

留学生の日本語作文データベース

中川 正弘

0. はじめに

コンピュータが普及した現在、ワープロを使うことで高品質の教材が誰にも簡単に作成できるようになってきているが、単純な文書作成以外にもコンピュータはさまざまに利用されている。

ワープロ専用機と違いコンピュータ上で動かすワープロの中には日本語、英語、欧文以外に中国語やタイ語、アラビア語など、文字の特殊な言語も混在する文書を作成できる多言語対応ワープロもある¹⁾。そこで、その言語が扱えるよう設定したコンピュータをこちらで用意すれば、コンピュータにそれほど慣れていない外国人留学生でも、協力を要請し、操作を練習してもらいながら少数の補助教材（語彙索引、対訳など）を作成することも可能だ²⁾。

また、コンピュータの得意とするデータ処理では、各種の試験の結果などを解析することで、日本語の研究、日本語教育の改善が図れよう。一方、コンピュータでの解析が容易な数値的データとはべつに、単純な言語テキストから膨大なデータベースが作られるようにもなってきた。

1. 日本語とデータベース

どんな文章でも、それを直線的に一度読むだけなら、わざわざコンピュータを使うことなどない。頁繰りなど取り扱いを考えれば、紙面で読むほうが簡単だ。しかし、その文章がテキストデータベースの形になっていけば、それを基に行う統計や分析まで考えなくとも、「読み方」そのものが変わる可能性もある。紙の形でしか保存できなかった言語資料がデータベース化されることで、これまでなかった利用法も出てくるだろう。

コンピュータのさまざまな機能の中でだれでも簡単に使える単純な機能は「検索」だ。瞬時に検索できることがどれほどありがたいか、これは書名、著者による図書検索や大事典の項目検索に限らない。分類、整理をせず、大容量ハードディスクに無造作に保存してある文書でも、文書名や作成日時、また文書中に使われている言葉から必要なものを探し出すこともできる。

日本語のデータベースとなるのは、まず出版された日本語のテキストであり、当然日本人の書いたものがほとんどである。日本語教育との関わりで言えば、読み物教材として使える文章である。また、日本語の文例や試験問題がデータベースとなっていれば、その中から必要に応じて随時選ぶことができるので、そのまま、あるいは、仕立て直して使った

りすることで、教科書や試験問題の作成にかかる時間がかなり節約できる。

他方、日本語教育に関わる者にはもう一つ目を向けなければならない日本語がある。日本語を学んでいる外国人の日本語である。そこには日本人の日本語には見られない間違いや書き方の傾向があり、ただその矯正についてだけではなく、どのような間違いがどのように、またなぜ起こるのか考えさせる。日本人の日本語だけで考えていたのでは見えない日本語の側面をわたしたちに見せてくれるのである³⁾。

間違いなど、言葉において何か異常を見せるのは何も外国人に限らない。日本人も同じだ。しかし、当然のことながら外国人の使う日本語ははるかに多くの「異常」を見せる。

会話ではその場の状況、また話している人の表情や身振りなどが情報を補足するため、言葉の異常が気にならず、話している人の人格の一部として、いわば表現として容認されやすい。一方、作文では、音声に関わる能力が見えない反面、省略や暗示などが多く、不完全なままで通用する会話と違い、内容の組み立ても文の組み立ても、はるかに慎重になされるため、日本語運用の濃密な記録となっており、書き手の文法についての理解など、書いた者の日本語がどの程度のものかよく分かる。

外国人の日本語は時には個性的な「スタイル」と感じられ、一種の表現価値を喚起する。そのようなものも含め、何か異常と感じられるとき、その事実はその外国人の理解、習得が不十分であるとか、母語の影響で文法がゆがんでいるとか、学習者について何かを語る。しかし、ただ学習者のことだけではなく、さらにその学習者が受けた日本語教育がどのように行われたか、使用した辞書や教科書の記述の不足、歪みなどの問題も間接的に語っているのではないだろうか。

このように考えれば、外国人の書いた日本語作文を読んで、「間違った日本語」を「正しい日本語」に書き直し、それで事は終わりとするだけではあまりにも惜しい。外国人の作文は、日本語を教える者、日本語について考える者にとって価値あるデータベースとなる可能性がある。

