視覚障害者向けかな漢字変換ソフトウェアに関する考察

平成 17 年 8 月 1 日

情報電子工学科 竹野研究室
山口 圭

目 次

1 はじめに **1**

2 日本語文書中の漢字説明ソフト **2**

2.1 フリーソフトウェア 2

2.2 視覚障害者の現状 2

2.3 視覚障害者の文字入力方法 4

2.4 有用なフリーソフトについて 5

2.5 UNIX 指令について 7

3 漢字の説明方法 **8**

3.1 日常で使われる漢字の説明方法 8

3.2 説明方法の候補 9

4 単漢字における統計 **13**

4.1 対象の漢字 13

4.2 読み方と漢字の種類の差異 14

4.3 統計の対象 16

4.4 可否の基準 16

4.5 統計結果 17

4.6 健常者に対してのアンケート 20

4.7 統計とアンケート結果についての考察 22

5 まとめ **24**

参考文献 **26**

概要

現在、視覚障害者の方が作成した WWW ページや電子メール等には同音異義語の誤字がよく見られる。そして、様々な OS が存在する中で、UNIX 上で動作する視覚障害者向けのかな漢字変換ソフトウェアはあまり普及していないと言える。そこで本研究では、この状況が多少でも改善されるように、 UNIX 上で動作する視覚障害者向けのかな漢字変換ソフトウェアに関する考察を行なう。本研究室では、視覚障害者向けのかな漢字変換ソフトウェアに関して、過去に熟語を用いて単漢字を説明する方法の研究が行なわれたが、本研究では熟語を用いず、単漢字のみの状態で画数や部首等を用いて説明する方法を取った。そして、その説明方法をいくつか提示し、プログラムを用いた統計を取り、その結果と実際に健常者に対して行なった漢字の実験的なアンケートの結果を考察する。

1　はじめに

視覚障害者の方が、パソコンやインターネットを使用するために必要だと思われるスクリーンリーダーソフトやかな漢字変換ソフト、点字を用いて説明するソフトは、 MS-Windows や MS-DOS を中心に普及されている。数年前まではあまり普及していなかった UNIX 常でも、多少は普及されつつあるようだが、前述の 2 種の OS に比べるとその数は明らかに少ない。したがって、現状では視覚障害者の方が UNIX を使用し、かつ日本語を使用する場合、それは非常に困難だと言える。これは、視覚障害者の方が作成した WWW ページや電子メールを健常者が見るとわかることだが、所々に同音異義語の誤字が見られる。全網の視覚障害者の方々にとって、キーボードから入力した文字は視認できないので、いくつかの同音異義語の中から目的のものを音声のみの情報により選ぶことが難しいためだと思われる。

さらに、視覚障害が先天的か後天的かによっても誤字となってしまう確率は変化する。後天的な視覚障害者の方の場合、その時期によるが多少は漢字を視覚的な知識としても備えていると思われる。したがって、スクリーンリーダーソフト等の説明ソフトを使用したとすると、その多少の漢字の知識を元に使用するので、それなりの結果が出せるのではないか、と思われる。しかし、先天的な視覚障害者の方だった場合、そもそも漢字そのものを健常者と同様に理かいすることが難しく、「同音異義語の判別」という行為は健常者の理解しえない苦労があると思われる。この場合、説明ソフトを使用した場合でも、同音異義語の誤字は発生してしまうと思われる。

しかし、本研究は健常者に値かい日本語と漢字に対する知識を備えた視覚障害者の方向けの漢字変換ソフトウェアの開発を目指すことにする。したがって、視覚媒体に頼らない漢字の説明をする場合、どのような方法ならば効率よく説明できるかを調べ、そのための実験や統計とその結果や考察を述べる。

ただし、漢字変換ソフトウェアを開発する場合、全ての漢字に対して、それぞれ固有のわかりやすいであろう説明を載せた辞書を用いることが最も効率よく説明することができる。しかし、本研究ではこのような辞書を用いないような説明方法を研究したいくことにする。

今回の研究は、以前同様の研究を同研究室で行なった井上氏 (2000 年) のものを参考にしつつ進めさせてもらった。井上氏の研究は、漢字のみで構成された熟語中の、どの位置に説明する対象の漢字があれば効率よく説明できるかを調べた。今回の研究も漢字を説明する場合、どのような方法が効率よく説明できるかを調べることに重点を置いた。しかし、井上氏の研究が熟語を用いて単漢字を説明するのに対して、今回の研究は熟語を用いずに説明するにはどのような方法が効率がよいかを調べ、それを考察した。

2　日本語文書中の漢字説明ソフト

2.1　フリーソフトウェア

フリーソフトには、

- 無料でインターネットで入手できる
- ソースが公開されており、誰でもそれを改良できる
- 自由に配布することができ、色々な人達に見てもらえる

等のメリットがある。このようなソフトウェアを PDS(Public Domain Software) と言う。しかし、著作権によって保護されており、いくつかの制限のある項目を持つソフトもある。

また、フリーのソフトにはデメリットもある。例えば、ソフトのしように関しては使用者が一切の責任を持たなければならない。仮に、ソフトが暴走してしまい何らかのデータが破壊されたとしても、自己責任なので文句は言えない。そして、メーカーの助言は得られないため、全ての処理は使用者が行なうことになる。

今回の研究で作成したソフトは、ユーザーがソースを自由に改良できるメリットが必要であると考え、フリーソフトウェアとして配布することを考えている。

2.2　視覚障害者の現状

現在、視覚障害者の方々がどのようにしてパソコンやインターネット等を使用しているのかを調べてみた。現在、ほとんどの視覚障害者の方は、MS-Windows をメインに使用している。ただし、一昔前まで主流だった MS-DOS を使用している方も少なくないようである。 UNIX を使用している方も多少はいるようだが、その数は前述の二種に比べると明らかに少ない。そこで、 MS=Windows 、 MS-DOS 、 UNIX の三種に分けて視覚障害者の現状をまとめてみた。

MS-Windows 環境

画面の内容を視認できない場合、スクリーンリーダーソフト (画面音声化ソフト) 等をインストールして使用することで、その音声を聞いて画面の様子を知ることができる。現在は視覚障害者の方々にとっても主流の OS のようで、様々なバージョンが存在する。中でも MS-Windows98 以降のものが好まれて使われている。基本的に、 MS-Windows95

