解　説

IPS

自然言語処理技術の応用

5. 文字認識における自然言語処理†

西　野　文　人††

1. はじめに

文字認識の研究・開発が進み，手書き帳票データや印刷文書などを読み取るための OCR（光学式文字読み取り装置）が実用化されている☆,☆☆．さらに，文字，図表，写真が混在した文書全体を読みとるための文書構造解析・理解技術[47]と結び付いた既存印刷物の電子ファイリング[42]や，ファックス入力なども期待されている．一方，入力デバイスの進歩などからオンライン手書き文字認識も実用的になり[28),30)]，これを利用したペンコンピュータも注目を浴びている．このように，文字認識技術を利用したさまざまな応用が期待されている．

しかし，1)日本語は文字の種類が多く，また複雑な文字や，類似文字（{口} と {ロ}，{り} と {リ} など），一文字が複数個のパターンから構成される分離文字（{加} と {力口} など）があることや，2)実際の文書では低品質文字（かすれ，にじみ，手書き文字における変形など）があり得ることから，個々の文字レベルでは文字切り出しの誤りや認識文字の誤りは避けられない．そこで文字列の言語的性質や背景知識といったさまざまな文脈情報を利用することで文字認識の読み取り精度の向上を目指す研究（文字認識後処理）が早くから行われてきた．

本稿では，以下主に日本語の一般文章を対象とした文字認識後処理について述べる．

2. 文字認識処理システムの現状

表-1に示すように，文字認識には活字文字認識，手書き文字認識，オンライン手書き文字認識がある．認識技術が異なるのは当然ながら，それらの認識技術を利用した主要アプリケーションの違いや個別文字認識精度の違いから，後処理が対象とするものも異なってくる．しかし，帳票中にも備考欄があって一般文があったり，ペンコンピュータで文章を作成することなどもあり，アプリケーションが広がるにつれて文字認識対象の違いに関係なく共通な後処理技術が必要になってきている．

一般に文字認識処理では，**図-1**に示すように，入力パターンに対してまず文字の切り出しを行い，個々の文字に対して文字の認識を行う．文字認識処理の結果としては複数の候補文字を認識距離値（その候補文字の辞書パターンと入力文字との間の類似性を示す）とともに与える．そして不使用文字の除去[38]や，認識距離値を比較した候補順位や各候補文字間の距離値の差や比などを利用して適当な候補文字数への絞り込みを行い（距離値を利用した各種の絞り込み方法の比較については文献 29)）☆，さらに言語的制約や背景知識を利用して候補文字を決定する後処理を行う．

3. 日本語文字認識後処理アルゴリズム

文が単語単位に区切られている英語のような言語では，単語単位の後処理が行われており，1)実際に辞書と照合する方法[2)]や，2)文字出現の統計情報としてたとえば binary *n*-gram を利用する方法[3)]などが提案されている．日本語においても，帳票のように氏名，住所，商品名など一記入欄が

† Natural Language Processing in Text Recognition by Fumihito NISHINO (Fujitsu Laboratories Ltd., Software Laboratory).

†† (株)富士通研究所ソフトウェア研究部

☆ 全般的な文字認識技術の動向については他の文献24),32),45),51)などを参照されたい．

☆☆ 参考文献は研究の流れが明確になるように発行年順に並べた．

☆ 類似性の尺度による評価値の大小のみによる判断は必ずしも妥当でないという指摘もある[36)]．

表-1 文字認識技術とその応用

文字認識技術	代表的応用	主な読み取り対象
活字文字認識	文書の読み取り	一 般 文
手書き文字認識	帳票の読み取り	住所，氏名，商品名
オンライン手書き文字認識	ペンコンピュータ	住所，氏名，商品名

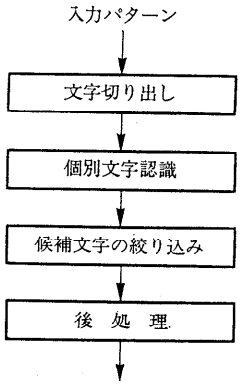


図-1 処理の流れ

一用語として扱えるものに対しては，その記入欄に依存した用語辞書を用いることが早くから行われている[6),8),9)].