2. 小規模データベース

外国人学習者の書いた作文は朱書きで添削した後、書いた本人に返却されるのが普通だ。コピー機がこれほど普及している現在、このような添削済みの作文はコピーでごく簡単に残すことができ、後で分析することもできる。しかし、このような資料はある程度量が溜まれば、内容や語彙、文法の誤用などの記憶を頼りとするだけでは探せなくなる。

そのような添削資料の中から、添削した時点で、あるいは後に読み返した時点で重要と思える箇所をカードのようなものに残すことはできよう。だが、作文の一部を書き出しただけでは当然の事ながら全体の内容や他の言葉との関係を検討することはできない。添削した作文のコピーに通し番号でも付け、カード資料とリンクさせればその不備をいくらか補えるかもしれないが、何度読んでも読み飛ばしていた言葉を後になって探すのは困難だ。

しかし、このような資料からどんな言葉でもその使用例が簡単に検索ができれば、どうだろうか。電子テキスト化され、コンピュータ上で容易に検索できれば、便利この上ない。

データベースは普通、規模が大きければ大きいほど利用価値が高く、受けられる恩恵は大きい。しかし、ここで取り上げている添削済みの日本語作文は、大量に収集して大規模データベースを構築する以前の、私家版的な小規模データベースでも固有の利点がある。

自分の添削した作文であれば、それを書いた学習者について有形、無形の情報が記憶に残っていることが多い。そして、テーマや内容、使われた語彙、間違いの傾向などもなにがしか記憶に残る。どんな人が書いたか分からない作文で、どんな人が添削したか分からないものでも、資料として自由に読むことができれば、それはそれで有効だ。だが、日本語作文のデータベースが自分の添削したものだけで小規模にまとまっているだけでも、記憶という関連情報と結びついていることで、特別な資料体となるだろう。

本稿では、その延長上に大規模データベースの可能性を考えながら、このような小規模データベースの作成を提案したい。大規模データベースを構築する前段階として、私家版的なものをまず試みることは、どのようなファイル形式が望ましいか、また、どのような利用方法があるかを検討するよい機会となるはずだ。以下においてデータベース作成の具体的な手順を紹介するとともに、その問題点も考えてゆきたい。

3. 元資料

このようなデータベースの作成を思い立ったのは、筆者がこれまで行ってきた授業の形式から、元となるものが残っていたためである。その授業は作文を主軸にしたものであるが、学習者にとって自分の書いた文章が幾分客観的に見え、作文と添削が読み比べやすくすることをまず考えた。そこで、提出された作文を添削してただ返却するのではなく、作文はワープロ文書化し、それをワープロ上で添削してから、元作文と書き直しの二つのバージョンを学習者に渡すことにしたのである。入力にいくらか時間がかかるとはいえ、一端入力してしまえば書き換えや一度書き換えたものの変更も容易であるし、読み違い、書き込みの間違いも極めて少なくなる。作文を書いた者にとって、手書きの添削では読み返すことも、元の文と比較検討することも難しいが、ワープロ文書化されていれば、自分の書いた文章でもいくらか距離ができ、教師が書き換えたものと読み比べやすくなる。

学習者にはワープロ入力した元作文と書き直しの対照版を渡す代わりに手書きの作文は返却しないことにしていた。これで、20名規模のクラスで学生が週に原稿用紙一枚ずつ書けば、一年30週で、600枚溜まる。この原稿と、ワープロ文書化した元作文と書き直し、これが元資料だ。

4. 使用機器とソフトウェア

• Macintosh コンピュータ (PM7600/200)

データベースはMacintoshでもWindowsでも使えるファイルメーカーProで作成する。そして、もとなるのがDOS系の日本語ワープロの文書ファイルなのだから、作業を最短距離ですまそうと思えば、Macintoshを使う必要はない。しかし、Windowsをほとんど使ったことがない筆者にはこのようなデータベースの作成と利用は困難と思えた。できあがったものもMacintoshで使うつもりであるため、遠回りになるとしても、作業は可能な限りMacintoshで行うことにした。

