

ペン入力校正支援システム

山田洋志 福島俊一 竹元義美
NEC 情報メディア研究所

1 はじめに

校正支援システムは新聞社や出版社など専門家向けに実用化されてきた[1, 2]。これらは、大量の文書を専門家が検査することを前提に設計されている。一方で、一般ユーザ向けの文章検査システムがあり、最近では文章検査機能がメジャーなワープロの標準機能として取り込まれている[3]。

筆者らは、一般ユーザを対象としたペン入力対応のWindows版校正支援システムを試作した。試作にあたっては、(1)Windows上で稼動、(2)ペン入力対応、(3)一般ユーザが使いやすい文章検査機能、の3点を方針とした。

Microsoft Windows 3.1は、すでに多くのユーザを獲得していることと、MS-DOS環境と比べてメモリ等のハードウェア制約が緩いことからターゲットとした。また、DLL(Dynamic Link Library)を利用して機能を部品化することで、他のアプリケーションから文章解析機能を利用できる。

入力デバイスとしては、キーボード、マウスのほかにペンを使えるようにした。ペン入力は紙と鉛筆に近い感覚で利用でき、人間の思考を助けるものとして注目されている。ただし、文章の入力ではキーボードに速度で劣っているという問題がある。そこで、本システムでは、文章の一部を入力すると自動的に残りを予測する、予測ペン入力方式[4, 5]を利用して効率向上を図っている。

専門家向けのシステムでは、誤りの可能性の低い箇所を検出してにおいて最終判断をユーザ(専門家)に任せることができた。しかし、一般ユーザが使用する場合には、指摘箇所が多すぎると検査機能を使う意欲をそいでしまう。そこで、専門家向けとは別に一般ユーザ向けの校正支援機能を考える必要がある。

本稿では、2節で試作したシステムの概要を紹介し、3節でペンに対応した操作、4節で筆者らの開発した文章入力方式である予測ペン入力方式、5節で文章検査機能を紹介する。6節では、形態素解析ライブラリを利用するためのツールについて述べる。

2 システムの概要

本システムはWindows 3.1上で動作する。ペンを使用するためには、ペン入力用のタブレットを接続するか、あるいは、ペンパソコン(PC-9801Pなど)を利用する。

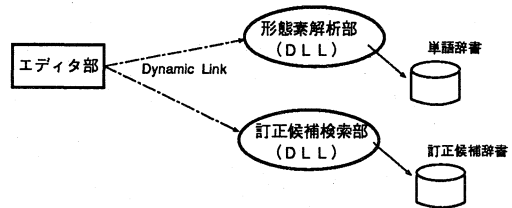


図1: ソフトウェア構成

用する。ペン対応には、Microsoft Windows for Pens 1.1とそのSDKを使用した。

本システムのソフトウェア構成を図1に示す。形態素解析部は、入力テキストを形態素解析し、文章中の検出箇所を判定する。訂正候補検索部は、検出語に対する訂正候補を訂正候補辞書から検索する。エディタ部では、ペンインタフェースなどユーザが文書作成をする環境を提供する。また、文章検査結果の表示や、訂正候補のユーザへの提示を行う。図1に示すように、形態素解析部および訂正候補検索部は、それぞれWindowsのDLLとして作成した。これにより、形態素解析による検査機能と訂正候補検索機能は、本システム以外のWindowsアプリケーションからも利用できる(6節)。

3 ペンによる操作

従来のキーボードとマウスを用いた校正支援システムでは、テキストを修正する際に、修正箇所カーソル移動をした後で文字入力や文字削除するような場合が多く生じる。キーボードではカーソルの移動が面倒である。また、カーソルの移動先をマウスで直接指定しても、その後でキーボードで文字入力することになり、キーボードとマウスを持ち変える手間が発生する。

ペン入力では、これらの操作が連続的に可能である。つまり、修正箇所への直接文字入力、あるいは、ジェスチャでの文字削除ができる。このように、ペン入力によって校正支援システムにおける作業効率を上げられる。

ペンを利用する機能は以下の3点に大別できる。

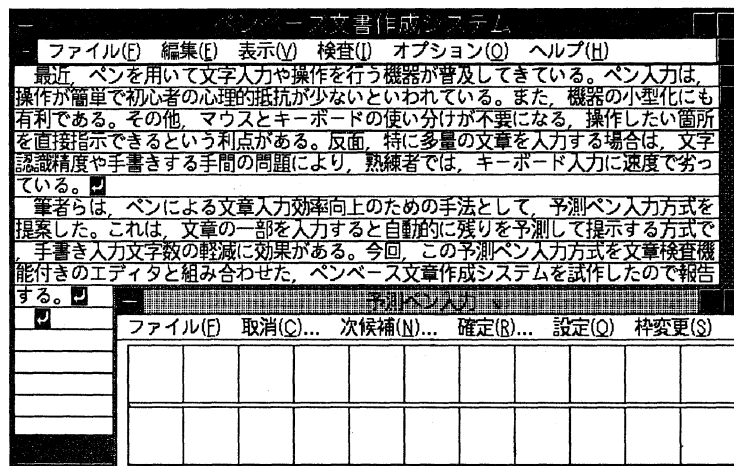


図 2: ペンベース校正支援システム (画面)