以降のバージョンであれば、使いやすいシステムを構築することが加納だが、 MS-Word や MS-Excel 等のアプリケーションを使用することを考慮するならば、 MS-Windows98 以降のバージョンを使用することが望ましい。

スクリーンリーダーソフトが対応しているアプリケーションも増えてきているようで、それに伴い全盲の視覚障害者の方が音声を頼りに MS-Windows 上で使えるソフトも増えている。ただし、視覚障害者の方は MS-Windows のような GUI 環境にはなかなかうまくアクセスできないという現状もある。 GUI 環境とは、画面に表示されているアイコンを見ながら、マウス等でパソコンを操作するといった形式のものである。よって、 MS-Windows を使用する場合でも、マウスなどを使用せず、キーボードのみで操作する必要がある。

MS-DOS 環境

前述の通り、ほとんどの視覚障害者の方は GUI 環境にはなかなかうまくアクセスできないのが現状である。このことから、キャラクタベースの MS-DOS が視覚障害者 (特に初心者視覚障害) のパソコンユーザーには向いているとされる。一昔前まで多くの視覚障害者の方が MS-DOS を使用し、現在でも比較的多くのユーザーが存在する。しかし、現在 MS-DOS 用のソフトは入手が困難で、製造元に問い合わせても在庫がすでにない、という状況も多々あるようだ。視覚障害者の方が使えるソフトには、漢字を入力するための日本語入力ソフト (FEP) 「 ATOK9 」等が、市販ソフトとして存在する。

UNIX 環境

NEC98 シリーズのパソコンで動く UNIX は FreeBSD や Plamo Linux 等が使用されているそうである。ただし、それらの UNIX 上でのスクリーンリーダーソフトはまだあまり普及していないようなので、全盲の視覚障害者の方が UNIX を使用するとなると、パソコンが 2 台と外付けの音声装置が必要になってしまう。点字ディスプレイもあった方が便利ということである。

パソコンが 2 台必要になるのは、 UNIX をインストールしたパソコンに直接音声装置や点字ディスプレイを繋いでも、音声や点字を出力できないからであり、そのため MS-DOS をインストールしたもうひとつのパソコンから、UNIX がインストールされたパソコンにアクセスし、音声や点字を出力している状況である。実際には telnet プログラムを用いて LAN からアクセスしたり、ETV という通信ソフトでシリアルポートからアクセスしたりしている。

しかし、以上の障害を乗り越えることができれば、他の OS 以上に視覚障害者の方も健常者と同様に使用することができるというメリットもある。さらに現在の MS-DOS の状況とは違い、ユーザーが多いので新しいソフトは現在も開発されている。したがって、現在は存在しないソフトであっても、今後開発される可能性を秘めていると言える。

2.3 視覚障害者の文字入力方法

視覚障害者の方がパソコンを使用する際、どのように文字を入力しているかを調べてみた。健常者は以下に説明するフルキーでのローマ字入力か、フルキーでの仮名字入力のどちらかの方式を取るのが一般的である。そこで、6 点入力を含めた三種の文字入力方法を説明する。

6 点入力

6 点入力とは、キーボードの特定のキー六つを点字の 6 点に見立てて、点字入力を行なう入力方式である。 6 点入力を行なうためには専用のソフトウェアが必要で、さらにキーボードによって 6 点入力が可能なものとそうでないものがある。

長所
- すでに点字を知っていれば、パソコンが身近なものとして使える。
- 漢点字を使うことで変換辞書に頼らず確実な漢字入力ができる。漢点字とは、6 点または 8 点の組み合せを数文字分使って漢字を直接入力するもので、これを使いこなせる一は現在それほど多くない。

短所
- 6 点だけでパソコンの全てを操作するには限界があり、フルキー・アルファベットや特殊キーを覚えることは必須となる。
- 6 点入力を行なう場合、指定されたキーを同時に押す必要がある。これにより、キーボードによっては認識できないものがあるため、特にノートパソコン等では使用できるものが制限されることがある。

フルキーでのローマ字入力

フルキーでのローマ字入力は、健常者も使用している最も一般的な文字入力方法である。日本語を入力する場合、子音と母音をキーボードからローマ字によって入力し、それが対応した平仮名となり、さらにそれを必要に応じて漢字、片仮名、記号、数字等に変換する。

長所
- 前述の 6 点入力のような、パソコンを使用する際にキーボードの心配がない。
- どのパソコンでもほぼ同じ入力が保証される。
- 日本語・アルファベット・記号を関連して覚えることができる。

短所
- ローマ字を知らない場合、新たにローマ字を覚え、さらにキー配置を覚えることが大変である。

フルキーでの仮名字入力

フルキーでの仮名字入力は、キーボード上の平仮名と濁点、半濁点、 Shift キーを用いて平仮名を入力する。必要に応じてそれを漢字、片仮名、記号、数字等に変換する。ローマ字入力よりもタイピングの数が少なくて済むものの、覚えなければならないキー配置の数がローマ字入力よりも多くなる。

長所
- ローマ字入力と同様に、パソコンを使用する際のキーボードの心配がない。
- 熟達した場合の入力速度は、他のものと比べると非常に速い。

短所
- キー配置を覚えるのが今回説明した三種の中で最も困難である。
- アルファベットキーをこれとは別に覚える必要がある。

2.4 有用なフリーソフトについて

視覚障害者向けのかな漢字変換ソフトウェアは、 MS-Windows や MS-DOS 用のものとして開発、販売されている。しかし、これらのソフトは UNIX では JavOICe 等、ソフトは存在しているようだが、情報が少なく、数も少ない。当然有料であり、フリーソフトではない。さらに、このようなソフトは、漢字を分かりやすく説明するために、漢字一字一字についてそれぞれ固有の説明を有する辞書が作成されている。

