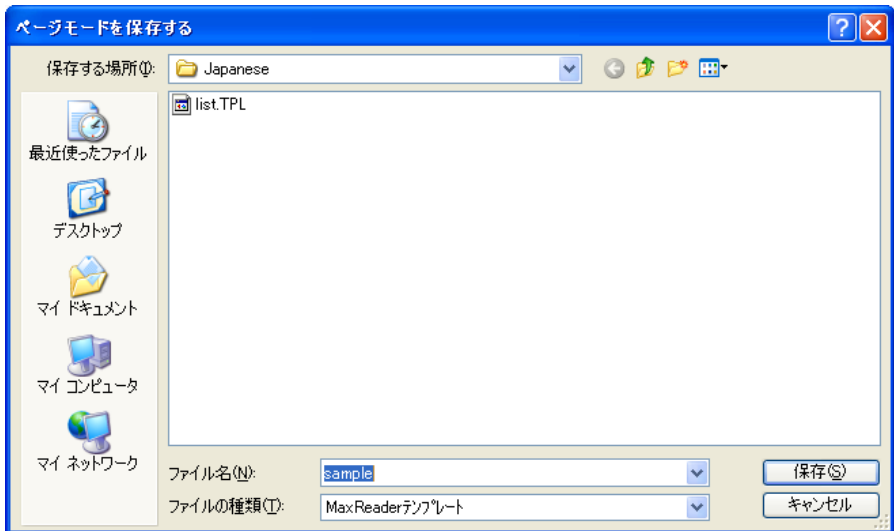
一度保存したファイルを開く方法については、第3章 – [プロジェクトを開く](#) を参照してください。またプロジェクトを「MaxReader中日英文OCR」がインストールされた他のコンピュータにコピーすると、編集を続行することができます。

## 頻繁に使うページレイアウトの保存

特定のページレイアウトを頻繁に使用するときには、そのレイアウトをテンプレートファイル (\*.TPL) として保存し画像ファイルに適用すると、設定の手間を省けるだけでなく、認識精度を高めることもできます。

ページレイアウトを保存するには：

1. ページ分析を実行するか、手動でブロックとブロックの属性を設定してください。
2. [フォーマット]–[テンプレートの保存]を選択します。



[ページモードを保存する]のダイアログボックス

3. ファイルパスとファイル名を選択し、[保存]をクリックします。

保存したテンプレートを使用する方法については、第4章 – [ページを開く](#)を参照してください。

## 印刷

「MaxReader 中日英文 OCR」では原稿と認識結果を印刷することができます。認識結果を印刷するには、[ファイル]–[印刷]–[認識結果の印刷]を選択します。原稿を印刷するには、[ファイル]–[印刷]–[原稿画像の印刷]を選択してください。

## 他のプログラムに送信

「MaxReader 中日英文 OCR」では画像ファイルをEメールに直接送信したり、認識結果や画像ファイルをWordPad、ペイント、Acrobat Readerなどのアプリケーションに送信したりすることができます。任意のアプリケーションを設定してファイル形式をエクスポートできるように、10個の送信設定グループを記録することができます。設定を選択したアプリケーションに直接適用し、ファイルを送信してください。


**ファイルをEメールで送信するには：**

[ファイル]–[転送]を選択して開いた画像ファイルをEメールソフトに送信します。

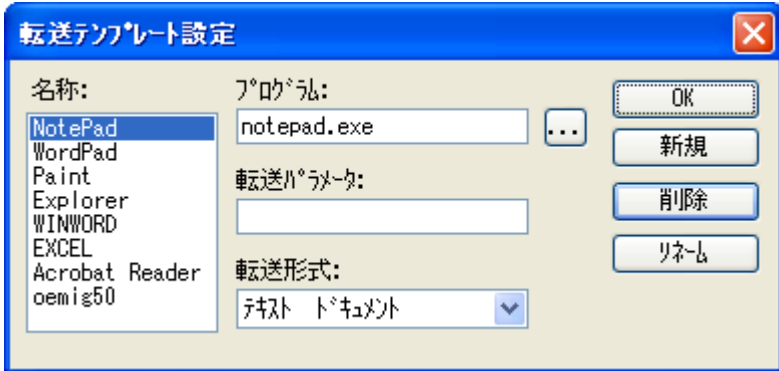


「MaxReader 中日英文 OCR」から直接Eメールソフトに送信できるように、使用するEメールソフトはExchangeやOutlook ExpressなどのようにMAPIメールシステムに対応したものでなければなりません。

**送信設定を行うには：**

- 次のいずれかの方法を選択してください。
  - [ファイル]–[転送テンプレート設定]を選択します。
  - コマンドバー上の4つ目のアイコンの下にある  をクリックし、リストから[エクスポート設定]を選択します。

2. [転送テンプレート設定]のダイアログボックスで任意のフィールドを設定してください。



[転送テンプレート設定]のダイアログボックス

- |                |                            |
|----------------|----------------------------|
| <b>新規</b>      | 新しい転送テンプレート設定を追加します。       |
| <b>削除</b>      | 選択した転送テンプレート設定を削除します。      |
| <b>リネーム</b>    | 選択した転送テンプレート設定の名前を変更します。   |
| <b>プログラム</b>   | 転送テンプレート設定のアプリケーションを指定します。 |
| <b>転送パラメータ</b> | 転送テンプレート設定のパラメータを指定します。    |
| <b>転送形式</b>    | 転送テンプレート設定のファイル形式を設定します。   |

3. [OK]をクリックします。

**転送テンプレート設定を適用するには：**

コマンドバー上の4つ目のアイコン  をクリックして[現在の開いているページのOCRの結果をエクスポートする]、[すべての結果をひとつの

ファイルにまとめてエクスポートする]、または[すべての結果を別々のファイルとしてエクスポートする]を選択し、アプリケーションを選択してください。以下は各オプションの説明です。

オプション	説明
現在の開いているページの OCR の結果をエクスポートする	現在開いているページの認識結果を選択したアプリケーションにエクスポートします。
すべての結果をひとつのファイルにまとめてエクスポートする	すべてのページの認識結果を選択したアプリケーションにエクスポートします。
すべての結果を別々のファイルとしてエクスポートする	すべてのページの認識結果を別々のファイルで選択したアプリケーションにエクスポートします。




エクスポート処理が行われている間、ファイルは一時的に「Pccrtemp」ディレクトリに保存されます。このディレクトリのファイルは定期的に削除して、ハードディスクのスペースを空けるようにしてください。

## 自動操作エクスポート設定

「MaxReader 中日英文 OCR」の自動操作で[スキャン>>認識>>エクスポート]と[開く>>認識>>エクスポート]を実行すると、自動的に認識結果がエクスポートされます。このセクションでは、エクスポートファイル形式とアプリケーションの設定を行う方法を説明します。

自動操作のエクスポートファイル形式とアプリケーションを設定するには：



1. 次のいずれかの方法でエクスポートアプリケーションを設定してください。
  - [ファイル]ー[自動操作エクスポート設定]を選択します。
  - コマンドバー上の最初のアイコンの  をクリックし、リストから[自動操作エクスポート設定]を選択します。
2. すると[自動操作エクスポート設定]のダイアログボックスが開きます。左側の[名前]アプリケーションリストから任意のエクスポートアプリケーションを選択します。
3. [フォーマット]から適切なファイル形式を選択してください。
4. [OK]をクリックします。

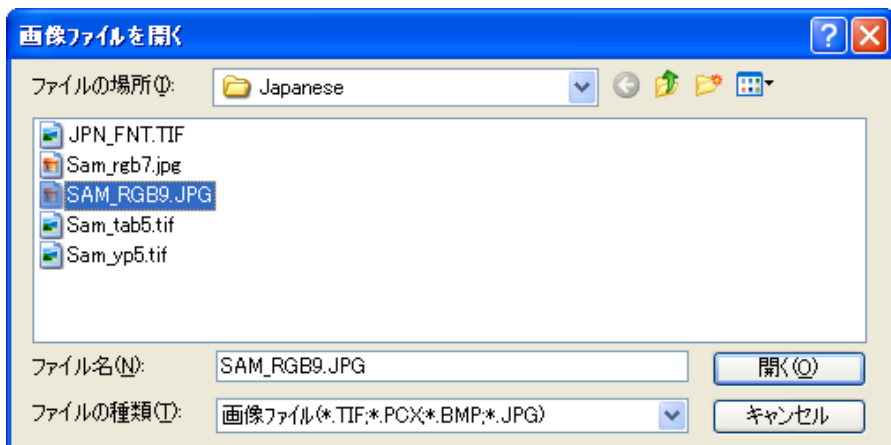
設定を変更するまで、[スキャン>>認識>>エクスポート]または[開く>>認識>>エクスポート]を選択するたびに、認識結果ドキュメントは選択したアプリケーションにエクスポートされます。



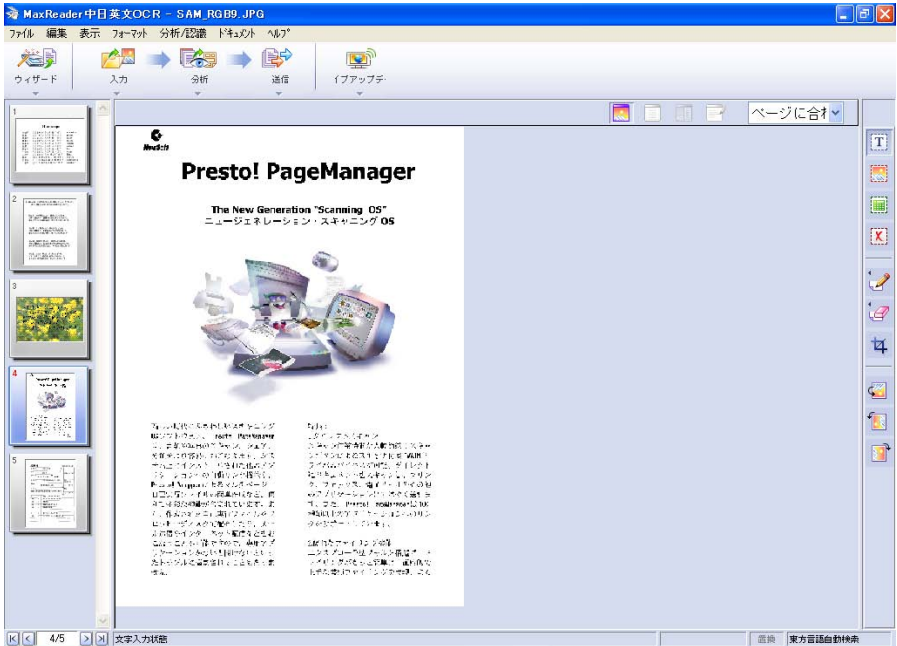
## 第7章 日本語認識サンプル

通常のドキュメントには、日本語ドキュメントに画像が含まれています。この章では、このようなドキュメントのページレイアウトと認識の例を紹介します。

1. コマンドバー上の2つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[画像を開く]を選択します。
2. 「MaxReader 中日英文 OCR」プログラムフォルダの「Example\Japanese」ディレクトリから SAM\_RGB9.jpg ファイルを選択します。



