動詞・形容詞での
単漢字説明ソフトウェアについて

平成 18 年 1 月 31 日

情報電子工学科
早川 道則

目 次

1 はじめに **1**

2 視覚障害者の現状 **1**

3 単漢字の説明方法 **2**

3.1 日常使われる説明方法 2

3.2 フリーソフトウェア 2

3.3 有用なフリーソフト 3

3.3.1 かな漢字変換辞書ファイル 3

3.3.2 yomi 5

3.4 有用な UNIX 上のコマンド 5

4 過去の卒業研究の問題点について **6**

4.1 過去の卒業研究について 6

4.2 過去の卒業研究結果について 7

5 説明方法の提案 **8**

5.1 動詞・形容詞での説明方法 8

5.2 動詞・形容詞について 9

5.3 問題点 10

5.4 説明ソフトについて 11

5.4.1 基本的な説明手順 11

5.4.2 説明ソフトのユーザーインターフェース 12

6 判別可能な漢字とアンケート **14**

6.1 説明可能な単漢字 14

6.1.1 対象の単漢字 15
6.2 判別方法 16
6.3 健常者に対してのアンケート 17

7 判別結果とアンケート結果 **18**

7.1 判別結果 18
7.2 アンケート結果 19
7.3 判別可能な単漢字とアンケートの考察 21

8 まとめ **23**

参考文献 **24**

概要

現在、 UNIX 上では単漢字の読み上げ説明ソフトはあまり普及していないといえる。そこで、本研究では熟語や単漢字を説明するためにそれぞれ固有の説明を持つ辞書ファイルを用いない別の説明方法を考察していくことにする。過去の卒業研究では、 kakasidict という辞書ファイルから対象の単漢字が含まれる単語を抜き出し、対象の単漢字がその単語の何文字目に使われているかという説明方法や、対象の単漢字の部首や画数で説明する方法を提案、考察していた。

本研究では、対象の単漢字を辞書ファイルから動詞・形容詞として収録されている対象の単漢字を辞書ファイルから抜き出し説明する方法を提案する。この説明方法では、どの程度の単漢字が説明可能なのかを調べるため JIS 第一水準と第二水準の単漢字を対象に判別していくと同時に、健常者に対してこの説明方法でどの程度理解ができるのかアンケートをとり、考察を行った。

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であろう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や音声ソフトは現在、 MS–Wiondows や MS–DOS などでは普及されつつあるも、 UNIX 上ではあまり普及されていないのが現状だといえる。

視覚障害者の人達にとって、パソコン上で単漢字を確認するためには対象の単漢字を一文字一文字音声で読み上げて説明してもらうか、点字で確認する等が必要となり、同音異義語を判別するのは難しいといえる。本研究では、ある程度漢字の知識を持っている視覚障害者を対象とした単漢字説明ソフトウェアの作成を目標とする。

漢字説明ソフトを作成する場合、単漢字や熟語の一つ一つにそれぞれ固有の説明を加えた辞書ファイルを作成し、熟語や単漢字ごとに固有の説明をする方法が一番良いが、今回はそのような辞書ファイルを用いずに別の方法で説明できないか考える。

過去の卒業研究では、井上氏 [1)] 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単漢字があれば効率よく説明できるかを調べ、山口氏 [2)] は単漢字を画数や部首を用いてどの程度説明できるか調べた。今回はそれらの方法とは別に現在あるフリーなソフトや辞書ファイル等を使用し、別の説明方法を考察していくことにする。

2　視覚障害者の現状

視覚障害者の人達が現在どのようにパソコンやインターネットを利用しているのか調べてみた。現在、ほとんどの視覚障害者の人は OS として、 MS–Windows を主に使用しているようである。一昔前まで主流だった MS–DOS や本研究が目標にしている UNIX を使用している人もいるが、 MS–Windows に比べると少ないようだ。しかし、 UNIX の開発は現在も進んでおり、今後さらに普及する可能性がある。

そこで、今回対象となる視覚障害者の UNIX 環境の現状についてまとめてみると、UNIX 環境の場合、 OS はパソコンで動作する FreeBSD や Linux 等が使用されているようであるが、これらに対応している UNIX 上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らしい。

全盲の視覚障害者の方が UNIX を使用する利点としては、 MS–Windows はマウス等を主に使用するグラフィカル ユーザー インターフェース (GUI) であり、支援ソフトは普及しているものの全盲の人達にはあまり向いていないといえる。

視覚的に対応する GUI に対し、 MS–DOS や UNIX ではコマンド等で操作を行なうキャラクター ユーザー インターフェース (CUI) なので、視覚的な操作を行なわないこち

らの方がむしろ視覚障害者の人達には向いているのではないかと考えられる。さらに現在は開発あまり行なわれていない MS–DOS の状況とは違い、 UNIX は現在もユーザーが増えつつあり、新しいソフトの開発は現在も行なわれている。

3　単漢字の説明方法

3.1　日常使われる説明方法

視覚障害者の人達に漢字を説明する場合、視覚的な情報で伝えることはできず、最も伝えやすいと思われるのは言葉で伝える方法である。つまり、我々が日常で言葉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。

市販の説明ソフトウェアでは、説明するためにそれぞれ固有の説明を持つ辞書ファイルを作成し、その中から熟語や単語を説明するときに固有の説明を持ち出し説明しているようである。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」

といった説明になる。つまり、説明するためにはその漢字の意味を伝えれば良い訳だが、そのような各熟語や単漢字に特有の説明を持たせているフリーの辞書ファイルは見当たらない。