一方日本語の一般文章や帳票などでも一記入欄が一用語として扱えないものでは，単語境界があいまいであることや日本語の単語の多くは1文字や2文字から成り立っていることなどから，文法的制約なども利用した後処理手法が研究されている．その手法としては，1)リジェクト文字を救済しようとするもの，2)一般文書校正支援システムを応用したもの，3)候補文字ラティスから正しい文を生成していくものがある（表-2）．

3.1 リジェクト文字救済方式

これは，個別文字認識の確信度の高い文字は候補を一つに絞り込むが，確信度の低い文字に対しては候補文字を一つに絞り込まずにリジェクト文字とし，このリジェクト文字のみの救済を試みようとするものである．2文字の連接情報を利用してリジェクト文字を含むすべての組合せについて接続可否を検定する手法[7)]や，リジェクト文字に対して候補文字集合から文字を選択して候補文を生成し，この候補文に対して単語辞書と照合したり単語の構文的制約から正しい文字を選択する手法[4),10)]が提案されている．

この方式では，1)リジェクト文字の数が多くなると候補文字置換えの組合せが増大して処理時間が増大する，2)リジェクト文字以外に認識誤りがあっても候補文字を置き換えないので正しい結果を見つけることができない，といった問題がある．したがって個別認識率が高く，リジェクト文字も正確に判定できることが前提となるが，手書き文字や，活字でもコピーなどによる低品質文字の個別認識率は十分とは言い難く，またリジェクト文字を正確に判定することの難しさも残している．

3.2 校正支援システムを応用した方式

これは，文字認識結果から得られる1位候補の文字列に対して，校正支援機能を応用して誤りを検出し，誤り指摘個所の近傍で次の候補との類似度が近い文字を次の候補に置き換えて再度校正支援機能により誤りを検出するというものである[41)]．

リジェクト文字として救済する文字を固定しない分柔軟ではあるが，やはり高い個別文字認識率が前提となっている．

3.3 候補文字ラティスからの単語列生成方式

文字認識結果は，各文字ごとに正解の可能性の高い候補文字から順に候補順位をつけて並べる

表-2 文字認識後処理方式

方 式	救済対象範囲	救 済 方 法
リジェクト文字の救済	リジェクト文字のみ	文の正当性を検査し，不適ならリジェクト文字を別の候補文字に置き換える
校正支援システムの応用	校正支援システムが誤りを検出した部分	誤り指摘箇所の近傍で候補文字を置き換えて再検査
候補文字ラティスからの単語列生成	全 体	候補文字ラティスから用語・構文的に正しい組合せを作り上げていく

第1位候補　一般的在意志疎通
第2位候補　メ股酌を章忘政適
第3位候補　ノ船曲な煮吉珠逸
第4位候補　J投灼社葦走球速

図-2　候補文字のラティス

と，候補文字のラティス（**図-2**）になる．本方式は，単語辞書と文法規則を参照しながら候補文字ラティスから適当な文字を選択することで正しい日本語文章を構築していくというものである[12),27),43)]．これには日本語形態素解析技術が応用されるが，単語境界のあいまい性の問題に加えて，各文字位置に対して文字候補が複数存在するという問題がある．そこで単語照合方法や単語列探索法に通常の形態素解析とは異なった手法が必要になってくる．

3.3.1　単語辞書との照合方法

単語照合方式としては，1)候補文字ラティスの各文字位置の候補文字を組み合わせてできる文字列が単語として存在するかどうかを辞書探索する候補文字主導型の方式と，逆に2)辞書に存在する単語に対して候補文字ラティスと照合する辞書主導型の単語照合方法がある．

候補文字主導型の単語照合では，単純にすべての組合せの照合を行ったのでは，各文字に対する候補文字数をm，最長単語長をlとすると，$\sum_{k=1}^{l} m^k$の辞書照合が必要になってしまう．文字種の変化点を見て文節ごとに分割することでlを小さくすることも行われているが，候補文字が複数の文字種に分かれている場合や「さ迷う」のような文字種にまたがった単語の存在の問題がある．しかし辞書構造としてトライ（trie）構造☆を採用[43)]して候補文字ラティスの要素のいずれかと適合する辺をたどっていくことで，無駄な照合を行わずにすべての長さの単語を検索できる．

辞書主導型の単語照合では，辞書の単語数が増加すると単語照合時間が増加するので，同一の先頭文字をグループ化する[26)]，などによって単語照合数を減らすことが行われている．また先頭文字から順に各位置において同一文字をグループ化すればこれはトライ辞書（前方一致圧縮辞書[27)]もトライの一種と考えられる）になり，高速な検索が可能である．