例えば、

家：人が住む建物、「家庭」の「か」声：人が話すときに口から発する音の振動、「音声」の「せい」

といった対応する説明が登録されている。これが最も良い方法のひとつではあるが、これにはそれなりの人手と時間が必要になる。しかし、フリーソフトとして作る場合、これでは不適切なこともあるので、この方法ではなく、何らかのプログラムを用いて人手と時間を少なく抑えつつ、漢字説明のソフトウェアの開発ができないかを考えていくことにする。そのために、まず目標にするフリーソフトの開発に当たって使用する UNIX 上の辞書ファイルを紹介する。

本研究に使用する辞書は、フリーの辞書である必要がある。そこで、kakasi の辞書である kakasidict と、それとは別に kakusu を使用する。辞書によって登録されている内容は異なる。当然、その内容によって得手不得手も発生する。したがって、ある辞書から有用な部分を使用し、有用でない部分は他の辞書に任せる、という方式を取る。

以下はそれぞれの使用目的と特徴の説明である。

kakusu

1 画から 30 画までの 6353 漢字が収められている辞書である。これは日本工業規格 (JIS) で定められた漢字の規格、正式には JIS X0208 の 1983 年度版に含まれた「情報交換用漢字符号系」(JIS 漢字水準) の第 1 水準と第 2 水準の漢字全てが登録されていることになる。本来この X0208 には、漢字以外に英数字、片仮名、平仮名、記号、罫線素片等があり、ギリシア文字やロシア文字も含まれている。いくつかのバージョンが存在し、それは数年に一度の割合で更新されている。更新される度に登録されている漢字の種類は増えている。ここでは、漢字をコンピュータなどで利用するために 2 バイトで 1 字を表す 2 バイト・コードを使っており、使用頻度を考えて JIS 第 1 水準 (2965 字) と JIS 第 2 水準 (3388 字) に分けられている。前者は基本的な漢字から構成され、後者は人名、地名等の特殊な固有名詞や旧漢字から構成される。これら全ての漢字がこの辞書に登録されている。

表示は以下の形式で登録されている。

11 晢 (にち) 11 移 (のぎ) 11 釈 (のごめ) 11 習 (はね) 11 翌 (はね)

登録されている漢字は全て単一の漢字のみであり、重複なく、熟語や送り仮名を含むものは登録されていない。漢字とその画数、部首は登録されているが、その漢字の読み方は登録されていない。登録されている部首名も、「へん」や「かんむり」といったものは省略され、例えば「くさかんむり」だった場合、「くさ」とだけ登録されている。ただし、一般的な漢和辞典などに登録されている部首名とは異なる名称で登録されているものも中には存在している。

表示形式は左から順に、対象の漢字の画数、漢字、部首名を示している。登録されている全ての漢字は、画数でソートされ、次に部首名でソートされており、検索がたやすくなっている。今回はここから画数と部首を検索する。

kakasidict

kakasi という、漢字かな読み上げソフト (漢字を平仮名や片仮名などに直すソフト) の辞書ファイルである。この辞書には 121795 語の熟語や単漢字、単語が適度にソートされた状態で登録されている。

表示は以下の形式で登録されている全ての単語を表示している。

えい 鋭するど i 鋭するど k 鋭するどさ 鋭さえいい 鋭意

左から順に読み方、漢字や平仮名、片仮名を含む熟語または単漢字を示している。ただし、三番目の候補の場合等の「するど k 」の「 k 」は送り仮名の始めのイニシャルを示す。これは送り仮名が多量に存在するので、この「 k 」は「く」や「かった」等の現在系、過去系を同時に表現させていることになる。

今回はこの中から、一番目の候補のような、送り仮名を必要としない単漢字だけを抜き出し、その 9729 字を使用する。今回はここから単漢字での読み方と熟語を検索する。

2.5 UNIX 指令について

UNIX には、 cat のようなテキストファイルの編集を行なうための多くの標準的なコマンドが存在する。その中で、今回使用するコマンドを紹介する。

grep ファイルに貯えられている文書データの中から、特定のパターンに合う文字列を含む行だけを抜き出すコマンドである。 1 番目のパラメタにパターンを、2 番目以降のパラメタにファイル名を与えて grep 指令を発すると、指定された文字列を含む行だけが出力される。

awk grep 指令による行の抽出はいわば「横方向の抽出」であるが、ファイルのデータの画行が幾つかの項目からなっているとき、特定の項目を抜き出す「縦方向の抽出」には awk 指令を使う (ただし、各行の項目は 1 個以上の空白又はタブで区切られているものとする) 。

sed 文書データに含まれる大文字の小文字への変換や、相続く空白を 1 個の空白に置換する。このような、あらかじめ定められた手順にしたがってデータの変形を行なうには、 sed という流れ処理方式のエディタが有用である。

sort ファイルに貯えられている文書データを、ある項目について順番に並べる、すなわち整列させるには sort 指令が有用である。

uniq ファイルに同じ内容の行が何行か含まれていると、整列した後ではそれらの行は連続した位置に整列される。連続した同じ内容の行のうち、ひとつの行だけを残す指令が uniq である。

3　漢字の説明方法

3.1　日常で使われる漢字の説明方法

視覚障害者の方に漢字を説明する場合、日常で健常者が言葉だけで漢字を説明するような方法が有効である、と考えられる。日常の会話で使われる「単語」とは、漢字や平仮名等が連なってひとつの漢字文字列になったものであり、ここから単語を用いた説明方法、画数や部首を用いた説明方法、訓読みでの説明方法等、様々なものが考えられる。このことを踏まえて、実際に健常者が日常生活の中で声だけで説明するときにはどのように説明するのかを考える。

熟語の場合

本研究室の 2000 年の卒研生である井上氏の研究では、単漢字を熟語を用いて説明する方法を用いていた。そこで、まずは井上氏が行なった説明の方法を含めて、熟語としての説明方法はどのようなものがあるか以下に挙げる。

例：安全

- 「安心のあんに、全部のぜん」 . . . 漢字を一字ずつ説明

  他の熟語等、そのときに使用している方法以外の、他の使い方を用いて説明する方法。

- 「全てを安ずるという単語」 . . . 単語ひとつとして説明

  意味を説明し、そこから連想させることで説明する方法。

- 「安らぐという漢字に、全てという漢字」 . . . 訓読みを用いた説明

  熟語ではなく、その熟語を構成している漢字を一字ずつにわけ、それを訓読みで説明する方法。

ここの一番目の説明方法が井上氏が用いた説明方法である。井上氏はこの方法で、熟語のどの位置に説明する対象の漢字があれば効率良く説明できるかを調べた。つまり、例の「安全」の「安」の場合、上記のような先頭に対象の漢字がある場合の他に、「大安のあん」のように末尾にある場合、「天安門のあん」のように中間にある場合の、どれが効率がよいかを調べた。