1. マウスによる操作の置き換え
ポインティングやドラッグなどの操作をペンで実行する。対象となる場所を直接操作できるのでマウスよりも使いやすい。
2. ジェスチャによる編集
改行、カット & ペーストなどの編集操作をジェスチャで行う。
3. 文字入力・文章入力
任意位置での枠なし入力と、枠付きの予測ペン入力部を用意した。枠なし入力は認識率が落ちるので、1文字の修正には簡易な枠なし入力、単語あるいは文章の入力には予測ペン入力機能を利用した枠付き入力と使い分けられる。

以下のジェスチャに対応した。ジェスチャの形状は、Windows for Pens 1.1 に用意されているものにしたかった。

- バックスペース、ペースト 現在のカーソル位置で実行する。
- スペース挿入、改行、タブ挿入 ジェスチャを書き込んだ位置で実行する。
- カット、コピー、削除 あらかじめ範囲指定しておき、指定した範囲に対して実行する。
- 単語削除 ペンでなぞった部分を削除する。
- 取消 編集操作を一つ取り消す。

今回の改造では、マウスやキーボードで実現していた機能をペンでも実現可能にして使いやすくすることを目標とし、新しい編集機能は加えていない。

今後、ペン入力の特長を活かした機能としては、訂正した箇所を手書きのまま残したり、遅延認識して色を変えたりして校正履歴を保存するといった機能も検討する [7]。

4 予測ペン入力機能

文字枠内に手書きで入力する。文字認識エンジンは、Windows for Pens 提供のものを利用している。文字認識後に、入力済みの文字列をもとにして続く文字列を予測し、画面に提示する。また、提示された予測文字列が誤っている場合は、誤っている文字以降を重ね書きする (図 3)。本方式には、事前に省略形を登録する必要がない、予測が誤っていても余分な修正操作の必要がないという利点がある [6]。

さらに、以下の機能を備えている。

- 単語の途中を省略した入力 (図 4)。予測ペン入力機能と文字の省略を明示する記号とを併用することでより柔軟に入力できる。
- 学習情報のファイル入出力。継続的な学習や、複数の学習ファイルの切り替えができる。
- 予測文字列候補の一覧の表示オプション。予測の第 2 候補以下を表示し選択入力する。
- 文字枠の大きさの選択。

5 文章検査機能

検査項目を図 1 に示す。未知語・誤用語・送りがな・かな書き推奨語・書き換え推奨語・同音類語・



図 3: 予測ベン入力の場合



図 4: 省略入力機能との併用

同音注意語・同音誤り語の8項目は形態素解析を利用する。すなわち、単語辞書内に文章検査用の校閲属性を与えておき、それをもとに入力の誤りや注意の必要な語を指摘する。なお、「未知語」は、単語辞書に登録されていない文字の並びを検出するもので、単語辞書に登録されている情報ではない。また、検出した語の訂正候補表示も行う。

同音誤り語検査は、かな漢字変換による入力時によく起こる、同音語選択の誤りを検出する機能である。誤り検出のために、2単語からなる用例を辞書に用意して、同音語をもつ単語が見つかったときに用例

表 1: 文章検査機能

校閲属性	説明
未知語	単語辞書にない語
誤用語	誤った表記
送り仮名	基準外の送り仮名
かな書き推奨語	ひらがなで表記した方が良い語
書き換え推奨語	書き換えた方が良い語
同音類語	紛らわしい同音語を持つ語
同音注意語	同音語を持つ語
同音誤り語	選択誤りの可能性の高い語
片仮名表記	カタカナ表記のゆれ
数字表記	漢数字と算用数字の混用
括弧	括弧の対応誤り
漢字レベル	教育漢字の学年レベル外使用

を参照して、誤りの可能性の高いものを検出する。

片仮名表記・数字表記・括弧・漢字レベルの4項目は字面解析による。

本システムをパソコンPC-9821An(CPU:Pentium 90MHz, メモリ:70Mbyte)に実装し、文章検査の速度を測定したところ414.7文字/秒であった。ただし、解析用50万語辞書をメモリにロードして実行した。これは、A4版文書1ページあたり2-3秒に相当し、一般ユーザ用としては十分な速度である。

6 文章解析ライブラリ

本システムで文章検査に使用している形態素解析部は、2節で述べたようにWindows用のDLLとして実現している。このDLLを呼び出すことで、他のシステムからも日本語解析や文章検査の機能を利用できる。ユーザごとにさまざまな文書作成アプリケーションを使用しているパーソナル環境では、文書作成アプリケーションとは独立に文章検査機能と呼び出せるような汎用性も重要である。