☆ トライ構造については文献53)などを参照のこと．

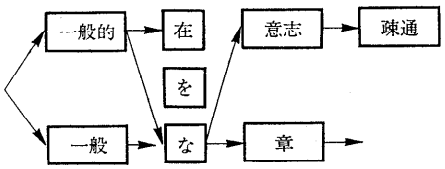


図-3　探　索　木

3.3.2　単語列の探索手法

一般文字を対象とした文字認識後処理では，単語照合による検査だけでなく構文的制約を利用した検査も行われる（**図-3**）．文節内の品詞構造を確率付き有限オートマトンで解析するもの[44)]や，各単語の頻度と品詞間の遷移確率と各文字候補の認識結果から得られる確信度とから計算されるコストに基づいて，最良優先探索を利用して一つの候補だけを導くもの[27)]，ビームサーチを基本として複数の解を得てパスのあいまい性を評価するもの[23),43)]が報告されている．

4.　日本語文章の認識後処理の問題点

日本語の一般文章の後処理を行っても誤読文字を検出・修正できないことがある．また時には，隣接する文字を誤訂正してしまって認識率の低下をもたらしてしまうこともある．その主な理由としては以下のものをあげることができる．

1. たまたま単語辞書に登録されている単語と照合し，文法的エラーも検出されなかった（特に1文字単語）．
2. 入力文中に辞書未登録語があった．
3. 候補文字集合の中に正解文字が存在しなかった．
4. 文字の切り出しが誤っていた．

4.1　後処理のための単語辞書と文法規則

日本語形態素解析処理は機械翻訳をはじめとするさまざまな日本語処理に使われているが，それらの多くは科学技術文書などの限られた種類の正しく入力された日本語文を解析することが目的であって，不適格な文が入力された場合のことは特に考えられていないことも多い．これに対して文字認識後処理では，さまざまな文書が読み込まれるので多様な言語現象に対して処理する能力をもたなければならないという要求があるのと同時に，さまざまな制約を利用して正しい候補文字を見つけるためには制約をきつくして不適格な文は

受理しにくく☆しなければならないという要求もある．単語辞書に関して言えば，さまざまな文を受理するあるいは正しい候補文字に置き換えるためには多くの単語を登録しておかなければならないが，しかし辞書の単語登録数を単純に増やしただけでは今度は認識結果が誤っていてもたまたま辞書に存在する単語と重なって候補文字の誤りを救済できなくなってしまう可能性も高くなってしまう．

このようなことから，文字認識後処理では単語を通常の品詞分類より細分類化し[12]，文法規則も接続の可否を判断するだけではなく品詞推移確率のような形式で文法自身に確率を付与する[27),43),44)]ことにより単語列探索のパス選択のときに利用するコストの一つとして取り入れることが行われている．また，特に強い制約をもたない1文字単語のようなものは，より詳細な分類やヒューリスティック規則によって特別扱いされることも多い[43),44)]．

4.2 未登録語処理

未登録語をどう発見して確定するかは，一般の日本語形態素解析でも大きな問題であり，さまざまな手法が提案されている[25)]．一般の形態素解析では入力文字列は正しい文と仮定されるので解析がうまくいかなかった場合には未登録語の可能性を強く疑うことができるが，文字認識の後処理では解析がうまくいかなかったのは未登録語の存在のせいなのか，個別文字認識がうまくいかなかったせいなのかが分からない．また未登録語があることが分かったとしても，1立候補文字が正解文字とは限らないので，その未登録語がどういう文字列からなっているのかがはっきりとしない．未登録語テンプレートを使って未登録語を発見し，文字の連接確率を使って文字列を限定する[27)]ことなどが行われている．

4.3 候補外文字に対する対処

個別文字認識精度が低くなると候補文字集合中に正解文字が入っていない場合もあり，救済不能あるいは誤訂正の一つの要因となっている．その対策としては，候補文字補完と類似検索（単語部分照合）がある．