単漢字のみの場合

次に、単漢字のみの場合はどのようなものがあるか以下に挙げる。熟語の場合と分けたが、これは単漢字のみの場合とでは、異なった説明方法が使えると思えたからである。また、単漢字として説明できれば、それをいくつかつなげることで熟語を説明することもできると思ったからである。

例：移動の「移」を説明したい場合

- 「のぎへんに多いという漢字」 . . . 部首に分けて説明

  部首別に分け、その名称、または分かりやすい他の漢字を用いて説明する方法。

- 「移転のいという漢字」 . . . 単語を用いた説明

  他の熟語等、そのときに使用している方法以外の、他の使い方を用いて説明する方法。

- 「動く方のうつるという漢字」 . . . 訓読みでの説明

  同音異義語としていくつか同じ読み方のものがある場合、目的の漢字の意味と共に、訓読みを説明する方法。

3.2 説明方法の候補

視覚障害者向けのかな漢字変換ソフトウェアを開発するに当たり、まずはその説明方法はどのようなものが候補として考えられ、どの候補が説明に有効であるか、という問題がある。また、それとは別にプログラムを作成する上ではどの候補適当か、という問題もある。現時点でもいくつかの候補が考えられるので、説明方法の候補を以下に挙げる。これら全てにおいて、それぞれの方法に添ったものを辞書ファイルから検索し、それを yomi を用いて発音させて説明する試験的なプログラムを作成することを念頭に置いているものとする。 yomi とは、本研究室で開発された、ごく単純な方法でテキストファイルを音声化するためのソフトである。さらに実験前に考えられる長所、短所も付随する。

そして、これらの説明は全て以下の条件を満たしている状態で使用されているものとする。

- 使用者は、全盲視覚障害者である。
- 使用者は、一般的な成人の健常者とほぼ同等の日本語と漢字に対する知識を備えている。

- キーボードから入力された平仮名を変換するものとする。つまり、求めたい漢字の読み方はわかっているものとし、その漢字の同音異義語の中から求めたい漢字を求める、という方式を取ることとする。
- 求めたい漢字は一字の単漢字で、送り仮名はなく、熟語、単語でもないもののみとする。

以上の条件下での説明方法の候補を以下に挙げる。

画数を用いた方法

入力された平仮名から、変換候補を探し出し、その候補の漢字の画数を辞書ファイルから検索し、それぞれを説明する方法。これだけでは数字を読み上げるだけになってしまうので、プログラムを作成する場合、先頭に「画数は (かくすうは) 」、末尾に「画 (かく) 」を付け加えたいと思う。画数の検索を行なう場合、 kakusu から検索する。

例：あつし

厚：画数は 9 画 (かくすうは きゅう かく)
淳：画数は 11 画 (かくすうは じゅういち かく)
敦：画数は 12 画 (かくすうは じゅうに かく)
篤：画数は 16 画 (かくすうは じゅうろく かく)

長所
- 画数を読み上げるだけなので、説明が短く済む。
- プログラムを作成する上で、辞書ファイルから検索し、先頭に「画数は」と末尾に「画」を付け加え、それを発生させるだけなので、今回候補に挙げるものの中では、比較的簡単に済む。

短所
- 変換候補の中に同じ画数の漢字があった場合、この説明方法だけでは判別が不可能になる。
- 求めたい漢字の画数は、入力する際に数えなければならないので、後述のものよりも時間がかかる。
- 画数が使用者の知識と登録されているものとで差異が発生することがあり得る。さらに、実際に対象の漢字を書く場合の画数と、登録されている画数とが異なる場合がありえるので、漢字に対する正確な知識を備えていない場合、判別は難しいと言える。

短所の三番目のものは、例えば「遠」の漢字の場合、しんにょう等見た目や実際書くときの画数と、登録されているものとで差異が発生する恐れのある部首を持った漢字全てに

言える短所である。例で挙げたしんにょうの部分だけに焦点を合わせると、これは 3 画である。しかし、 4 画として覚えているひともいるかもしれない。同様のことが 2 画としても言える。

部首を用いた方法

入力された平仮名から、変換候補を探し出し、その候補の漢字の部首を辞書ファイルから検索し、それぞれを説明する方法。これだけでは部首を読み上げるだけになってしまうので、プログラムを作成する場合、先頭に「部首は (ぶしゅは) 」を付け加えたいと思う。部首の検索を行なう場合、kakusu から検索する。

例：あつし

厚：部首は がんだれ (ぶしゅは がんだれ)
淳：部首は みず (ぶしゅは みず)
敦：部首は とまた (ぶしゅは とまた)
篤：部首は たけ (ぶしゅは たけ)

ただし、上記の例を見ればわかるが、この辞書ファイルに登録されている部首の説明は、「さんずい」であろうと「みず」、「たけかんむり」であろうと「たけ」のように登録されている。この点もプログラムを作成する際は置換させて、わかりやすいものにしたいと思う。具体的には、「みず」は「さんずい、またはみず、みずへん」のように、その部首に含まれる読み方を全て説明させる方法を取れば、理解しやすくなると思う。

長所
- プログラムを作成する上で、部首名を検索し先頭に「部首は」を付け加え、発生させるだけなので、比較的簡単に済む。
- 求めたい漢字の部首は、入力する際に確認すればいいので、あまり時間がかからないように思える。

短所
- その部首に含まれる読み方を全て説明させるという方法を取るとすれば、元々登録されている状態のものを説明させる場合よりも説明が長くなる。
- 変換候補の中に同じ部首の漢字があった場合、この説明方法だけでは判別が不可能になる。
- 部首が、使用者の知識と登録されているものとで差異が発生することがあり得る上、実際に対象の漢字を想像した場合でも、その客観的な部首と登録されている部首が異なる場合がありえるので、正確な知識を備えていない場合、判別は難しいと言える。