3. [開く]をクリックしてファイルをインポートします。

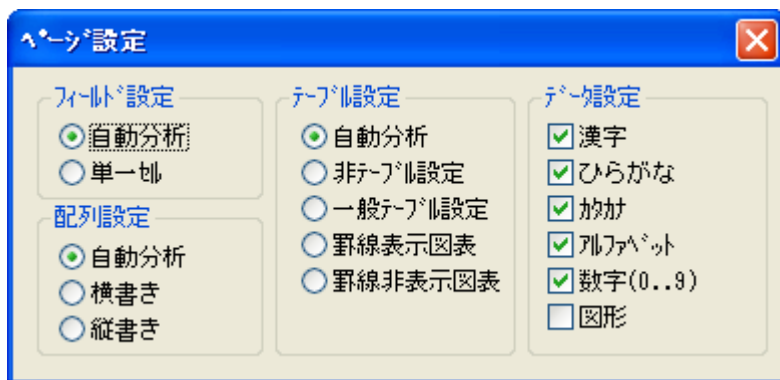





4. [フォーマット]—[認識ライブラリ設定]を選択してサブメニューから[日本語]を選択するか、ステータスバーの右側をクリックして[日本語]を選択します。

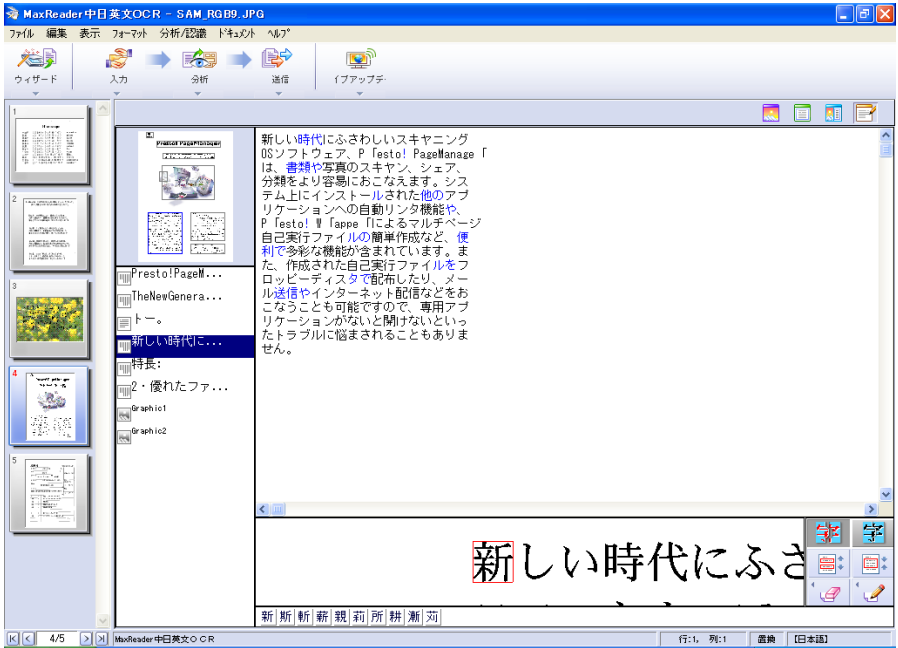
東方言語自動検索  
【繁体字中国語Ⅰ】  
【繁体字中国語Ⅱ】  
【簡体字中国語】  
【日本語】


5. [フォーマット]—[ページレイアウト]を選択します。
6. [ページレイアウト]のダイアログボックスの設定を次のように調整してください。

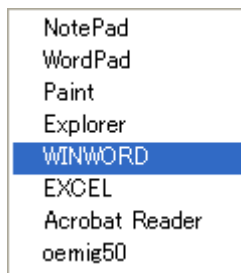
アイテム	設定
フィールド設定	自動分析
配列設定	自動分析
テーブル設定	自動分析
データ設定	漢字、ひらがな、カタカナ、アルファベット、数字 (0..9)、図形



7. コマンドバーで3つ目のアイコン  をクリックし、[現在開いているページを認識]を選択すると認識が開始されます。
8. 表示ツールバーからページモード  と編集モード  を自在に切り替えて、認識結果を表示し、必要に応じて編集してください（ドキュメントの校正については、第5章を参照してください）。



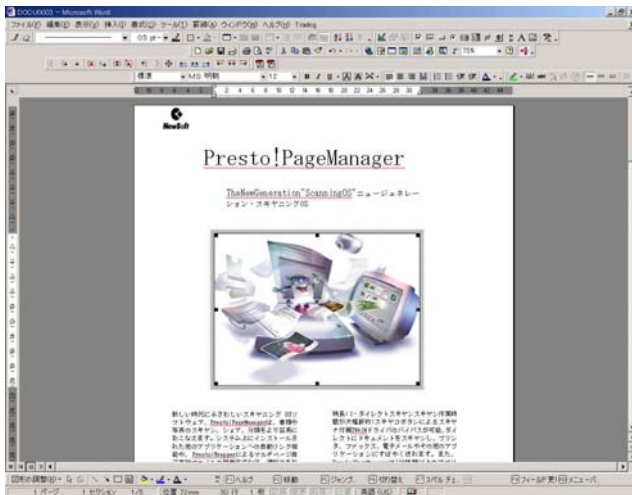
9. コマンドバーの4つ目のアイコン  をクリックし、[現在開いているページのOCRの結果をエクスポートする]を選択した後、[Winword]を選択します。



- (1) 先にMicrosoft Word をインストールし、[転送テンプレート設定]でこれをテンプレートとして設定しておかなければ、プルダウンリストから4つ目のアイコンを選択することはできません。[転送テンプレート設定]についての詳細は、第6章— [他のプログラムに送信](#) を参照してください。

(2) オリジナルのページフォーマットを維持するには、[ファイル]–[システム設定]を選択して Microsoft Word ドキュメントに改行を入れるよう設定してください。ページを手動で編集し、並べ替えたい場合は、改行を入れないでください。



10. 認識結果が直接 Microsoft Word に転送されます。[表示]–[フォーマット設定]を選択し、ドキュメント全体を表示します。

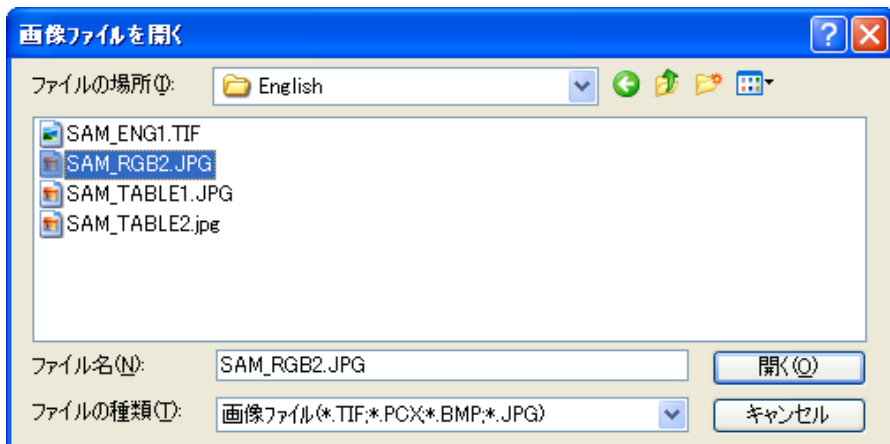


Microsoft Word に転送された認識結果

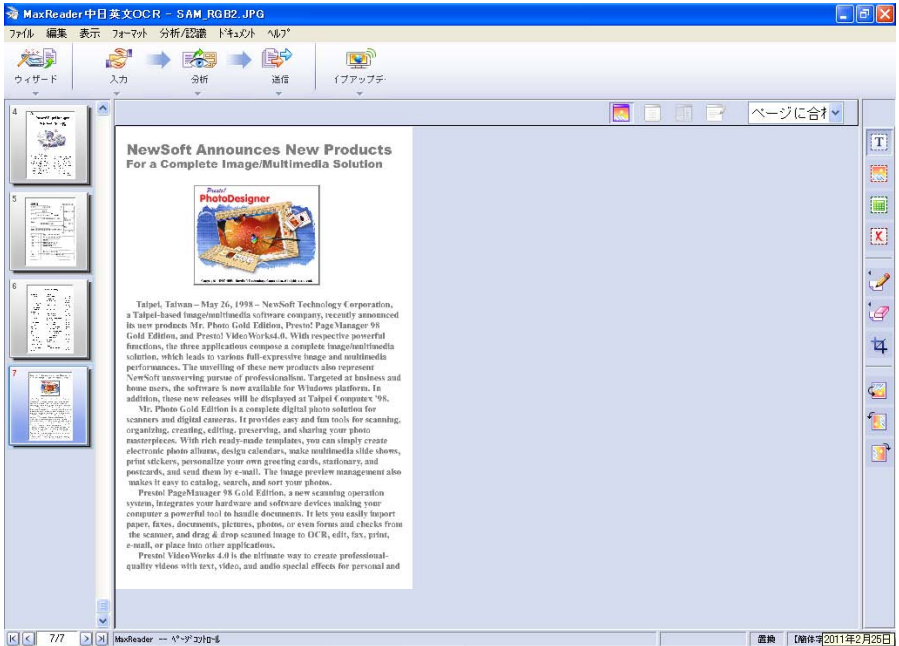
## 第8章 英語の認識サンプル

日本語ドキュメントのほかにも「MaxReader 中日英文 OCR」では英語ドキュメントも認識することができます。この章では、英語のみのドキュメントを認識する手順を説明します。