候補文字補完とは，あらかじめ文字認識が誤り

☆ もともと誤りをもった文が読み込まれているかもしれないので，完全に拒絶はできない．

入力文字列　デ　コ　タ　ル

↓　↓　↓　↓

w11　w13　w11　w11

w12　w25　w12　w32

w13　　w32

w51

a）各文字位置から単語を検索する

単語コード	単　語	判定値
w11	デジタル	3
w12	データ	2
w13	デコーダ	2
w32	レンタル	2

b）判定値（この例では一致文字数）を集計する

図-4　類似検索の例

そうな類似文字の組（たとえば，「力」と「カ」）を定義しておき，文字認識結果の候補文字に対して無条件にこの類似文字を候補として付け加えてから後処理を行おうというものである[12),26)]．

類似検索は，単語辞書との照合を行う際に完全に一致したものだけではなく類似した（部分的に一致した）単語も検索しようというものである[5)]．住所や商品名のように用語が限定されている場合には，辞書の全単語と照合を行って，その中から一致文字数の多いものや距離値の合計の小さいものを選ぶことが行われている[22)]．また，全単語との照合ではなく，各文字位置の文字から単語へのインデックスを保有することで高速化を図り(図-4)[16)]，判定値には文字認識結果などから算出される各候補文字の確からしさなどが用いられている[16),39)]．

しかし，このような類似検索アルゴリズムは住所や商品名などの単語の長さが長い場合には有効であるが，一般的な日本語単語を対象とした場合には単語長が短いので類似単語が非常に多く検索されてしまうという問題がある．そこで，短い単語に対しては完全一致による照合で，比較的長い単語に対してのみ類似検索も行われている[27),38)]．しかし，一般文の類似検索は，処理速度と精度の面であまり効果があがってはいない．

4.4 文字切り出し誤りに対する対処

漢字には分離文字が多く含まれており，特に枠なしの手書き文字列認識では文字を正しく切り出すための多くの努力がなされている．言語的情報を利用したものとしては，複数の文字の切り出し候補の中から文字連接情報や単語照合などにより正しい切り出しを見つけようとするいくつかの試みがある[13),18),35)]．また，記載内容に大きな限定

がある郵便物の住所を対象にして，文字切り出し，文字認識，単語照合をトップダウン的に統合する方式も提案されている[20]．

5. 文字認識後処理の課題

5.1 精度向上

現在のところ，個別文字の認識率が90%程度以上の高い認識率であれば，後処理によって95%程度以上に認識率を引き上げることができているがまだ十分な精度とは言えない．また，個別文字認識の能力がかなり上がってきたとはいえ，まだ乱雑に書かれた手書き文字，コピーなどによってかすれた文字などの認識率は低い．このように個別文字認識率が低くなると現在の後処理システムでは後処理効果はほとんど期待できない．しかし，人間はさまざまな情報を利用することでかなり低い認識率の文章でさえも正解を推測することが実験されている[15]．この差は，現在の後処理がほとんど形態素解析レベルの言語処理に留まっていることによるもので，もっとさまざまな知識を利用した高度な処理技術が必要であることを示している．

5.1.1 共起関係の利用

構文的な制約だけでは「入口」と「人口」のように同じ品詞の単語間の誤読を修正することはできない．そこで，「入口—入る，作る，狭い，…」「人口—増える，増加する，多い，…」のような関連語を記述した共起辞書を用意することで，誤読の修正を図ろうとしているものもある[44]．ただし現在は共起辞書が利用されている範囲は限られており，辞書の整備が課題となっている．

5.1.2 特定知識の利用

入力対象を特定した知識処理はかなり以前から試みられており，手書きのFortranプログラムに対する知識処理[1]，帳票の氏名欄のふりがな情報を利用したもの[22]，住所構造知識を利用したもの[34]などが報告されている．一般文章の中にも，住所が書かれていたり，振り仮名（ふりがな）が振られていたりすることがある．このような特定知識を集め，どの部分にどの特定知識が適用できるのかを見つけ出して，その知識を適用するということが必要である．

5.1.3 大局処理

局所的な制約だけでなく大局的な観点からの情報も利用しようという試みとしては，文書の内容に含まれている情報をキーワードとして抽出してこの知識を利用しようという報告[37]がある程度である．文書の分野を推定しての処理や文書構造情報の利用などが今後期待されている．

5.1.4 距離情報と言語情報との統合

文脈を利用した後処理も完全ではないので，あまり文脈情報に頼り過ぎると，個別文字認識は正しかったのに，後処理で誤訂正してしまうということがおこる．個別文字認識から与えられる確からしさと，文脈情報処理の確からしさの双方を利用することが必要になる．