- 登録されている部首の説明を、「みず」から「さんずい、またはみず、みずへん」のように置換した場合、「さんずい」のものと「みず」のものの区別が付けられないことになる。

短所の三番目のものは、例えば「処」の漢字の場合、見た目の部首は「すいにょう」に見える。しかし、登録されている部首は「つくえ」である。

対象の漢字一字での読み方を全て説明させる方法

変換候補に挙がった漢字を、個別に単漢字の状態での読み方を辞書ファイルから検索し、説明する方法。プログラムを作成する場合、最初に説明する読み方の先頭に「読み方は（よみかたは）」と付け加えたいと思う。読み方の検索を行なう場合、kakasidict から検索する。

例：あつし

厚：読み方は「あつ」「あつし」「こう」
淳：読み方は「あつ」「あつし」「きよし」「じゅん」
敦：読み方は「あつ」「あつし」「とん」
篤：読み方は「あつ」「あつし」「じゅん」「とう」「とく」

長所
- 対象の漢字の音読み、または訓読みで送り仮名をつけない状態の読み方で説明するため前述の二つの説明方法よりは判別が容易であると考えられる。
- 全ての読み方が同じものしかない漢字は少ないと思われるので、他のものと判別がしやすいと思われる。

短所
- 画数を用いた方法に比べて読み上げられる量が多いので、説明に時間がかかる。
- 数種類の読み方が存在する漢字ならば、特定することも可能かもしれないが、特定するには少ない数しか読み方が存在しない場合、判別が困難になると思われる。
- 求めたい漢字の区別と、全て同じ読み方を持つ漢字だった場合、この方法だけでは判別できないことになる。

短所の三番目のものは、例えば「漢」と「勘」の漢字の場合、いずれも単漢字での読み方は「かん」のみである。このような漢字を対象とした場合で、かつ対象の漢字一字での読み方を全て説明させる方法を用いる場合、これらの漢字の判別は不可能となる。

対象の漢字の含まれる熟語とその読み方を用いた説明

対象の漢字の含まれる熟語とその読み方を熟語ファイルから検索し、それを説明する方法。このとき、対象の漢字一字で構成されるものは除外する。プログラムを作成する場合、対象の漢字がその単語の中のどの位置にあり、そのときの読み方は何であるか、というところまで説明させたいと考えている。熟語の検索を行なう場合、 kakasidict から検索する。

例：あつし

厚：厚意（こうい）の 1 文字目
淳：淳子（じゅんこ）の 1 文字目
敦：敦賀（つるが）の 1 文字目
篤：危篤（きとく）の 2 文字目

この説明方法は、その読み方の漢字候補のひとつ目の説明をした時点で、使用者が理解でき、さらに求めている漢字であれば決定、求めている漢字でなければ次の候補へ、という流れになる。理解できなかった場合、その候補の他の熟語を用いて説明し、同様に決定、または次へ、と繰り返すことになる。

長所
- 熟語で説明するので、一般的に使用される形式に近いため、どのような漢字なのか連想しやすい。
- 対象の漢字の含まれる単語、という条件のものを羅列するので、使用できる説明の種類が多くなり、使用者にとってその漢字を連想しやすいものが説明される可能性が高くなると考えられる。

短所
- 熟語と、その読み方を説明すると言う方法を用いているので、前述のものと比べると説明が長くなる。
- 熟語が存在しない漢字もあるので、その場合説明できないことになる。

短所の二番目のものは、例えば「咸」の漢字の場合、単漢字としての読み方以外、 kakasidict には全く登録されていない。このような場合、対象の漢字の含まれる熟語とその読み方を用いた説明では説明できないことになる。

4　単漢字における統計

4.1　対象の漢字

今回の統計の対象は、 kakasidict から単漢字のみのものを検索し、その中から送り仮名を必要としないもののみとした。

例：視

- し
- み n
- み r
- み t

この漢字の場合、熟語を除く単漢字として kakasidict に登録されているのは以上の四種である。しかし、末尾がローマ字である場合、それは送り仮名を必要とするものであるので、ここでは除外する。この場合、「視」という漢字の一字の状態における読み方は四種あるが、今回の統計の対象となるのは「し」のみとなる。

4.2 読み方と漢字の種類の差異

対象の漢字は前述の通りであるが、その単漢字の種類は 6339 種、読み方の種類は 1966 種であった。しかし、ここで対象とした単漢字の種類と、JIS 第 1 水準と JIS 第 2 水準の合計である 6353 種に差異が生じている。この差異である 14 種は kakasidict に単漢字であり、かつ送り仮名を必要としない形で登録されていなかったものである。つまり、送り仮名を必要とする読み方か、 2 字以上の熟語としてしか登録されていないか、または登録されていない漢字ということになる。

この問題は、他の辞書ファイルを使用することで回避できるものなのかもしれない。しかし、今回の時点では、他の辞書ファイルを使用していないので、結論として報告することはできない。

この 14 種の漢字を表 1 にまとめる。そして、表のそれぞれの行について説明する。単漢字は対象の漢字である。画数と部首は、その漢字におけるそれぞれを示す。これは kakusu から調べたものである。音読みは大修館書店から出版されている「漢語新辞典」という漢字辞典より調べ、それを記載した。ただし、その辞典にすら単漢字における音読みが載っていなかったものがあるので、それらの漢字は「なし」と記載した。熟語等は対象の漢字が含まれる熟語の例である。これは kakasidict から検索したもので、そこに登録されていないものは「なし」と記載した。読み方は前項目である熟語等の読み方である。

表 1．単漢字として kakasidict に登録されていない漢字

単漢字	画数	部首	音読み	熟語等	読み方
匕	2 画	さじのひ	ひ	匕首	あいくち
				匕首	ひしゅ
悚	8 画	こころ	なし	悚	こら e
秉	8 画	のぎ	へい	秉	と r
苹	9 画	くさ	なし	なし	なし
唸	11 画	くち	てん	唸	うな r
				唸	うな t
				唸り	うなり
閊	11 画	もん	なし	閊	つか e
				閊	つか h
				閊	つっか e
				閊	つっか h
赧	12 画	あか	だん	赧	あから m
			ねん		
萍	12 画	くさ	ひょう	なし	なし
			びょう		
			へい		
楝	13 画	き	れん	なし	なし
裲	13 画	ころも	でい	裲襠	うちかけ
			ない		
			ね		
錬	16 画	かね	なし	教練	きょうれん
				試練	しれん
				精錬	せいれん
				洗錬	せんれん
				鍛錬	たんれん
				百戦錬磨	ひゃくせんれんま
錙	16 画	かね	なし	錙	かざ r
薜	17 画	くさ	ばい	なし	なし
			はく		
			ひゃく		
			へい		
			へき		
鶫	19 画	とり	とう	なし	なし