1. コマンドバー上の2つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[画像を開く]を選択します。
2. 「MaxReader 中日英文 OCR」プログラムフォルダの「Example\English」ディレクトリから SAM\_RGB2.jpg ファイルを選択します。

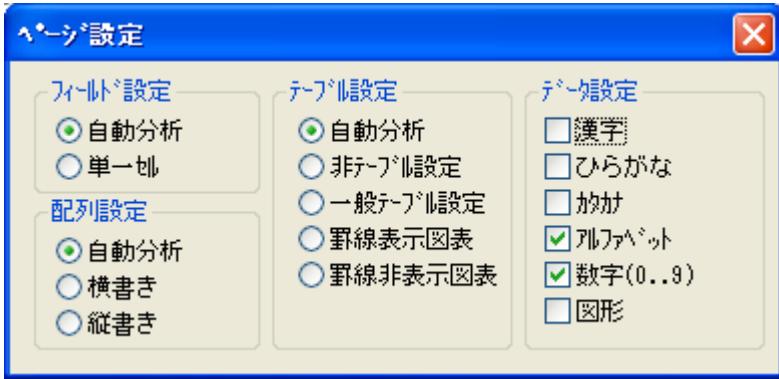




3. [開く]をクリックしてファイルをインポートします。

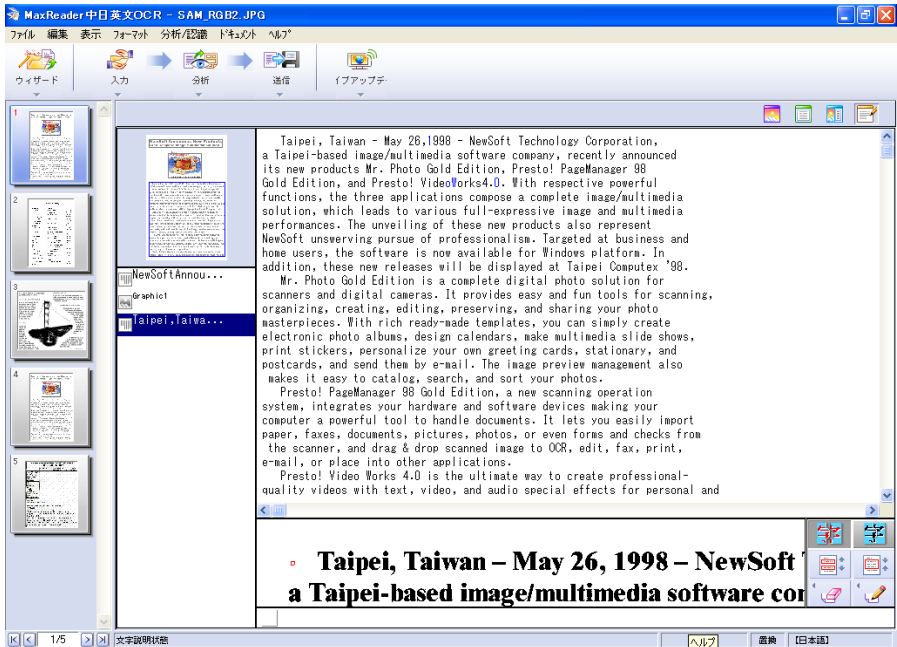




4. [フォーマット]—[レイアウト設定]を選択します。
5. [ページレイアウト]のダイアログボックスの設定を次のように調整してください。

アイテム	設定
フィールド設定	自動分析
配列設定	自動分析
テーブル設定	自動分析
データ設定	アルファベット、 数字 (0..9)




6. コマンドバー上の3つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[現在開いているページを認識]を選択します。

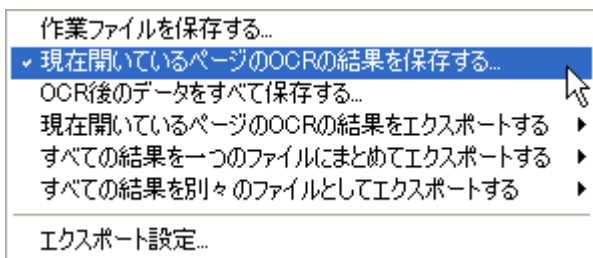


7. 表示ツールバーからページモード  と編集モード  を自在に切り替えて、認識結果を表示し、必要に応じて編集してください（ドキュ

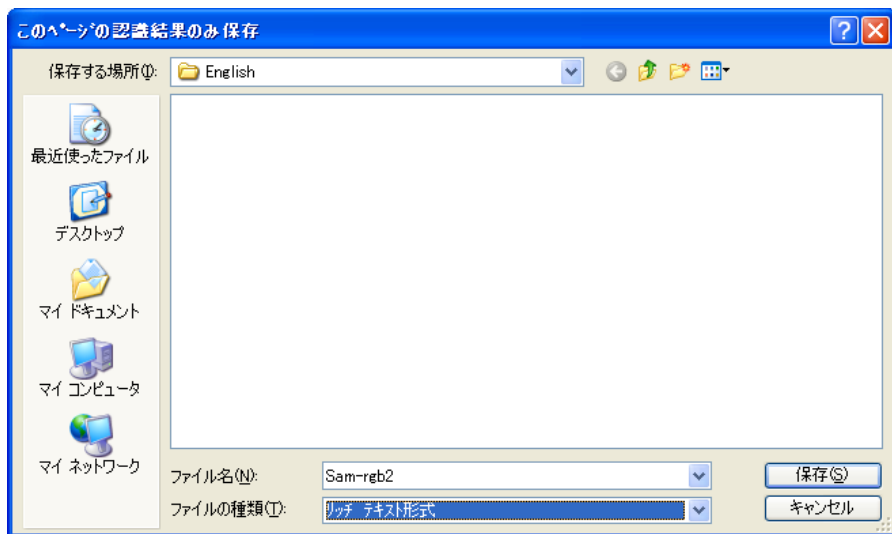


メントの校正については、第5章を参照してください)。


8. 校正が完了したら、コマンドバーで4つ目のアイコン  をクリックし、[現在開いているページの OCR の結果を保存する]を選択してください。



9. [このページの認識結果のみ保存]のダイアログボックスでファイル名を「Sam-rgb2」と入力し、リッチテキスト形式 (RTF) で保存します。



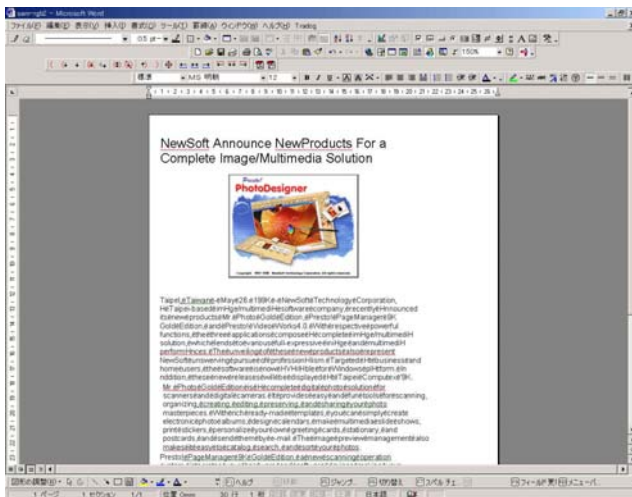
オリジナルのページフォーマットを維持するには、[ファイル]-[システム設定]を選択してRTFファイルに改行を入れるよう設定してください。ページを手動で編集し、並べ替えたい場合は、改行を入れないでください。

10. またファイルを関連するソフトウェアに転送することもできます。コマンドバーの4つ目のアイコン  をクリックし、[現在開いているページの OCR の結果をエクスポートする]を選択した後、[Winword]を選択します。



先にMicrosoft Word をインストールし、[転送テンプレート設定]でこれをテンプレートとして設定しておかなければ、プルダウンリストから4つ目のアイコンを選択することはできません。[転送テンプレート設定]についての詳細は、第6章— [他のプログラムに送信](#)を参照してください。

11. 認識結果が直接 Microsoft Word に転送されます。[表示]—[フォーマット設定]を選択し、ドキュメント全体を表示します。



Microsoft Word に表示された認識結果





画像を保存する必要がないときには、テキストのみが含まれたドキュメントを書式なし文書 (.txt) として保存することもできます。このファイルはワードパッドやメモ帳でも開くことができます。

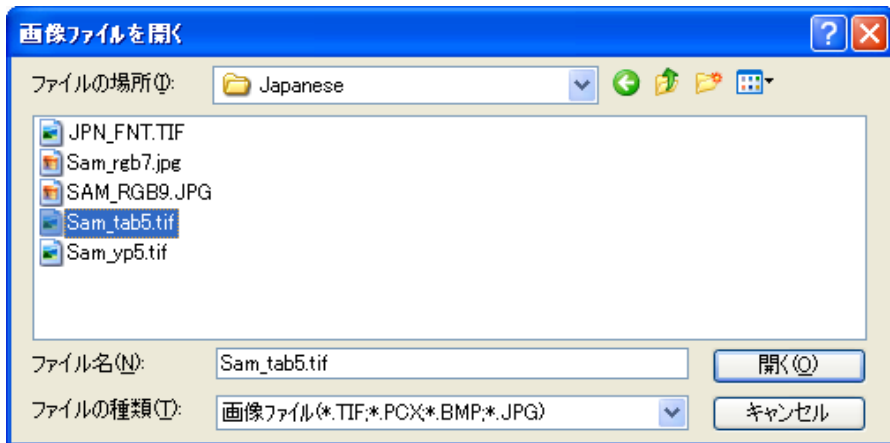
## 第9章 フォームの認識サンプル

「MaxReader 中日英文 OCR」では画像とテキストの両方を含む日本語／英語のドキュメントのほかに、公式文書、住所録、履歴書、レポートカードなど、さまざまなフォームも認識することができます。フォームのレイアウトとテキストはオリジナルのまま維持することができます。認識結果は Word や Excel に転送して編集することができます。