文字認識の距離情報に関しては，確信度をきっちりとした値に変換する手法が検討されている[17),36]．一方，日本語の形態素解析においても確率的な手法を採り入れようという試みもある[14),33]．そして，文字認識から与えられる尺度である距離値と言語情報から与えられる尺度とを統一して取り扱おうという試みもある[11]．しかし，大量の言語データに対して信頼できる言語情報を与えることの難しさなどから実際の後処理システムでは経験に基づく評価式が用いられることが多い．言語データの整備が望まれている．

5.2 多様な文書への対応

これまでの自然言語処理ではごく限られた種類の文を対象にしてきた．しかし，文字認識を使って読み込ませたい文書は，外国語が挿入された文書（論文の要約，参考文献など），ひらがなやカタカナばかりの文書（小学校の教科書など），旧仮名使い，文語，話し言葉などを含む文書などさまざまである．これらも扱えるようにするための辞書整備と文法整備が必要とされている．

5.3 ユーザインタフェース

後処理を含めた文字認識技術がいくら向上しても，100%の認識率は望めない．したがって完全な結果を得るためには人間による修正作業が必要になる．多くのOCRシステムで，認識結果に対して十分な確信度が得られない文字は利用者に警告して，候補文字を表示して選択させるようなユーザインタフェースをもっている．候補表示のレイアウトや選択方法の検討[40]や，人手による訂正履歴を自動的に辞書や他の部分の認識結果に反映する研究[50]もあり，今後このような認識結果の確信度に基づいた表示の仕方，訂正の指示の（単

に文字レベルでの修正ではない）仕方，訂正履歴の学習，辞書などの管理といった統合的にユーザの使い勝手を良くするための実用性を高める研究が望まれている．

6. おわりに

文字切り出しや個別文字認識機能はモジュール化され，これらをソフトウェアだけで処理できるシステムも出始め，文字認識をツールとして考えることができるようになってきた．これによって，他のアプリケーションとの連携も容易になってきた．これまでの文字認識に関する研究・開発は文字の読み取り精度の向上に主眼が置かれてきたが，これからは単に文字認識精度の向上だけでなく，文字認識をどのように利用するかという観点から後処理を含めたシステムとして考える必要があろう．

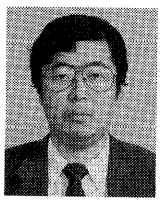
研究開発のための環境を整備することも重要な課題である．研究開発者が文書をスキャナから読み込ませて文字認識をかけてから後処理の実験を行っていたのでは，大量の文書に対する実験はなかなかできない．文字認識実験のためのデータベースもいくつか報告されている[46]が，特に後処理の実験を行うためには単なる文字イメージデータだけではない文書全体のデータベース[48]の構築が必要である．また，ユーザが容易にさまざまな種類の OCR 装置を自分のパソコンやワークステーション上から利用できるようなソフトウェア環境[49]の整備も重要である．このような環境では複数の OCR 結果を比較することによりエラー率を減少できるという voting アルゴリズムも開発されている[31]．日本ではまだ文字認識をツールとして利用する環境があまり整っていないが，文字認識をツールとして利用できる環境が揃えば，さまざまなアプリケーションと連携される．そのとき，精度の向上を含めて，自然言語処理への期待はますます大きくなることであろう．