• Windows コンピュータ (富士通 FM bibloDC/16NAIII)

ワープロ専用機で作成した文書フロッピーはMacintoshでは読めない。今回富士通ワープロのオアシス文書で保存していた下線付きのテキストを下線付きのままMacintoshで読めるようにしたかったのだが、それには、Windows用で出ているテキスト・コンバータを使うしかないため、作業の一段階にWindowsコンピュータが必要だ。オアシス文書をWindows版のMicrosoft Word文書に変換すれば、そのファイルはMacintosh版のMicrosoft Wordで読み込むことができる。

• 小型スキャナー (VISIONEER PaperportVx for Mac)

手書き原稿の画質をよくしようと思えば、フラットヘッド型のスキャナーを使うところだが、このデータベースでは過度に高画質な画像を求めても意味がない。モニタ画面ではっきり読め、プリントアウトしたものが必要な場合も、手書き文字が判読できるぐらいで十分である。画質をよくすれば、それだけファイル容量が大きくなる。画像データの容量をできるだけ小さくし、データベース全体の容量を小さくすることで、動作速度を上げるほうがいい。キーボードの半分程度の面積で設置できるこのタイプの小型スキャナーは入力も高速で、取り込んだ原稿用紙の画像を簡単に束ねたり分類できるので非常に便利だ。

• クラリス ファイルメーカーPro3.0 for Mac

テキスト・データと画像データを組み合わせてデータベースを作成するためのソフトで、Macintosh版で作ったものはWindows版からも読める。

• Microsoft Word 6.0 for Mac

オアシスのワープロ文書からテキストを直にファイルメーカーにコピー・アンド・ペーストで張り付けようとしても、テキストデータ以外の情報、つまり下線や文字種など、特殊な属性はすべて飛んでしまう。このようなものを残したい場合、ファイルメーカーが読み込み可能な文書形式、Microsoft Wordの文書に一旦変換するしかない。

• アンテナハウス リッチ・テキスト・コンバータ 97R2

日本語ワープロのほとんど全機種 of 文書を他の文書形式に変換できる。文字属性を含まない単純テキストとしてならMacintosh用にも変換できるが、ここでは富士通オアシスで作成した文書の文字属性などをそのまま残したいので、MacintoshとWindowsの両方で使