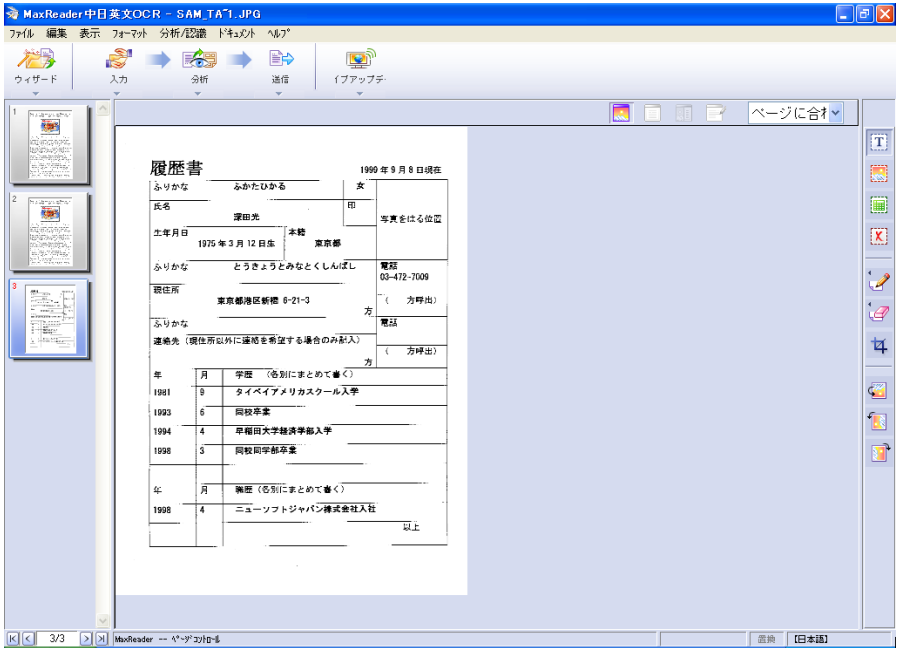
- [フォームの認識](#)
- [罫線非表示図表](#)

### フォームの認識

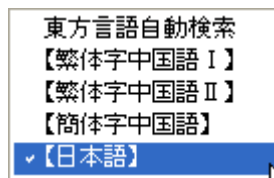
1. コマンドバー上の2つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[画像を開く]を選択します。
2. 「MaxReader 中日英文 OCR」フォルダから「Example\Japanese」ディレクトリを開き、Sam\_tab5.tif ファイルを選択します。



3. [開く]をクリックしてファイルをインポートします。

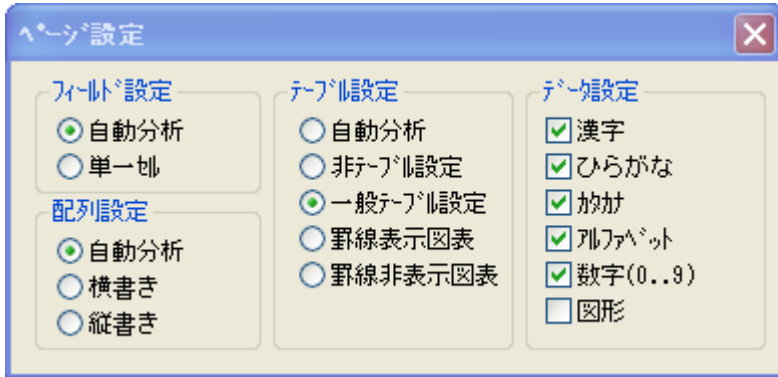







4. [フォーマット]-[認識ライブラリ設定]を選択してサブメニューから[日本語]を選択するか、ステータスバーの右側をクリックして表示されたコマンドバーから[日本語]を選択します。

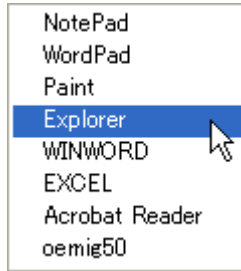


5. [フォーマット]-[レイアウト設定]を選択します。
6. [ページレイアウト]のダイアログボックスの設定を次のように調整してください。

アイテム	設定
フィールド設定	自動分析
配列設定	自動分析
テーブル設定	一般テーブル設定
データ設定	漢字、ひらがな、カタカナ、アルファベット、数字 (0..9)

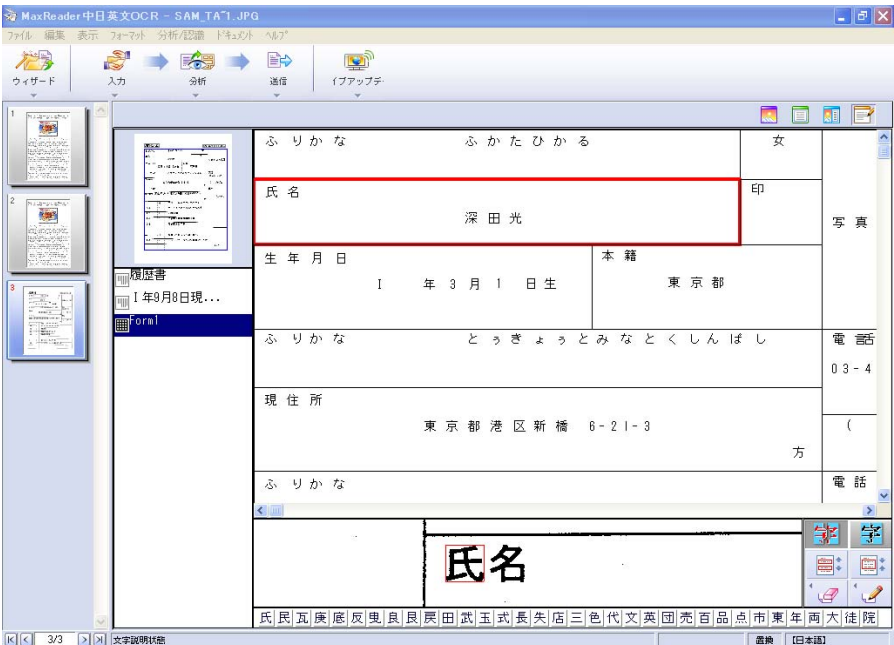


7. コマンドバー上の3つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[現在開いているページを認識]を選択します。
8. 表示ツールバーからページモード  と編集モード  を自在に切り替えて、認識結果を表示し、必要に応じて編集してください（ドキュメントの校正と編集については、第5章を参照してください）。
9. ファイルを関連するソフトウェアに転送することもできます。コマンドバーの4つ目のアイコン  をクリックし、[現在開いているページのOCRの結果をエクスポートする]を選択した後、[Explorer]を選択します。




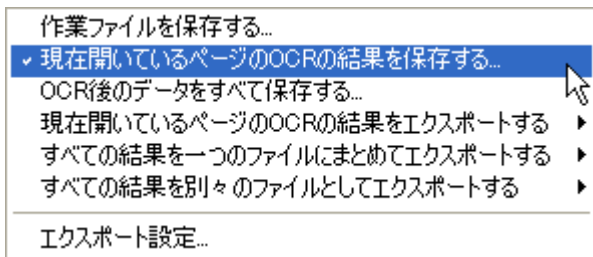
先にWebブラウザをインストールし、[転送テンプレート設定]でこれをテンプレートとして設定しておかなければ、プルダウンリストから4つ目のアイコンを選択することはできません。[転送テンプレート設定]についての詳細は、第6章— [他のプログラムに送信](#)を参照してください。

10. 認識結果は自動的に Internet Explorer Web ブラウザに転送されます。

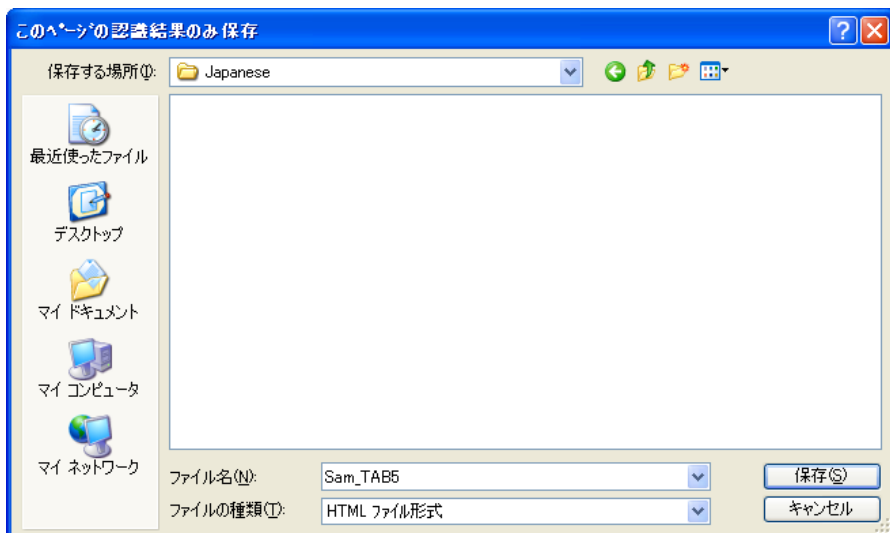


Internet Explorer に表示された認識結果

11. またファイルを保存することもできます。コマンドバー上の4つ目のアイコン  をクリックし、リストから[現在開いているページのOCRの結果を保存する]を選択します。





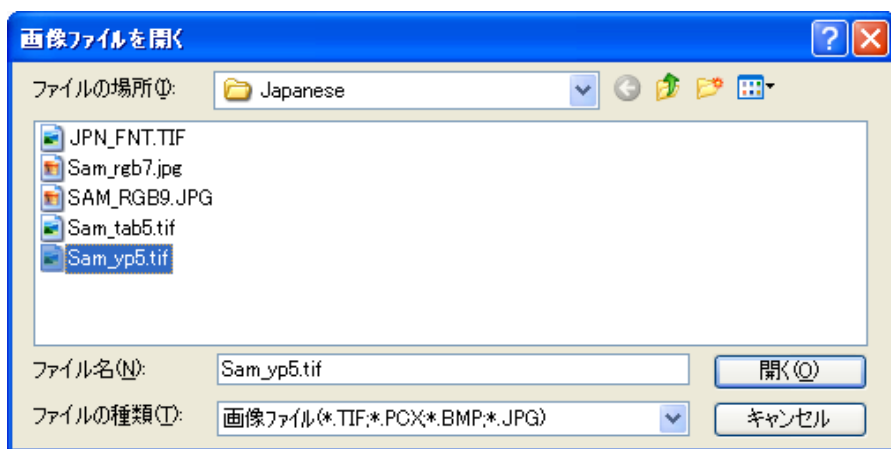
12. [このページの認識結果のみ保存]のダイアログボックスでファイル名を「Sam\_TAB5」と入力し、HTML形式で保存して[保存]をクリックします。するとファイルがHTML形式で保存され、ファイル内の画像は順番に\*.JPGファイルとして保存されます。