4.3　統計の対象

全ての読み方について、単漢字における画数を用いた方法と、部首を用いた方法の二つを使用した場合の、判別の可否の統計をとることにする。例えば、以下の三種の候補があったとする。

例：あおい

- 葵
- 青い
- 青井

本来はさらに数種の候補があるが、ここでは割愛する。このような候補があった場合、今回の対象となるものは一番最初の「葵」だけで、他の二種は対象外となる。

4.4　可否の基準

可否の判別の基準については、以下の通りである。

画数を用いた方法　：その読み方において、前後 x 画以上の差が開いているか

> 例えば x を 3 とした場合、ある読み方について様々な画数の漢字が存在するの中で、 5 画の漢字と 7 画の漢字が存在する場合のように画数同士の差が 2 画以下のものがある場合は不可とし、同様に最も差が少ない漢字同士でも 5 画の漢字と 8 画の漢字だった場合のように、その差が 3 画以上の場合は可とした。前者のように可の条件を満たしていないものがひとつでも存在した場合、その読み方について画数を用いた方法では説明が不可とした。

部首を用いた方法　：その読み方において、同じ部首のものが存在するか

> ある読み方について、同じ部首の漢字がひとつでもあった場合は不可とした。この条件に当てはまるものがひとつでもある場合、その読み方について部首を用いた方法では説明不可とした。

次に、それぞれの説明方法での可否の例を以下に挙げる。

例：あつし

厚：画数は 9 画、部首はがんだれ
淳：画数は 11 画、部首はみず
敦：画数は 12 画、部首はとまた
篤：画数は 16 画、部首はたけ

上記のような例で x を 3 とした場合、画数を用いた方法では不可とし、部首を用いた方法では可とした。局部的に見れば、使用者が「篤」の漢字を使用したい場合で、かつ x を 4 以下と設定していた場合のみ、画数を用いた方法でも説明することはできる。しかし、今回はそのような例外は考えないものとして統計を取ることにする。

4.5 統計結果

単独の場合

まず、全ての読み方を画数を用いた方法と部首を用いた方法を、それぞれ単独で使用した場合の統計を取った。その結果の内の、部首を用いた方法を使用した場合の説明可能な種類をまとめ、全ての読み方の種類も付随したものを以下の表 2 に示す。

表 **2** ． 単独での統計 部首を用いた方法

	数（種類）	割合（％）
全読み方	1966	
部首の可能数	1372	69.8

次に、画数を用いた方法を使用した場合の説明可能な種類を以下の表 3 にまとめ、全ての読み方の種類も付随する。今回は x を 1 から 5 としてそれぞれの統計を取った。

表 **3** ． 単独での統計 画数を用いた方法

	数（種類）	割合（％）
全読み方	1966	
$x = 1$	1559	79.3
$x = 2$	1386	70.5
$x = 3$	1273	64.8
$x = 4$	1201	61.1
$x = 5$	1143	58.1

さらに、画数を用いた方法と部首を用いた方法の、双方共通で説明可能なものを調べた。つまり、その読み方において、全ての漢字の画数の差が最小のものでも x 以上であり、どの部首についても同じものが他に存在しない読み方である。

同様に、双方共通で説明不可能なものも調べた。つまり、全ての漢字の画数の差が最小のものが x 未満であり、かつどれかの部首について他に同じものが存在している読み方である。これらの結果を以下の表 4 にまとめる。

表 **4** . 単独での統計 双方共通の結果

	双方可能（種類）	割合（%）	双方不可能（種類）	割合（%）
全読み方	1966			
$x = 1$	1291	65.7	326	16.6
$x = 2$	1212	61.6	420	21.4
$x = 3$	1159	59.0	480	24.4
$x = 4$	1124	57.2	517	26.3
$x = 5$	1089	55.4	540	27.5

さらに、画数を用いた方法でのみ説明可能な読み方と、部首を用いた方法でのみ説明可能な読み方を調べた。前者は、全ての漢字の画数の差が最小のものでも x 以上であるものの、どれかの部首について他に同じものが存在している読み方である。後者は、全ての漢字の画数の差が最小のものが x 未満であるものの、どの部首についても同じものが他に存在しない読み方である。この結果を以下の表 5 にまとめる。

表 **5** . 単独での統計 単独で可能な部分

	画数のみ可能（種類）	割合（%）	部首のみ可能（種類）	割合（%）
全読み方	1966			
$x = 1$	268	13.6	81	4.1
$x = 2$	174	8.9	160	8.1
$x = 3$	114	5.8	213	10.8
$x = 4$	77	3.9	248	12.6
$x = 5$	54	2.7	283	14.4

併用の場合

画数を用いた方法と部首を用いた方法を、併用した場合の統計を取った。この方法は、双方の方法を同時に使用することで、どの程度判別できる読み方の数が変化するのか調査

することを目的とした。画数を用いた方法の x の部分は、こちらも 1 から 5 とした。併用する場合の例を以下に挙げる。

例：あお

青：画数は 8 画、部首はあお
蒼：画数は 14 画、部首はくさ
碧：画数は 14 画、部首はいし
靜：画数は 16 画、部首はあお

上記のような場合で、別々に使用する場合、部首を用いた方法はもちろん、x の値をいくつにしようと画数を用いた方法でも判別は不可となる。しかし、併用した場合、 2 文字目の「蒼」と 3 文字目の「碧」は、画数が同じでも部首が異なっているので、判別可となる。同様に、 1 文字目の「青」と 4 文字目の「靜」が部首が同じであるが、 x の値が 8 以下であれば画数は十分に離れていると言えるので、判別可となる。