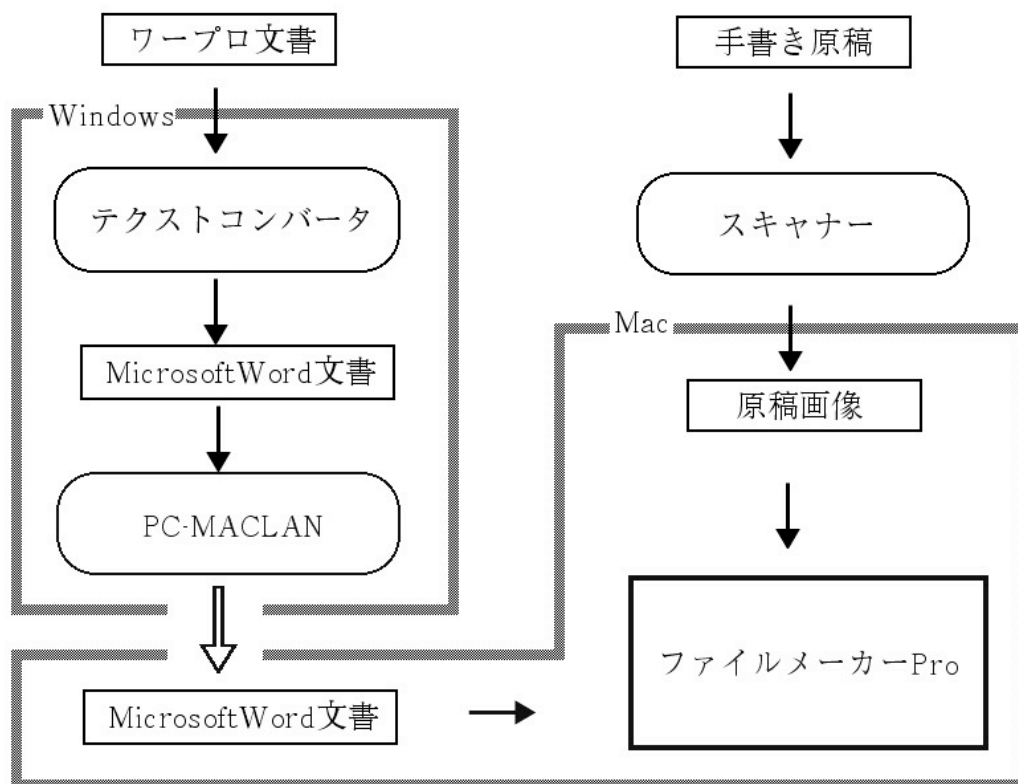
える MicrosoftWord 文書に変換する。

・ DIT PC-MACLAN for Windows95 v.6.1

ネットワークで接続された Windows コンピュータのハードディスクやフォルダーを Macintosh にマウントし、文書のやり取りをするために用いる。大量の文書をフロッピーディスクを使って Windows から Macintosh に移すのは繁雑だ。ディスク容量が大きい ZIP や MO を使うこともできるが、Windows と Macintosh がネットワークで接続できるなら、このほうが楽だ。

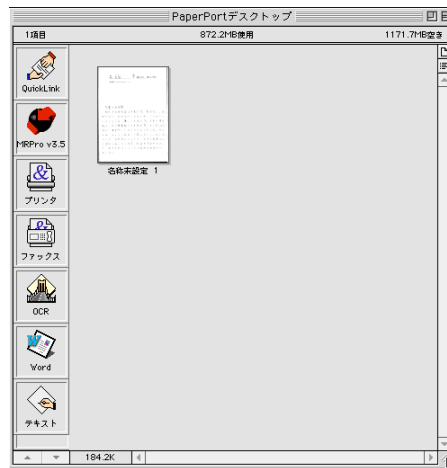
5. データベース作成作業手順

作業手順は以下のようになる。



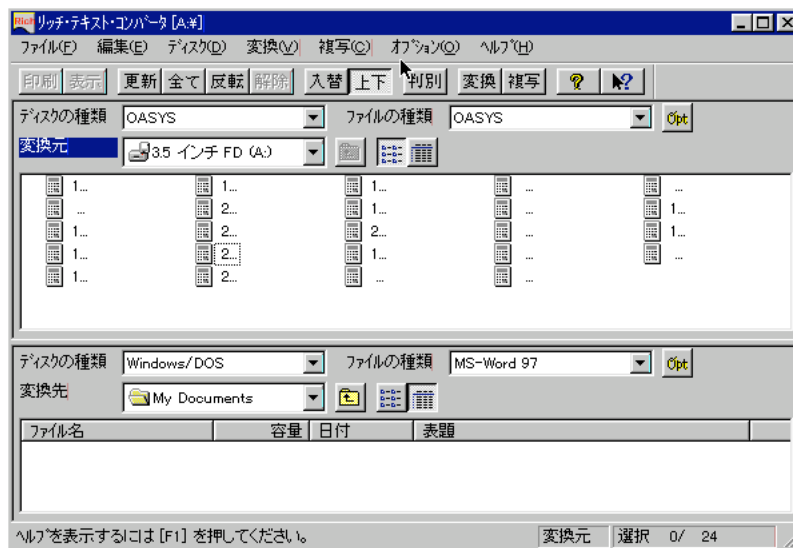
① 手書き原稿を画像として取り込む

Macintosh に接続した小型スキャナーで手書き原稿を画像として取り込む。1枚の読み込みは十秒ほどで、このスキャナー専用のウィンドウに自動的に画像ファイルのアイコンが並んでいく。このアイコンをダブルクリックすると、画像ファイルが開き、基本的な画像処理が行える。



②ワープロ文書を MicrosoftWord 文書に変換する

Windows でリッチ・テキスト・コンバータを起動し、ワープロ文書フロッピー（富士通オアシス）の作文テキストを Windows 版 MicrosoftWord 文書に変換する。変換したものは Windows コンピュータ内に作成したフォルダーに保存する。