このページの認識結果のみ保存

## 罫線非表示図表

1. コマンドバー上の2つ目のアイコンが  になっているときには、これをクリックしてください。これ以外のアイコンの場合は、 をクリックしてリストから[画像を開く]を選択します。
2. 開いたダイアログボックスで「MaxReader 中日英文 OCR」フォルダの「Example\Japanese」ディレクトリから「Sam\_yp5.tif」ファイルを選択します。

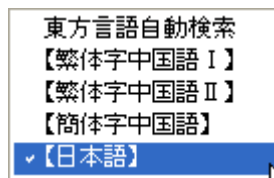


3. [開く]をクリックしてファイルをインポートします。



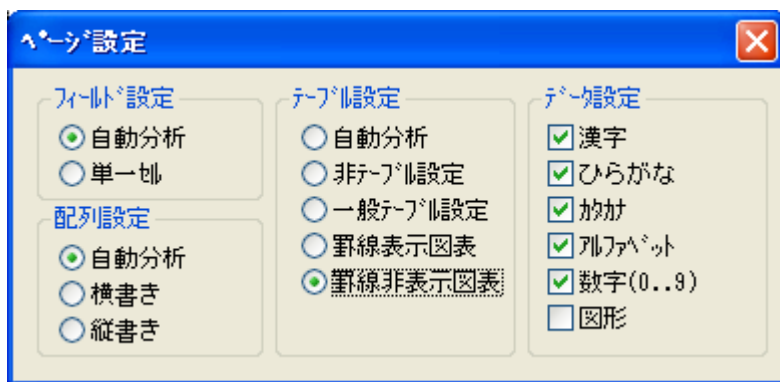



4. [フォーマット]-[認識ライブラリ設定]を選択してサブメニューから[日本語]を選択するか、ステータスバーの右側をクリックして表示されたコマンドバーから[日本語]を選択します。

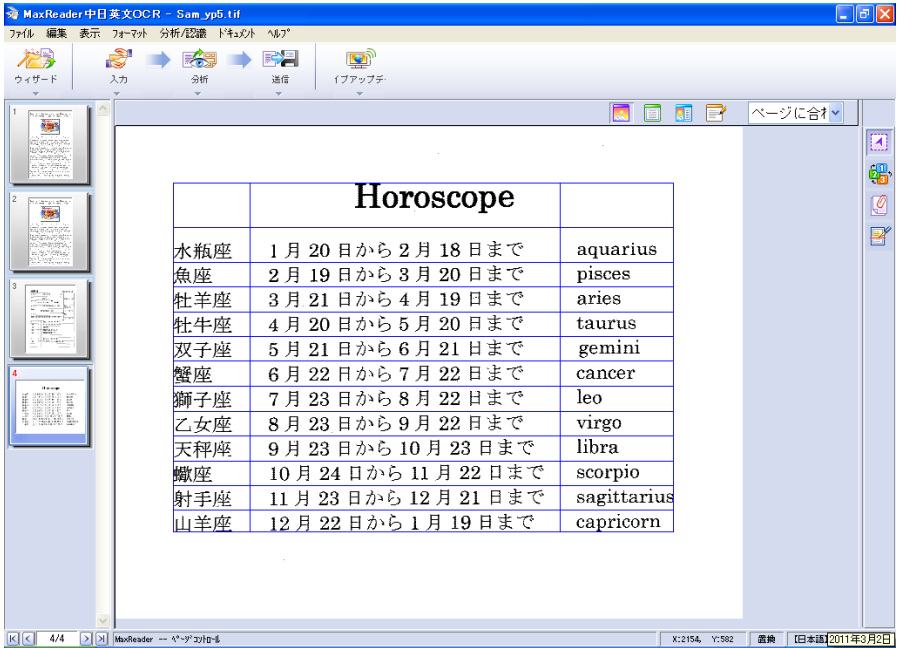





5. [フォーマット]-[レイアウト設定]を選択します。
6. [ページレイアウト]のダイアログボックスの設定を次のように調整してください。

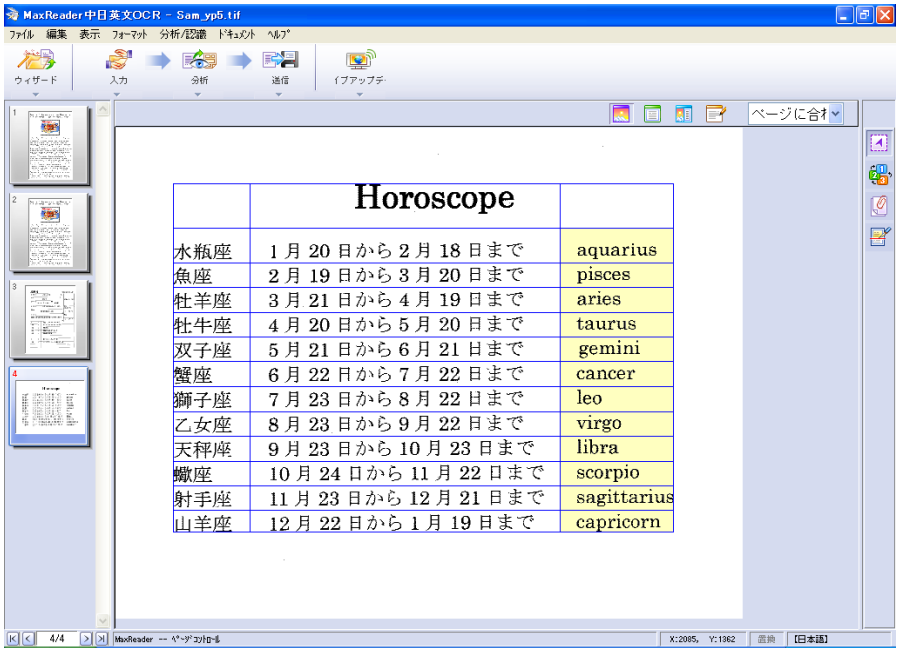
アイテム	設定
フィールド設定	自動分析
配列設定	自動分析
テーブル設定	罫線非表示図表
データ設定	漢字、ひらがな、カタカナ、アルファベット、数字 (0..9)



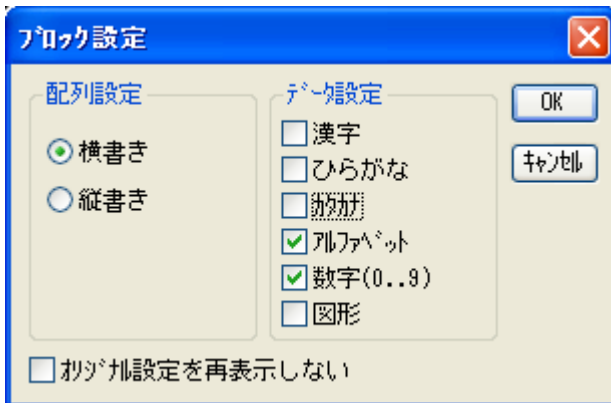
7. コマンドバーで3つ目のアイコン  をクリックし、[現在開いているページを認識]を選択すると認識が開始されます。





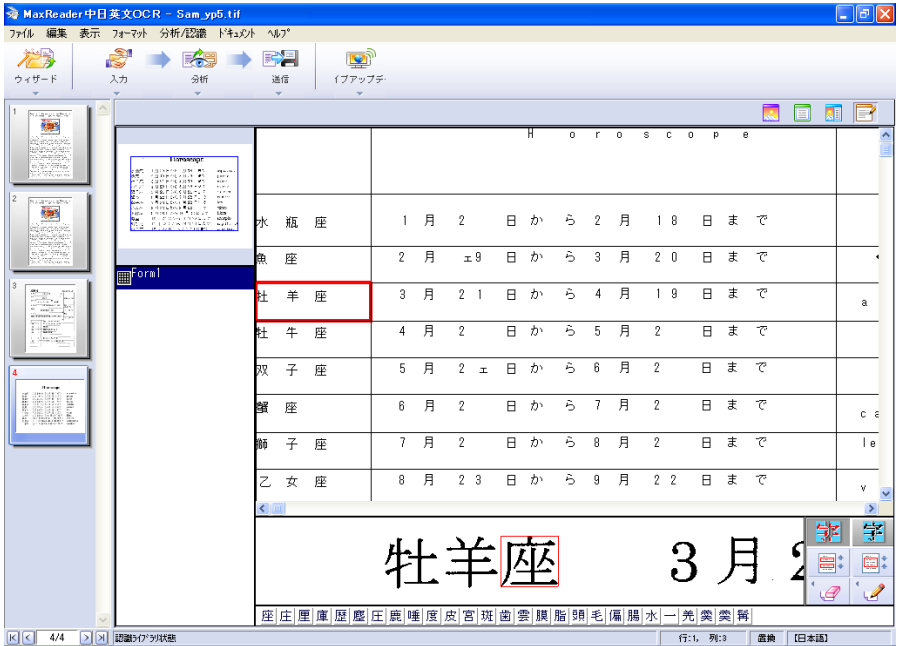
8. 表示ツールバーからページモード  と編集モード  を自在に切り替えて、認識結果を表示し、必要に応じて編集してください（編集と校正についての詳細は、第5章を参照してください）。
9. 右端の欄には漢字が含まれていませんので、ブロックの属性をリセットしておくことと認識精度を高めることができます。編集ツールバーから [ブロックをマーク]  を選択し、再認識したいブロックを選択します。選択したブロックがハイライト表示されます。