このように、それぞれの漢字を画数だけ、部首だけで見るのではなく、「画数 x 画で、部首は y の漢字」という情報を持った漢字として可否の判別を行なった。その結果を以下の表 6 にまとめる。

表 **6** ． 併用での統計 全体の結果

	数（種類）	割合（％）	不可能な数（種類）
全読み方	1966		
$x = 1$	1785	90.8	181
$x = 2$	1669	84.9	297
$x = 3$	1580	80.4	386
$x = 4$	1518	77.2	448
$x = 5$	1477	75.1	489

この結果と、画数を用いた方法単体の結果と、部首を用いた方法単体の結果を比較する。画数を用いた方法と比較する場合、 x の値が等しいものを比較する。部首を用いた方法と比較する場合、画数を用いた方法単体の平均値として挙げた $x = 3$ の条件の双方を併用した場合のものと比較する。

前述の表を見ればわかることだが、双方を併用した場合の結果だけに焦点を合わせれば、今挙げた条件と画数、部首を用いた方法単体とを比較した場合、全て単体のものよりも優れていることになる。したがって、ここでは正負の符号は不要で、全てプラスということになる。その結果を以下の表 7 にまとめる。

併用での統計 単独との差

	数 (種類)
画数 x = 1	226
画数 x = 2	283
画数 x = 3	307
画数 x = 4	317
画数 x = 5	334
部首	208

4.6 健常者に対してのアンケート

健常者に対して画数と部首についての漢字のアンケートを取った。対象は当大学に在学中の学生 10 人である。対象の単漢字と kakusu に登録されている画数、部首は以下の表 8 の通りである。部首について、カッコに囲まれている部分は、本来 kakusu には登録されていない部分だが、今回はわかりやすく読めるように追加した。

表 8 . アンケート対象の漢字

単漢字	画数 (画)	部首
処	5	すいにょう
遠	13	しんにゅう
登	12	はつがしら
関	14	もん (がまえ)
就	12	まげあし
端	14	たつ (へん)
秘	10	のぎ (へん)
験	18	うま (へん)
海	9	みず (さんずい)
蓮	15	くさ (かんむり)

画数の結果と部首の結果は別々に示す。

まず、画数についての結果は以下の表 9 にまとめる。単漢字の列の− 2 から + 2 は、解答された画数と登録されている画数との差を示してしる。それぞれの漢字の列にある数字は、解答された人数を示している。

表 9．アンケート結果 画数

単漢字	− 2 以下	− 1	0	+ 1	+ 2 以上
処	0	0	8	1	1
遠	0	8	2	0	0
登	0	0	9	1	0
関	0	1	8	1	0
就	0	0	10	0	0
端	0	0	8	2	0
秘	0	0	10	0	0
験	0	0	10	0	0
海	0	0	6	4	0
蓮	10	0	0	0	0

次に、部首についての結果を以下の表 10 にまとめる。正答者とは解答した部首部分が登録されているものと同じだった人数で、部首名称無解答とは正答者の内で、名前を挙げずその部首をマルで囲む等の方法で示した人数である。誤答者は登録されている部首名と違うものを答えた人数である。

表 10．アンケート結果 部首

単漢字	正答者	部首名称無解答	誤答者
処	0	0	10
遠	10	3	0
登	9	9	1
関	10	6	0
就	3	3	7
端	10	9	0
秘	10	5	0
験	10	9	0
海	10	3	0
蓮	10	4	0

4.7　統計とアンケート結果についての考察

今回取った統計についての考察を、それぞれ以下に述べる。

画数を用いた方法

画数を用いた方法のみを使用した場合、判別可能な読み方は x の値次第で変化する。しかし、おおまかに数えて平均値を取る $x = 3$ の場合でも 1273 種類、約 64.8 %と部首を用いた方法を下回っている。

統計を取る前の段階では、部首を用いた方法よりも画数を用いた方法の方が説明には適していると予想した。これは、部首を説明されたところで使用者がその部首を思いつくことが難しいと考えたためである。逆に、画数を説明されることを考えると、使用者はその都度数えなければならないという手間はあるものの、ほぼ確実に求めたい漢字にたどり着けると考えたためである。

実際、健常者に対して行なったアンケートの結果から、 1 画以内の正答率を見ると、ほぼ全ての漢字について 90 %以上の正答が求められている。一部実際に書く場合の画数と登録されている画数との差異が激しいものもあるようだが、その漢字もあまり多くはないようなので、その部分だけ特例の説明を作成する等の処置を行なえば良いと思う。

しかし、統計を取ってわかったことだが、今回対象とした単漢字の画数は 1 画から 30 画である。言い換えれば、画数は 30 種類しか存在しないことになる。しかし、部首は 214 種類存在する。したがって、部首よりも画数の方が、同一の種類として登録され、判別不可能とされる割合が高いことになる。しかも、 x の値次第ではさらに種類が圧迫されることになる。

画数を用いた方法単体で行なった統計結果の表を見てもらえばわかると思うが、x の値を小さくすればするほど可となる読み方が増えている。今回統計を取った中で、不可となる範囲を最も狭くした場合である $x = 1$ の状態での判別可能な読み方は 1559 種類で、全体の 79.3 %である。この条件ならば、部首を用いた方法よりも優れた結果を出せることになる。$x = 2$ の状態でも部首を用いた方法とほぼ同じ数が可能と判別された。しかし、実際に使用することを考えると、 x の値を小さくしすぎるのはあまり現実適ではないと言わざるを得ない。使用者の知識に関わることになるが、全ての漢字をほぼ正確に記憶しておくことのできる人間は非常に少ないと思われるからである。これは、健常者に対して行なったアンケートの結果からもわかると思う。平均値として挙げた $x = 3$ の条件の場合、統計の結果では 693 種、約 35 %の読み方に対してこの方法単体では説明不可能となる。しかし、大半の漢字に対して 90 %以上の正答率が得られている。 $x = 1$ とした場合、つまり登録されている画数との誤差が 0 画とした場合、その正答率は約 70 %程度になる。統計として得られる可能な範囲が広かったとしても、正答率が低ければ無意味と言える。