参考文献

1) Duda, R.O. and Hart, P.E.: Experiments in the Recognition of Hand-Printed Text: Part II—Context Analysis, *AFIPS Conference Proceedings*, Vol. 33, pp. 1139-1149 (1968).
2) 阿部圭一，秦野和郎，福村晃夫：辞書を利用する文字認識系の能力の評価，電子通信学会論文誌，Vol. 52-C, No. 6, pp. 305-312 (1969).
3) Riceman, E.M. and Hanson, A.R.: A Contextual Postprocessing System for Error Correction Using Binary n-Grams, *IEEE Transactions on Computers*, Vol. C-23, No. 5, pp. 480-493 (1974).
4) Kawada, T., Amano, S. and Sakai, K.: Linguistic Error Correction of Japanese Sentence, *COLING 80*, pp. 257-261 (1980).
5) Hall, P.A.V. and Dowling, G.R.: Approximate String Matching, *ACM Computing Surveys*, Vol. 12, No. 4, pp. 381-402 (1980).
6) 印牧直文，中島健造，荒川弘熙：単語辞書を活用した文字認識法の一検討，信学技報，PRL 81-91, pp. 69-76 (1981).
7) 杉村利明，斉藤珠喜：文字連接情報を用いたリジェクト文字の判定処理（文字認識への応用），信学技報，PRL 81-105, pp. 73-79 (1981).
8) 飯田行恭，杉村利明：パターン認識における単語照合処理の一検討，信学技報，PRL 82-77, pp. 93-98 (1982).
9) 蕪山幸和，菅原秀明，山本栄一郎，中西道明：手書き漢字認識における単語情報の利用，昭和 57 年度電子通信学会総合全国大会 1341, pp. 5-326 (1982).
10) 新谷幹夫，梅田三千雄：複合後処理法による文字認識精度向上の評価，信学技報，PRL 83-42, pp. 25-34 (1983).
11) 長田一興，牧野 保，日高 達：日本語の文脈情報を用いた文字認識，電子情報通信学会論文誌，Vol. J 67-D, No. 4, pp. 520-527 (1984).
12) 池田克夫，大田友一，上野恵美子：手書き原稿認識における語彙および構文の検定，情報処理学会論文誌，Vol. 26, No. 5, pp. 862-869 (1985).
13) 村瀬 洋，新谷幹夫，若原 徹，小高和己：言語情報を利用した手書き文字列からの文字切り出しと認識，電子情報通信学会論文誌，Vol. J 69-D, No. 9, pp. 1292-1301 (1986).
14) 松延栄治，日高 達，吉田 将：確率文節文法による形態素解析実験について，九大工学集報 59-6, pp. 799-804 (1986).
15) 西野文人：文字認識後処理の可能性，自然言語研究会 62-10, pp. 69-76 (1987).
16) 松尾比呂志，佐藤哲司，津田伸生：連想統合型照合による単語あいまい検索，第 34 回情報処理学会全国大会 4E-7, pp. 1845-1846 (1987).
17) 瀬川英生：複合類似度法における類似度値の分布について，信学技報，PRU 87-18, pp. 1-10 (1987).
18) 西野文人，高尾哲康：日本語文書リーダ後処理の実現，自然言語研究会 64-6, pp. 45-52 (1987).
19) 津雲 淳，浅井 紘：文字認識技術の最近の動向，信学技報，IE 88-5, pp. 31-38 (1988).
20) 佐瀬慎治，辻 善丈，津雲 淳：制限付文字列読み取りの一検討，信学技報，PRU 88-115, pp. 49-56 (1988).
21) 鈴木 章，宮原末治，小橋史彦：住所認識装置の選択後処理方式，信学技報，PRU 88-154, pp. 57-64 (1988).
22) 鈴木 薫，麻田治男：ふり仮名付き単語の文字認識後処理方式，第 36 回情報処理学会全国大会 6V-