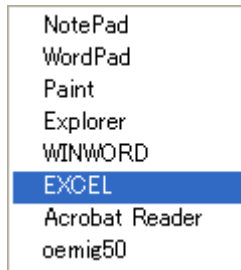
10. [分析/認識]—[ブロック再認識]を選択します。次に[データ設定]オプションで[アルファベット]と[数字]以外をクリアし、[OK]をクリックすると、選択したブロックが再認識されます。



11. 表示ツールバーからページモード  と編集モード  を自在に切り替えて、認識結果を表示し、必要に応じて編集してください（編集と校正についての詳細は、第5章を参照してください）。

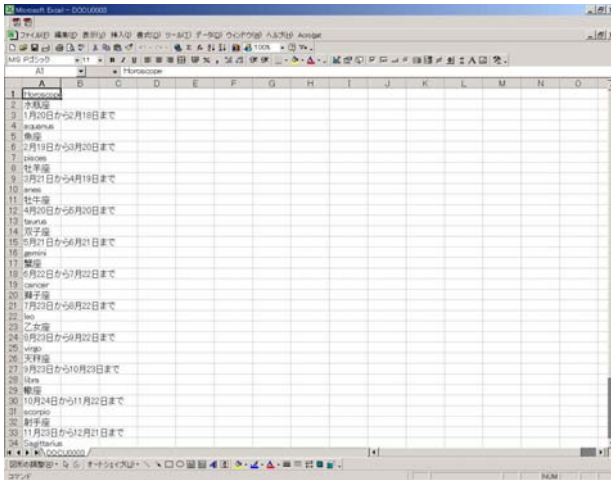


12. ファイルを関連するソフトウェアに転送することもできます。コマンドバーの4つ目のアイコン  をクリックし、[現在開いているページのOCRの結果をエクスポートする]を選択した後、[Excel]を選択します。




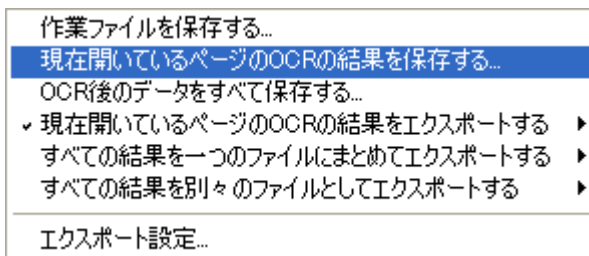
先にMicrosoft Excel をインストールし、[転送テンプレート設定]でこれをテンプレートとして設定しておかなければ、プルダウンリストから4つ目のアイコンを選択することはできません。[転送テンプレート設定]についての詳細は、第6章— [他のプログラムに送信](#) を参照してください。

13. 認識結果が Microsoft Excel に転送されます。

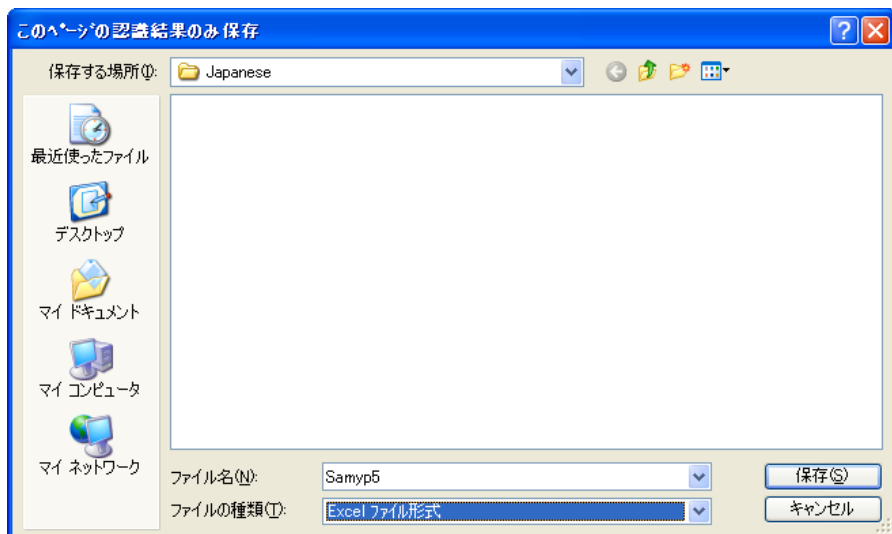


Microsoft Excel に表示された認識結果

14. またファイルを保存することもできます。コマンドバー上の4つ目のアイコン  をクリックし、リストから[現在開いているページのOCRの結果を保存する]を選択します。



15. [このページの認識結果のみ保存]のダイアログボックスでファイル名を「Samyp5」と入力し、認識結果を Excel 形式で保存します。





認識結果を Excel 形式で保存

## 第10章 自動認識のサンプル

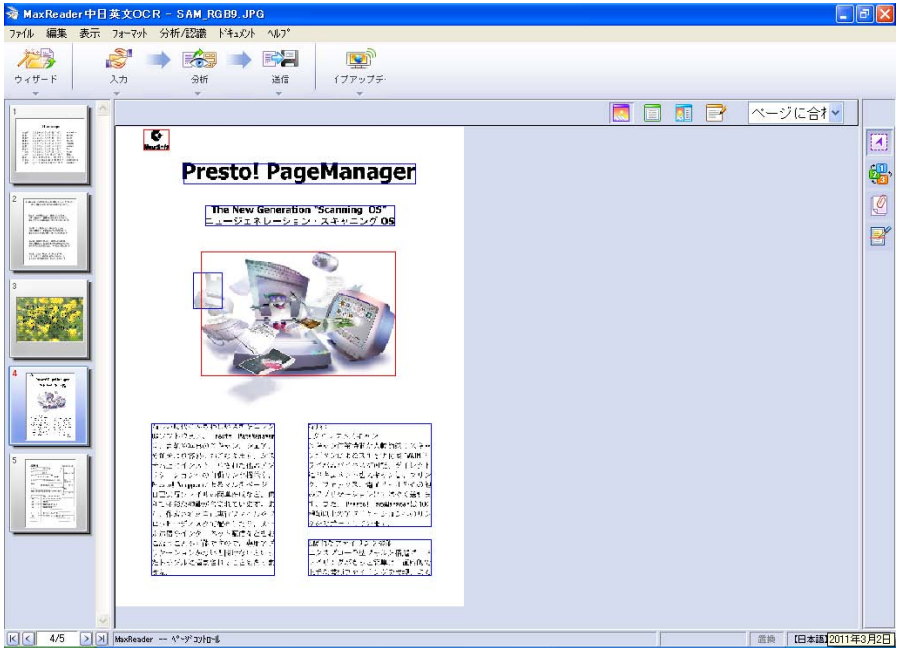
「MaxReader 中日英文 OCR」では手早く作業を自動処理することができます。この機能を活用するとインポートから認識にいたるまで、各種タスクを自動で実行し、高精度の認識を簡単に行うことができます。

1. [ファイル]ー[自動操作]を選択します。
2. [開く >>認識]を選択します。
3. 開いたダイアログボックスで「MaxReader 中日英文 OCR」フォルダから「Example\Japanese」ディレクトリを開き、Ctrl キーを押しながら SAM\_RGB9.jpg、Sam\_tab5.tif、Sam\_yp5.tif などのファイルを選択します。

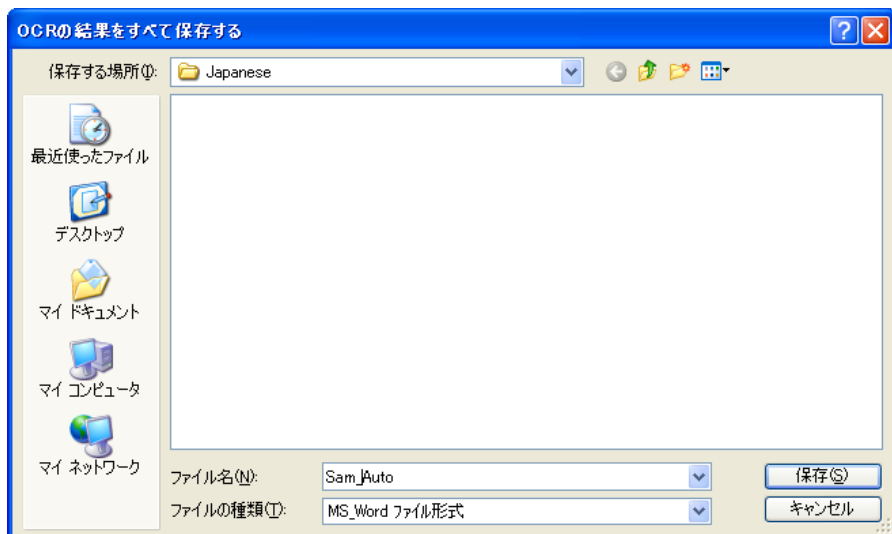


4. [開く]をクリックすると、それぞれのファイルが開き、認識されます。
5. 認識が完了したら、表示ツールバーからページモード  と編集モード  を自在に切り替えて、認識結果を表示し、必要に応じて編集してください（編集と校正についての詳細は、第5章を参照してください）。






6. 校正が完了したら、[ファイル]–[OCR後のデータをすべて保存する]を選択します。
7. [OCRの結果をすべて保存する]のダイアログボックスでファイル名を「Sam\_Auto」と入力し、MS\_Word形式で保存します。3個のファイルはページ区切りで区切られ、1個のファイルにまとめられて保存されます。



OCRの結果をすべて保存する

## 付録A 用語集

用語	説明
ブロックの属性	ページレイアウトやデータコンテンツの特長。縦書き／横書き、使用言語、フォームの有無、単一／複数セルなどがあります。
クリップボード	テキストや画像が一時的に保存される場所で、[切り取り]や[貼り付け]コマンドで貼り付けたオブジェクトを一時的に保存しておきます。
初期設定	プログラムのすべてのオプションの最初の値です。
DPI	解像度の単位です。プリンタとスキャナの解像度は1インチ毎に再現されるドットの数で測定されます。DPIが高くなると、解像度も高くなります。
メモリ	RAM (Random Access Memory) と呼ばれ、コンピュータが一時的にデータを保存する場所です。メモリのコンテンツはハードウェアやディスクにコピーすることができます。
プルダウンメニュー	メニューバーやコマンドバーの  を選択すると表示されるコマンドリストです。
ページレイアウト	ページレイアウト、配列設定、フォーム、言語など、ドキュメント内容のプロパティを設定します。
ページの分析	システムが自動的に画像ファイルの認識エリアを分析し、ブロックの順番を変更したり、ドキュメントの属性を設定したりします。