したがって、画数を用いた方法のみを使用した場合、 $x = 3$ の条件で約 65 %の読み方を説明することができ、そのうち約 90 %が正しい漢字にたどり着けると言える。

部首を用いた方法

部首を用いた方法のみを使用した場合、判別可能な読み方は 1372 種類、約 69.8 %であった。これは、統計結果だけを見た場合、画数を用いた方法の、 $x = 2$ の状態とほぼ等しく、 $x = 3$ 以上であればこちらの方が優れていることになる。しかし、優れていると言っても、 594 種類、約 30 %の読み方に対してこの方法単体では説明不可能となる。そして、健常者に対して行なったアンケートの結果、部首を用いた方法での正答率は約 80 %であると言える。

これらを踏まえると、約 50 %程度は求めたい漢字を求めることができることになるが、これでは約 1000 種類もの読み方が不可能となる。しかも、これは仮定であり、実際は更に不可能な数は増えると思われるので、この方法単体での説明は難しいと言える。

双方を併用した場合

双方を併用した場合、今までの説明単体よりも情報量が増える。したがって、x の値が等しい条件であれば、当然併用した場合の方が優れた結果が出た。部首を用いた方法でも同様に、 x の値をいくつにしたとしても、双方を併用した場合のほうが優れた結果が出た。

しかし、あくまで「画数、または部首を用いた説明方法よりも優れていた」という結果を出しただけに過ぎず、平均値として挙げた $x = 3$ の条件では 386 種類、約 20 %の読み方が説明不可能となっている。

この方法で説明不可能と判別される読み方の種類が 50 種類前後であれば、それらに対して特例の説明を作成することも考えられた。こうすることで、この方法で全ての単漢字における読み方を説明できることになるからである。しかし、これほどの量になると、手作業で特例の説明を作成するにはかなりの時間と労力が必要となる。それでは今回の研究の趣旨から逸脱してしまうので、画数と部首を用いる方法の双方を併用した場合であっても、視覚障害者向けかな漢字変換ソフトとして、単漢字を説明するには難しいと言える。

アンケートについての考察

まず、画数についてだが、表を見ればわかる通り、「遠」と「海」と「蓮」の漢字の正答率があまり高くない。これはそれぞれ、

- 「遠」のしんにゅう部分の画数は本来 3 画だが、アンケート対象者の大半は 2 画として書いていたものと思われる。
- 「海」のつくりの下の部分、つまり「母」のような部分は、本来 4 画だがアンケート対象者は 5 画と書いたひとが多かったようだ。これは、「口」の漢字と同様に、「母」を書いたため、 1 画多く書いてしまったものだと思われる。
- 「蓮」の場合、まず「遠」と同様にしんにゅう部分である。しかし、「蓮」の場合、しんにゅうは 4 画として登録されている。これは、しんにゅうが旧字体の場合、点が二つついているからである。次にくさかんむり部分である。これは一般的に漢字を書く場合、 3 画で書ける部分であるが、旧字体では横棒が離れており、 4 画として登録されている。これらの条件により、アンケート対象者の全員が画数を大きく間違えてしまったものだと思われる。

このような理由によるものだと思われる。しかし、漢字辞典によってはこのような漢字の画数は、今回基準として用いた kakusu とは異なって登録されていることもある。

次に部首についてだが、表を見ればわかる通り、「処」の漢字は正答者がいなかった。さらに、「就」の漢字についても正答率が高いとは言えない。これはそれぞれ、

- 「処」の漢字の場合、前述の 3.2 節の説明方法の候補の部首を用いた方法の項目にある例で説明した通り、見た目の部首は「すいにょう」であるが、登録されている部首名は「つくえ」である、という理由。
- 「就」の漢字の場合、つくりが部首であるが、この「まげあし」でなく、へんの「京」の部分の方が部首だと思ったアンケート対象者が多かった。これは、「まげあし」という部首を一般生活ではあまり見かけないためだと思われる。

以上のような理由のためだと思われる。さらに、部首名称無解答が多いという結果も出た。これは、単純に部首の名前が思い出せなくて解答できなかったというひとと、漢字を図として見て「正確な知識はないものの、予想としてはこの部分だと思う」という方法で選んだひとに分かれている。前者は部首の名称を言われれば気付ける場合も多いと思われるが、後者の場合言われても気付けない場合が多いと思われる。

5 まとめ

今回の研究は、ある読み方について、多数存在するであろう単漢字の画数、または部首、またはそれら双方を説明させると仮定した場合、どの程度の数の読み方が説明可能であるかを、統計として調べた。これは、視覚障害者向けかな漢字変換ソフトを作成する上で、基礎段階として必要だと思われることをしたにすぎない。他に挙げた説明の候補であ

る、対象の漢字一字での読み方を全て説明させる方法や、対象の漢字の含まれる熟語とその読み方を用いた説明については、統計すら取ることができなかった。

今回の統計結果では、$x = 3$ を画数を用いた方法の平均値とした場合、それぞれ単体での説明だったとすれば、部首を用いた方法の方が優れているという結果が出た。また、双方を併用した場合の結果は、説明可能となる数は増えるものの、全ての読み方を説明するには不足であると言えるという結果が出た。

また、今回の研究では送り仮名を必要としない単漢字のみを対象としたが、実際に視覚障害者向けかな漢字変換ソフトを作成するとすれば、それら除外した条件のものも説明ソフトとして必須であるので、様々な課題が残されていると言える。

残念ながら今回の研究では、最後まで進めることができなかったが、今後は他の説明候補の統計を取り、また、さらに新たな説明方法を提示し、それの統計も取り、適切な説明を配置し、全ての読み方について説明できるかな漢字変換ソフトの開発ができれば、と思う。

参考文献

[1] 井上 直大：日本語文書中の漢字の説明ソフトの開発について (新潟工科大学卒業論文 pp1-28 2000 年)

[2] 小山 裕司、佐々木 浩、斉藤 靖、中込 知之： UNIX 入門 (株式会社トッパン 1998 年)

[3] Paul DuBois：入門 csh & tcsh (鈴鹿 倫之、福澤 康裕 訳 オライリー・ジャパン 2002 年)

[4] 鎌田 正、米山 寅太郎：大修舘 漢語新辞典 (大修舘書店 2001 年)