③Windows から Mac へデータを移す

Windows コンピュータに PC-MACLAN をインストールすると、Macintosh の AppleTalk に接続が可能になる。そこで、Windows コンピュータ内の作文テキストフォルダーをネットワークで接続された Macintosh コンピュータにマウントし、これをハードディスクにコピーする。

④ファイルメーカーPro でファイルのフィールドを決める

必要なフィールドはテキスト用が二つ（元作文と書き直し）、画像用が一つ（手書き原

稿)。日付や書いた者の名前用のフィールドも作ろうと思えば作れるが、後から追加することもできるので、とりあえず三つだけとする。

⑤ テキストデータをファイルメーカーに流し込む

ハードディスクにコピーした MicrosoftWord 文書の作文テキストを Macintosh 版 MicrosoftWord6.0 で開き、ドラッグ・アンド・ドロップ (コピー・アンド・ペーストでもよい) でファイルメーカーの新規ファイルのテキスト・フィールドに張り付ける。

⑥ 画像データをファイルメーカーに張り付ける

PaperPort デスクトップ内の原稿の画像を一枚ずつ開き、マウスの操作で必要な箇所を選び、それをファイルメーカーの画像フィールドにペーストする。この作業はテキスト・データをすべて流し込んでから行うほうがいい。画像がどのテキスト・データのものであるかはきちんと整理できていない限り分からない。そこで、ファイルメーカーの検索機能を使い、手書き原稿の冒頭部なりをたよりに探せばいい。そして、見つけた当該ファイルの画像フィールドに張り付ける。

⑦ レイアウトをデザインする

今回は表紙、手書き原稿を主に見るレイアウト、それに元作文と書き直しを読み比べやすくするため画像を置かないレイアウトの三つだけにした。

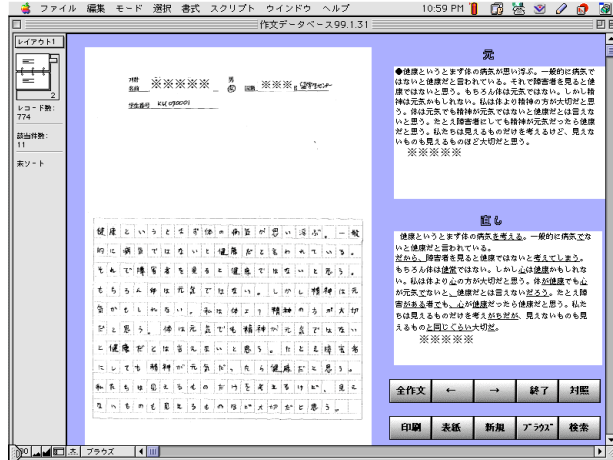
手書き原稿を見るレイアウトにも元作文と書き直しの小さめのテキストフィールドを置き、これは作文が長い場合のためにスクロール・ウィンドウにしてある。

ボタンはレイアウトの移動、ページめくり、印刷、検索ぐらいがどうしても必要なぐらいで、使っていて必要があれば後から追加することも可能だ。

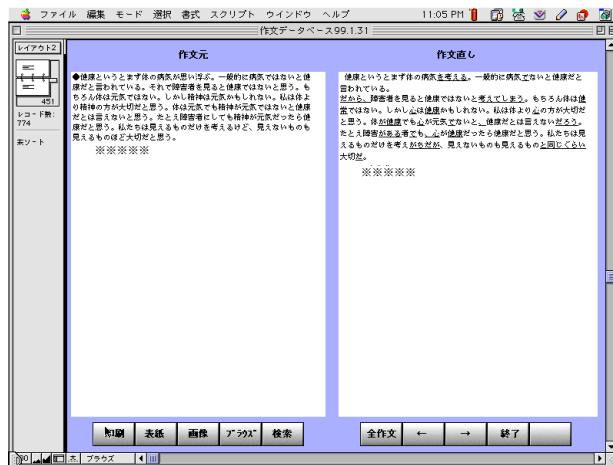
表紙は、手書き原稿画像へ行くボタン、元作文／書き直しの対照ウィンドウへ行くボタン、終了ボタンの三つ。手書き原稿を主に見るウィンドウと元／直しの対照ウィンドウには、レイアウトの移動、ページめくり、印刷、検索以外に、検索レコードから全レコードに戻るためのボタン、新規レコードを作成するためのボタンなどを置いた。