7, pp. 1795-1796 (1988).
23) 黒澤由明：日本語文章を対象とする文字認識後処理方式，第36回情報処理学会全国大会，7V-2, pp. 1801-1802 (1988).
24) 坂井邦夫：文字・文書の認識と理解，電子情報通信学会誌，Vol. 71, No. 11, pp. 1182-1191 (1988).
25) 吉村賢治，武内美津乃，津田健蔵，首藤公昭：未登録語を含む日本語文の形態素解析，情報処理学会論文誌，Vol. 30, No. 3, pp. 294-301 (Mar. 1989).
26) 杉村利明：候補文字補完と言語処理による漢字認識の誤り訂正処理法，電子情報通信学会論文誌，Vol. J 72-D-11, No. 7, pp. 993-1000 (1989).
27) 高尾哲康，西野文人：日本語文書リーダ後処理の実現と評価，情報処理学会論文誌，Vol. 30, No. 11, pp. 1394-1401 (Nov. 1989).
28) Nakagawa, M.: Non-Keyboard Input of Japanese Text (On-Line Recognition of Handwritten Characters as the Most Hopeful Approach), *Journal of Information Processing*, Vol. 13, No. 1, pp. 15-34 (1990).
29) 磯山秀幸，木谷 強：OCR の認識結果に対する文字認識後処理方式の検討，第40回情報処理学会全国大会 2E-3, pp. 329-330 (1990).
30) Tappert, C. C., Suen, C. Y. and Wakahara, T.: The State of the Art in On-Line Handwriting Recognition, *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol. 12, No. 8, pp. 787-808 (1990).
31) Bradford, R. and Nartker, T.: Error Correlation in Contemporary OCR Systems, *Proc. First International Conference on Document Analysis and Recognition*, pp. 516-523 (1991).
32) 坂井邦夫：文書認識・理解システムの基礎技術，第27回東北大学電気通信研究所シンポジウム論文集，pp. 39-49 (1991).
33) 丸山 宏，荻野紫穂，渡辺日出雄：確率的形態素解析，日本ソフトウェア科学会第8回大会論文集，E 2-1, pp. 177-180 (1990).
34) 磯山秀幸：住所文字列に対する文字認識後処理方式の検討，自然言語研究会 82-3 (1991).
35) 小林弥生，津雲 淳：文字連接情報を利用した手書き文字列認識，信学技報，NLC 91-32/PR/PRU 91-67, pp. 39-46 (1991).
36) 阿曽弘具，越後和徳，木村正行：正確な個別文字認識の検討，信学技報，NLC 91-33/PRU 91-68, pp. 47-54 (1991).
37) 丹羽寿男，萱島一弘，〆木泰治：文字認識後処理法と後処理による効果の分析，信学技報，PRU 91-135, pp. 71-78 (1991).
38) 木谷 強：手書き文書の文字認識結果に対する後処理方式，自然言語研究会 86-1 (1991).
39) 小黒雅己，中村 修，北村 正：手書き複合語文字列識別のための最適単語組探索方式，電子情報通信学会論文誌，Vol. J 75-D-II, No. 1, pp. 96-102 (1992).
40) 宮脇俊一，宮原末治：認識結果に対する候補文字の表示と選択に関する検討，1992年度電子情報通信学会秋季大会 D 327 6-329 (1992).
41) 山中紀子，出野崎康雄，齋藤裕美，小林賢一郎：日本語テキストリーダにおける日本語校正支援機能，第44回情報処理学会全国大会 3Q-1, Vol. 3, pp. 243-244 (1992).
42) (財)関西文化学術研究都市推進機構：学術研究支援のための高度情報システムに関する研究（平成4年3月）.
43) 伊東伸泰，丸山 宏：OCR 入力された日本語文の誤り検出と自動訂正，情報処理学会論文誌，Vol. 33, No. 5, pp. 664-670 (May 1992).
44) 紺野章子，本郷保夫：日本語 OCR の後処理に関する一手法，信学技報，PRU 92-21, pp. 23-30 (1992).
45) Mori, S., Suen, C. Y. and Yamamoto, K.: Historical Review of OCR Research and Development, *Proc. of IEEE*, Vol. 80, No. 7, pp. 1029-1058 (1992).
46) Nagy, G.: At the Frontiers of OCR, *Proc. of IEEE*, Vol. 80, No. 7, pp. 1093-1100 (1992).
47) Tsujimoto, S. and Asada, H.: Major Components of a Complete Text Reading System, *Proc. of IEEE*, Vol. 80, No. 7, pp. 1133-1149 (1992).
48) Nartker, T., Bradford, R. and Cerny, B.: A PRELIMINARY REPORT ON UNLV/GT 1: A Database for Ground-Truth Testing in Document Analysis and Character Recognition, *Proc. Symposium on Document Analysis and Information Retrieval*, pp. 300-315 (1992).
49) Rice, S.: The OCR Experimental Environment, Technical Report, Information Science Research Institute, University of Nevada (1992).
50) 村木一至，浜田和彦：OCR の認識誤り訂正に於けるテキスト適合性の評価，信学技報，NLC 92-27/PRU 92-41, pp. 47-52 (1992).
51) 藤澤浩道，栗野清道，嶋 好博：OCR における手書数字認識の技術と応用の動向，電子情報通信学会「手書き文字認識技術の過去・現在・未来」シンポジウム講演論文集，pp. 46-53 (1993).
52) 津雲 淳：手書き漢字の OCR の開発動向と今後，電子情報通信学会「手書き文字認識技術の過去・現在・未来」シンポジウム講演論文集，pp. 72-77 (1993).
53) 青江順一：トライとその応用，情報処理，Vol. 34, No. 2, pp. 224-251 (Feb. 1993).

（平成5年5月20日受付）



西野 文人（正会員）

1956年生．1979年東京工業大学理学部情報科学科卒業．1981年同大学院修士課程修了．同年(株)富士通研究所入社．機械翻訳をはじめとする自然言語処理の研究・開発に従事．ACM 会員．