用語	説明
ブロック	認識済みの矩形ブロックのことです。
文字の分割	誤って結合された行を分割し、認識します。
文字の結合	誤って分割された行を結合し、認識します。
行の分割	誤って結合された隣り合う行の中の文字を分割し、認識します。
行の結合	誤って2行に分割された文字を結合し、認識します。
ブロックの分割	誤って結合されたブロックを分割します。
ブロックの結合	誤って分割されたブロックを結合します。
候補文字	選択した文字に形や意味が似ている文字を表示します。
登録文字の学習	よく誤認識される文字をライブラリに登録して認識精度を高めます。

## 付録B メニューコマンド

ここではメニューバー上の各コマンドとその機能について説明します。


ファイルメニュー	
新しいプロジェクトを開く	新しいドキュメントを開き、すでに開かれているドキュメントを閉じます。
プロジェクトを開く	未完成のドキュメントを開きます。
プロジェクトを保存する	未完成のドキュメントを保存します。
自動操作	ファイルをインポートし、自動的にドキュメントを認識、校正します。
自動操作エクスポート設定	自動操作で認識結果をエクスポートするとき に使用するアプリケーションとファイル形式 を設定します。
認識結果の保存	現在開いているページの認識結果を保存し ます。認識結果をdoc、rtf、html形式で保存 すると、すべての内容（画像とテキスト）が 保存されます。html形式ではファイル内の 画像に番号を付けて、それらを順番に保存 します。ファイルを書式のみ文書で保存 すると、ファイル内のテキストしか保存 されません。
OCR後のデータをすべて保存する	すべての認識結果を保存し、ページの区 切りを入れて各ドキュメントを別々に分 割します。認識結果をdoc、rtf、html 形式で保存すると、すべての内容（画像 とテキスト）が保存されます。html形 式ではファイル内の画像に番号を付け て、それらを順番に保存します。ファ イルを書式のみ文書で保存すると、フ ァイル内のテキストしか保存されませ ん。

ファイルメニュー	
画像を開く	画像ファイルを開きます。
ページの削除	ドキュメントから現在開いているページを削除します。
原稿画像の保存	現在開いているページ画像を以前のままの名前で保存します。
スキャン	スキャナの画面を開きます（詳細は、スキャナの取り扱い説明書を参照してください）。
ソースの選択	使用するスキャナを選択します。
印刷	現在のファイルを印刷します。
プリンタの設定	使用するプリンタを選択します。
転送	画像ファイルをEメールアプリケーションに転送します。
転送テンプレート設定	受け取り側のアプリケーションと、エクスポートのファイル形式を設定します。
プリファレンス	「MaxReader中日英文OCR」プログラムの環境を設定します。
終了	「MaxReader中日英文OCR」を終了します。

編集メニュー	
元に戻す	前に実行した操作を元に戻します。
切り取り	選択したエリアを切り取り、クリップボードに保存します。
コピー	選択したエリアをクリップボードにコピーします。
貼り付け	クリップボードの情報を現在のファイルに貼り付けます。
クリア	選択したエリアを削除します。
すべて選択	文書校正ウィンドウのすべてのテキストと画像を選択します。
抜き取り	選択したエリアの外側の画像を切り取ります。
反転	画像のすべての色を補色に反転します。
回転	文書校正ウィンドウの画像ファイルを（反）時計廻りに90度、180度、または角度指定で回転させます。
全ファイルのゆがみを自動補正	OCRを行う前にファイルのゆがみを自動補正します。
全ファイルの方向を自動補正	ファイルの方向を自動補正します。
検索	ドキュメント内で指定した単語を検索します。
次を検索	ドキュメント内で次の指定した単語を検索します。

編集メニュー	
置換	指定した単語を別の単語に置き換えます。
疑わしい文字を検索	現在開いているページの中で最初の疑わしい文字を検索します。
次の疑わしい文字を検索	次の疑わしい文字を検索します。



表示メニュー	
幅全体	画像を画面に合わせて拡大／縮小します。
全ページ	ページ全体を画面に表示します。
実際の大きさ(100%)	画像を実物大で表示します。
縮小表示	画面の画像を12.5%、25%、50%の倍率で縮小します。
拡大表示	画面の画像を2倍から16倍の倍率で拡大します。
拡大	画面の画像を拡大します。
縮小	画面の画像を縮小します。
画像モード	インポートした画像を表示します。
ページモード	認識結果画像／テキストファイルを表示します。
編集モード	認識結果を編集モードで表示します。   比較モードでは、取り込んだイメージのオリジナルの状態と、OCR 処理後の状態を並べて表示します。
コマンドバーを表示	コマンドバーを表示／非表示にします。
サムネイル表示	サムネイル表示を表示／非表示にします。
編集ツールバーの表示	編集ツールバーを表示／非表示にします。
クイックインプット	編集モードで頻繁に使用する記号や特殊文字をすばやく入力することができます。

フォーマットメニュー	
ページフォーマット	ドキュメントのデータのプロパティとページレイアウトを設定します。
認識ライブラリ設定	認識時にシステムが使用する認識ライブラリを設定します。
テンプレートの保存	現在の画像のブロックとレイアウト（文字校正ウィンドウのデータの属性、ブロックの位置、順番を含む）を保存します。
テンプレートを開く	保存したテンプレートを適用します。
認識結果をキャンセルする	現在の認識結果をキャンセルして再び認識します。
段落を分割しない	ブロックを1個の認識済みブロックとして維持します。このオプションを選択すると、ブロックを囲む線が単線から二重線に変わります。
フォント設定	テキストブロックのフォントを設定します。
フォントサイズの設定	テキストブロックのフォントサイズを設定します。
入力設定	編集モードで入力する文字を「挿入」するか、「上書き」するかを指定します。

分析／認識メニュー	
ページ分析	自動的に画像を分析し、OCRブロックをマークします。
現在開いているページを認識	現在開いているページを認識します。
すべてのページを認識	すべてのページを認識します。
全ページを認識ライブラリで校正	指定したライブラリで再認識します。
認識環境設定	頻繁に使用するテンプレートを将来再度使用できるように、また認識時に参考できるように設定を行います。
辞書編集設定	辞書を参考にスペルチェックを行います。
登録文字の学習	ライブラリの中の新しい文字を参考に認識します。
登録文字の削除	ライブラリから登録文字を削除します。
学習文字ライブラリ設定	参照用の学習文字ライブラリを設定します。
ブロックの再認識	選択したブロックを再認識します。
このページのみキャンセル	現在の分析と認識結果をキャンセルして、再度分析と認識を行います。
すべてキャンセル	すべての分析と認識結果をキャンセルして、再度分析と認識を行います。





ドキュメントメニュー	
最初のページ	ドキュメントの最初のページを表示します。
最後のページ	ドキュメントの最後のページを表示します。
次ページ	現在のページの次のページを表示します。
前ページ	現在のページより1つ前のページを表示します。
ページ指定	指定したページを表示します。
ページ情報	画像サイズや解像度など、現在表示しているページの情報を表示します。

ヘルプ	
ヘルプ	ヘルプの使い方を表示します。
「MaxReader中日英文OCR」について...	「MaxReader中日英文OCR」プログラム、バージョン、著作権情報を表示します。

## 付録C コマンドツールバー

コマンドツールバーには認識中によく使用するコマンドが配置されています。プルダウンメニューを開くと、操作に応じたコマンドがリストされます。通常アイコンはプルダウンメニューから選択したコマンドによって変わります（ただしいくつかの例外があります）。コマンドのアイコンをクリックすると、相当するコマンドが実行されます。




### ウィザードアイコン

アイコン	コマンドと説明
 ウィザード	<b>[スキャン&gt;&gt;認識]</b> スキャナーの画面を開き、ドキュメントのスキャンをおこないます。スキャンが完了すると、自動的にドキュメントが分析、認識されます。
 ウィザード	<b>[開く&gt;&gt;認識]</b> 画像ファイルを開き、自動的に分析、認識をおこないます。
 ウィザード	<b>[スキャン&gt;&gt;認識&gt;&gt;エクスポート]</b> スキャナーの画面を開き、ドキュメントのスキャンをおこないます。スキャンが完了すると自動的にドキュメントが分析、認識され、ドキュメントがあらかじめ指定しておいたアプリケーションにエクスポートされます。
 ウィザード	<b>[開く&gt;&gt;認識&gt;&gt;エクスポート]</b> 画像ファイルを開き、自動的に分析、認識、エクスポートをおこないます。
----	<b>自動操作エクスポート設定</b> 受け取り側のアプリケーションと、自動操作でエクスポートするファイル形式の設定を行います。

## 入力アイコン

アイコン	コマンドと説明
 入力	<b>ドキュメントのスキャン</b> スキャナに画像をインポートします。先に[ソースの選択]を選択して使用するスキャナを選択しておく必要があります。
 入力	<b>画像ファイルを開く</b> 画像ファイルを開きます。
 入力	<b>プロジェクトを開く</b> 認識、校正、編集が未完了のファイルを開きます。
----	<b>ソースの選択</b> 使用するスキャナのドライバを選択します。

## 分析アイコン

アイコン	コマンドと説明
 分析	<b>ページの分析</b> 自動的に画像を分析し、ブロックをマークします。
 分析	<b>現在開いているページを認識</b> 現在開いているページを認識します。
 分析	<b>すべてのページを認識</b> すべてのページを認識します。