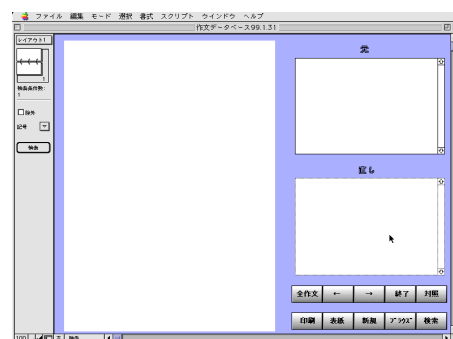
レイアウト 1



レイアウト 2



検索ウィンドウ



検索はどちらのレイアウトでも行える。元作文フィールドか作文直しフィールドに検索したい言葉を書き込み、リターンを押すだけだ。また、テキストには留学生の名前も含まれているので、使われた言葉だけでなく名前でも検索できる。

レイアウトのサイズはデスクトップの17インチモニタの標準表示(1024×768)用とパ

ワーブックの標準表示（800×600）で切り替えられるようにした。Lサイズ表示にすれば原稿用紙の画像がほぼ実寸大になり非常に見やすくなる。Sサイズは50%程度の縮小表示だが、これでも手書き文字が画面上で十分読めるほどの画像精度であり、印刷してもかなり鮮明なプリントが得られる。

6. おわりに

先にも述べたように、誤用など、外国人の用いる日本語について研究するための資料として作文が保存されていても、紙のままでは資料として生かされにくい。これがコンピュータ上で検索の容易なデータベースとなっていれば、どれほど役立つであろうか。本稿で紹介した小規模データベースはひとりの日本語教師が関わる日本語作文の量的な限界を越えないものであるが、だからこそ作文を書いた学生についての記憶、作文の内容や書き直し時に考えたことの記憶が残っている特別なデータベースとなる。

この規模でもさまざまな誤用の例に当たることができるため、日本語研究者や日本語教育に関わる方々に利用していただこうと考えている。また、同様のデータベースを自作しようと思われる方々は、このデータベースのファイル形式、レイアウトだけをテンプレートとして使い、必要に合わせて作り替えていただいても結構だ。この作業がご自身でできない場合には、ワープロ文書化も含め、アルバイトを使って作成できなくもない。

こうして作られた小規模データベースは将来、作成者同士が合意すれば一つに合わせてより大きな規模のデータベースとしていくこともできるだろう。（了）

注

- 1) 以前から多言語対応で定評のある NisusWriter（マーキュリーソフトウェア）に加え、I-Write（Han Mac ソフトウェア）、MicrosoftWord（98 から多言語対応）などがある。
- 2) Macintosh コンピュータには中国語、韓国語、アラビア語、ロシア語などの Language Kit が用意されている。また、インターネットから無料でダウンロードすることができるモジュールもいろいろある。タイ語は <http://thaigate.rd.nacsis.ac.jp/ftp/upload/kritbeta/MacOSThaiEnable8b6.hqx> で、ビルマ語は <http://w3mb.kcom.ne.jp/~ishi/> で手に入る。
- 3) 中川正弘、**作文の誤りと文体**、広島大学留学生センター紀要、第3号、1993年。
中川正弘、**作文と解釈行為**、広島大学留学生センター紀要、第4号、1994年。
中川正弘、**作文の添削と文体差**、『広島大学留学生日本語教育』第7号、1995年。
中川正弘、**「は／が」と助詞選択の零度**、『広島大学留学生日本語教育』第8号、1996年。
中川正弘、**添削文が語る日本語のスタイルー (1) 完了表現と時制**、『広島大学留学生教育』第1号、1997年。

中川正弘、添削文が語る日本語のスタイルー (2) 言葉の選択と序列、『広島大学留学生教育』第2号、1998年。