Windows上の種々のアプリケーションから利用できる文章解析環境を提供するために、クリップボードを介してデータをやり取りするインタフェースツールを試作した(図5)。本ツールによって、日本語文章の形態素解析機能(文章を単語分割し読みがなと品詞を付加する)と文章検査機能を、クリップボードとテキストの受け渡しができるアプリケーションから利用できる。

本節では、試作したインタフェースツールの機能について説明する。

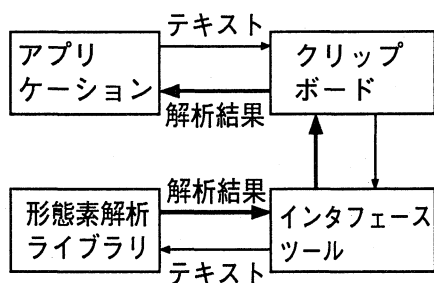


図 5: 解析ツールの動作

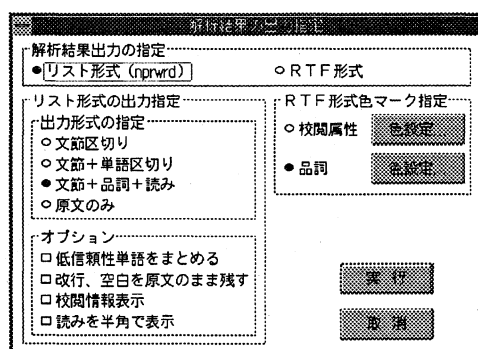


図 6: インタフェースツールのメニュー

本ツールは、形態素解析機能と文章検査機能を備えている。出力形式としては、単語情報を付加したテキスト形式と、単語に色を指定する RTF 形式とを用意した。RTF 形式の出力は、RTF に対応したアプリケーション (MS-WORD など) を利用して、特定の品詞や誤り箇所の色を付けて表示できる。機能と出力形式はメニューで選択する (図 6)。

形態素解析機能によって、入力テキストを単語に分割して、各単語に読みがな、品詞コード、校閲属性を付加する。各情報は形態素解析に利用する単語辞書内に登録されている。試作ツールで使用した単語辞書は、約 50 万語を収録している。なお、校閲属性は文章検査機能で利用している。

テキスト形式で出力した形態素解析結果は、ワープロやエディタでそのまま利用できる。単語表記 (入力テキスト) と、品詞コード、読みがなを列挙する形式など 4 通りの形式を用意して、そこから選択するようにしている (ユーザが出力形式を決められるように

改良中)。RTF 形式で出力すると品詞 (13 種類に大別) で単語を色分けできる。

文章検査機能では、表 1 の検査機能のうち形態素解析を利用する 8 項目を実行して RTF 形式で出力することができる。

本ツールの今後の予定としては、現在の機能の改良 (特に多様な出力形式のサポート) を行う。また、形態素解析の応用例 (キーワード抽出など) の有効性を検討した上で、本ツールに追加したい。

7 おわりに

ペンに対応したパーソナル文章校正支援システムを試作した。

キーボード、マウスに加え、ペンを文字入力・文書編集・コマンド操作に使用できる。また、文章の入力効率を上げるために予測ペン入力機能をもつ入力部を備えている。

文章検査機能は、形態素解析を用いた入力誤り検出や、同音語選択誤りの検出、用語の統一など 12 種類を備えている。

文章検査に用いる形態素解析部は Windows 用の DLL として実現し、50 万語辞書使用時でも 400 文字 / 秒と高速である。また、形態素解析部を独立して使用するためのインタフェースツールを用意することで、任意のアプリケーションから文章解析機能と文章検査機能を利用できる。

今後は、本システムの使用経験を積み重ねながら、評価 / 改良を行っていく。

参考文献

- [1] 奥村, 脇田, 金子: “日本語校正支援システムにおける校正知識”, 情処 48 回全国大会, 5Q-6(1994)
- [2] 福島, 佐々木, 宇都宮, 竹元: “校正支援システム St. WORDS”, NEC 技報, Vol.47, No.8(1994)
- [3] “特集『誤った日本語に気付き始めたワープロ』”, 日経バイト 1 月号, pp.148-154, (1995)
- [4] 山田, 福島, 竹元: “ペンベース文書作成 (1): 試作システム”, 情処 49 回全国大会, 1S-6(1994)
- [5] 福島, 山田: “ペンベース文書作成 (2): 予測ペン入力方式”, 情処 49 回全国大会, 1S-7(1994)
- [6] 福島, 山田: “予測ペン入力インタフェースとその手書き操作削減効果”, 情処 HI 研究会, 59-2(1995)
- [7] 曾谷, 福島, 高橋, 中川: “遅延認識方式による手書き原稿作成プロトタイプシステム”, 情処 HI 研究会, 43-4(1992)