アイコン	コマンドと説明
----	<b>認識環境設定</b> データのプロパティとページレイアウトを設定します。

### 送信アイコン

アイコン	コマンドと説明
	<b>作業ファイルを保存する</b> 編集途中の認識結果と画像ファイルを将来校正、編集できるように保存しておきます。
	<b>現在開いているページのOCRの結果を保存する</b> 現在開いているOCR結果を編集可能なワープロファイルとして保存します。
	<b>OCR後のデータをすべて保存する</b> すべてのOCR結果を編集可能なワープロファイルとして保存します。
	<b>現在開いているページのOCRの結果をエクスポートする</b> 現在開いているページの認識結果を指定したアプリケーションにエクスポートします。
	<b>すべての結果をひとつのファイルにまとめてエクスポートする</b> すべての認識結果を指定したアプリケーションにエクスポートして、それらを1つのファイルにまとめます。
	<b>すべての結果を別々のファイルとしてエクスポートする</b> すべての認識結果を個別のファイルに分割して指定したアプリケーションにエクスポートします。

アイコン	コマンドと説明
----	<b>エクスポート設定</b> エクスポートするアプリケーションを追加、削除、設定します。










### Live Update アイコン

アイコン	コマンドと説明
 Live Update	<b>ライブアップデート</b> NewSoft ホームページにリンクし、「MaxReader中日英文OCR」ソフトウェアをアップデートします。
 Live Update	<b>NewSoftホームページにリンク</b> NewSoft ホームページにアクセスします。








## 付録D 編集ツールバー

### 画像モード（認識前）









ツール	説明
	<b>テキストの選択</b> マウスを使って画像上をドラッグして認識したいエリアを選択します（フォームを除く）。このエリアは青い線で囲まれます。
	<b>画像の選択</b> マウスを使って認識が不要な画像を選択します。「MaxReader中日英文OCR」はこの部分を画像とみなし、認識を行いません。このエリアは赤い線で囲まれます。
	<b>表の選択</b> マウスを使って認識が必要な表を選択します。このエリアは緑の線で囲まれます。
	<b>削除</b> 選択したブロックを削除します。
	<b>ペン</b> ドキュメントで欠けた部分をペイントします。ペン先のサイズを変更するには、左上隅の  をクリックしてください。
	<b>消しゴム</b> 画像の黒い点を消します。消しゴムのサイズを変更するには、左上隅の  をクリックしてください。
	<b>抜き取り</b> マウスを使って維持したい画像の一部を選択します。[編集] - [切り取り]を選択すると、選択エリアの外側の部分が切り取られます。

ツール	説明
	<b>180度回転</b> 画像を180度回転させます。
	<b>反時計廻りに90度回転</b> 画像を左に90度回転させます。
	<b>時計廻りに90度回転</b> 画像を右に90度回転させます。

### 画像モード（認識後） / ページモード

ツール	説明
	<b>ブロックをマーク</b> 処理するブロックを選択します。
	<b>ブロックの順番</b> ブロックにはそれぞれ処理番号が付けられます。このツールを使って順番を変えることができます（詳細は、「第4章ー <a href="#">ブロックの順番を変える</a> 」を参照してください）。
	<b>ブロックの結合</b> 分割したブロックを結合し、そのブロックを再認識します。
	<b>ブロックの分割</b> 結合したブロックを分割し、そのブロックを再認識します。
	<b>文字の校正</b> ページモードで疑わしい文字を表示します。これらの文字は青い背景に黄色で表示されます。このツールはページモードでしか使用できません。

## 編集モード

ツール	説明
	<p><b>文字の分割再認識</b> 誤って結合された文字を分割し、再認識します。</p>
	<p><b>文字の結合再認識</b> 誤って分割された文字を結合し、再認識します。</p>
	<p><b>行の結合再認識</b> 誤って結合された行を分割し、再認識します。</p>
	<p><b>行の分割再認識</b> 誤って分割された行を結合し、再認識します。</p>
	<p><b>ペイント</b> ドキュメントで欠けた部分をペイントします。ペン先のサイズを変更するには、左上隅の  をクリックしてください。</p>
	<p><b>消しゴム</b> 画像の黒い点を消します。消しゴムのサイズを変更するには、左上隅の  をクリックしてください。</p>

## 付録E スキャンヒント

スキャンするとき、原稿画像の画質がスキャン結果に大きく影響します。別の画像をスキャンするときには、スキャン解像度を変える必要があります。一般的に標準ファイル（文字の高さが約 3mm）をスキャンする場合は 400dpi を使用するようお勧めします。これよりもフォントサイズが小さい場合は、解像度を上げてください。

認識精度を高めるために、認識を実行する前に消しゴムツールを使って画像にある黒い点を消してください。

認識品質には次の 3 つの要素が関係します。画質、スキャン解像度（dpi）、スキャン明度です。

- 画質：ドキュメントは正しく配置し（上部の傾き具合が 3 度以内）、画像は鮮明でなければなりません。
- スキャン解像度（dpi）：通常画像ファイルのフォントサイズが 40 から 50 ピクセルのとき、最高の結果が得られます。フォントサイズがこれよりも大きすぎたり、小さすぎたりすると、認識精度は低下します。
- たとえば新聞のフォントサイズ（3mm）であれば、400dpi でスキャンすることができます。これよりも大きいフォントの場合は、300dpi でスキャンするようお勧めします。詳しい情報は、図 1 - スキャン解像度の提案を参照してください。
- スキャン明度：画像が明るすぎたり（行が切れる - 図 4 参照）、暗すぎたりすると（文字がぼける - 図 3 参照）、認識精度が低くなります。スキャン明度を調整するときには、単純な文字（線が切れないように）と複雑な文字（ぼけないように）に特別の注意を払う必要があります。どちらの要素も該当しない場合は、優先的に単純な文字の横線を明確にしてください。

➤ [図 1：スキャン解像度の提案](#)

➤ [図 2：標準](#)

➤ [図 3：暗すぎる場合](#)

➤ [図 4：明るすぎる場合](#)

## 図 1 : スキャン解像度の提案

提案値 : 400 dpi

NewSoft Technology Corp. (10 ポイント)

NewSoft Technology Corp. (11 ポイント)

NewSoft Technology Corp. (12 ポイント)

提案値 : 300 dpi

NewSoft Technology Corp. (12 ポイント)

NewSoft Technology Corp. (14 ポイント)

NewSoft Technology Corp. (16 ポイント)

提案値 : 200 dpi

NewSoft Technology Corp. (18 ポイント)

NewSoft Technology Corp. (20 ポイント)

NewSoft Technology Corp. (22 ポイント)

図 2 : 標準

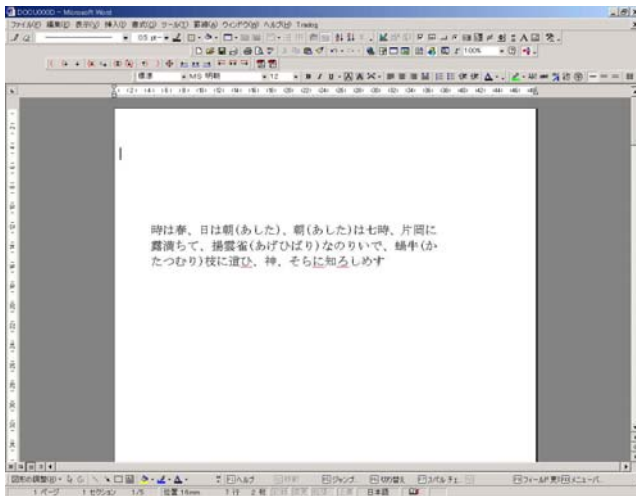
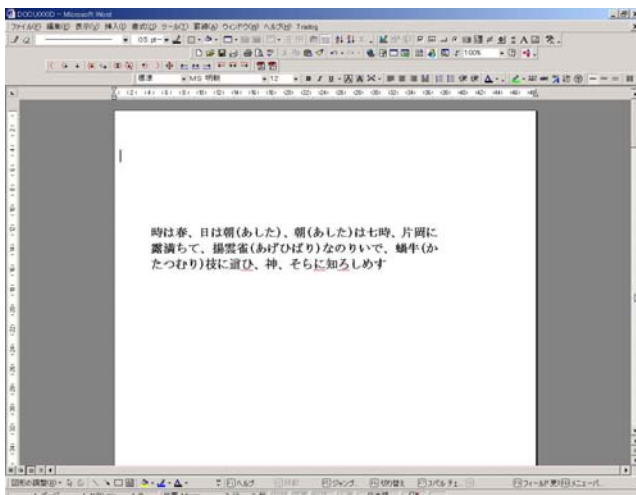
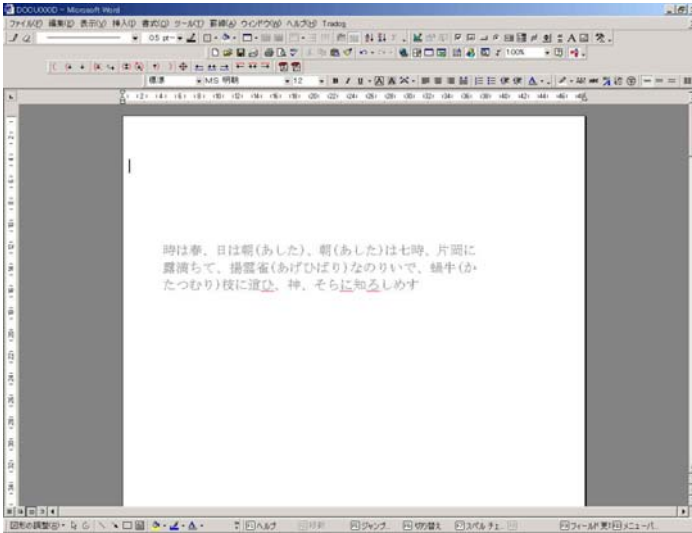


図 3 : 暗すぎる場合



## 図 4 : 明るすぎる場合



理論上はドキュメントの背景が白く、テキストの色が黒の場合、スキャンドライバの[明度の自動調整]ボタンを使ってスキャンの明度を設定することができます。ドキュメントの背景が黄色で、テキストの色が黒の場合(例：新聞)、明度を上げることができます。それでも思い通りの効果が得られない場合は、レベルを上げることができます。

一般的に「MaxReader 中日英文 OCR」の認識品質が標準に満たないようであれば、まずスキャン解像度と明度を調整してみてください。正しい明度が得られるまでには、何度もやり直す必要があるでしょう。また明度はスキャナによっても変わります。さらにブロックに漢字、ひらがな、カタカナ、アルファベット、数字が含まれているなど、ブロックの属性を正確に設定することによって、認識精度を高めることができます。