本来はそのような辞書ファイルを用いた説明ソフトを作成すれば良い訳だが、今回はそのような辞書ファイルや説明方法を用いず、別の説明方法を考察する。

3.2　フリーソフトウェア

フリーソフトウェアとはユーザがその扱いを「自由」にできるソフトウェアのことで、「自由」の定義は文脈によって異なる。この「自由」という内容には主に 2 つの意味があり、「無償で利用できる」又は、「ソースコードが入手でき、改変・再配布が制限なく行なえる」という意味で使われることが多い。

「無償で利用できる」とは主に、1990 年代の日本で、パソコン通信を利用して形成されたソフトウェア開発者のコミュニティで定着した概念で、従来の商業ベースのソフト

ウェアとの対比で「無料で利用できる」ことに力点が置かれているソフトウェアのことである。

「ソースコードが入手でき、改変・再配布が制限なく行なえる」とは主に、インターネットの普及とともに国境を越えて形成されたソフトウェア開発者のコミュニティで広まってきた概念で、無料で利用できるだけでなくソースコードの入手や、その改変・再配布、派生ソフトウェアの開発と公開、さらには、ソフトウェアを「販売する自由」さえ含む、一切の制約が排除された状態を重視するソフトウェアのことである。また、フリーソフトウェアの中でも特に、ソースコードが公開されているソフトウェアのことを「オープンソースソフトウェア」と呼ぶこともある。

3.3 有用なフリーソフト

ここでは今回の研究で使用した、又は今回は使用しなかったが、さらに研究を進めるにあたって有用であろうフリーソフトウェアや辞書ファイルを紹介していく。

3.3.1 かな漢字変換辞書ファイル

canna

canna は今 昭氏と NEC に勤務する数人によって開発された日本語入力システムである。canna は通常、システム辞書を用いて漢字変換を行うが、各ユーザーが辞書を作成し、システム辞書の内容を変更することができる特徴を持つ。付属されている辞書ファイルはフリーの辞書 pubdic+ をベースに作成されたものを使用している。

辞書ファイルには以下の形式で登録されている。

あぶな 危な #KY 23
あぶなっかし 危なっかし #KY 0
あぶのう 危のう #KYU 2
あやう 危う #KY 2

kakasi

kakasi とは、漢字仮名まじり文をひらがな文やローマ字文に変換することを目的として作成されたプログラムと辞書の総称であり、高橋 裕信氏が作成したものである。また付属の辞書ファイル、 kakasidict は高橋氏及び佐藤氏等が SKK version7 に付属する辞書

と kakasi1.2 のために作成された辞書から kakasi のために若干改良を加えたものである。また、 kakasidict には読み方が多い単語には優先順位がついており、単語を変換したときに先に出力される読みはある程度決ってくる。

辞書ファイルには以下の形式で登録されている。

き 危
あぶな k 危
きぐ 危ぐ
きがい 危害

kakusu

1 画から 30 画までの 6353 個の単漢字が収録されている辞書である。登録されている漢字は全て単漢字のみで、熟語や送り仮名を含むものは登録されておらず、単漢字とその画数、部首は登録されているが、そ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。

辞書ファイルに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)

SKK

SKK は emacs 上に高速で効率的な日本語入力環境を提供するシステムで、 GNU General Public License に従ったフリーソフトウェアとして配布されている。SKK の辞書ファイルは、約 80 名の SKK ユーザーから提供してもらった辞書ファイルと、 JUNET の記事で使用頻度の高い単語に読みを与えたものを元に作成されたものであり、この辞書はフリーの辞書としては最大規模の単語の登録数となっている。

辞書ファイルには以下の形式で登録されている。

あや m /殺; 人を殺める/危; ≒殺める/怪;–む (いぶかる)/
あや b /危/
あぶな s /危; 非正則 (あぶ n)/
あぶな k /危; 非正則 (あぶ n)/

3.3.2 yomi

yomi とはこの研究室で開発されたテキストファイルを単純な方法で音声化するためのソフトウェアである。音声出力の方法は、単にテキストファイルをひらがなの文書に直し、それに対応するひらがなの音声データをつなぎ、音声デバイス等に出力する単純な方法である。

3.4 有用な UNIX 上のコマンド

本研究では、 UNIX 上で動作するソフトを目標にしている。したがって、 UNIX 特有のコマンドを使用したソフトウェアを作成できないかと考えた。ここでは、今回使用する UNIX 上のコマンドを紹介していく。

grep 特定の文字列を含む行を参照したい場合に使うコマンド。テキストファイルから指定されたパターン (文字列) を含む行を抜き出し、結果を標準出力に出力する。ファイル名はワイルドカード等を用いて一度に複数のファイルを指定できる。

awk テキストデータの中から指定パターンを検索する場合に使うコマンド。読み込んだテキストデータ中に指定されたパターンがないか照合し、一致するパターンが見つかった場合、指定された処理を行なう。ファイルが指定されるとそのファイルからデータを読み込み、ファイルが指定されない場合は、標準入力からデータを読み込み、処理結果は標準出力に出力する。

sort テキストファイルを行単位にソートする場合に使うコマンド。ファイル名に指定されたテキストファイルをソートして、その結果を標準出力に出力する。ファイル名が指定されない場合は標準入力からデータを読み込み、他のコマンドの出力をパイプで sort コマンドに連結すると、他のコマンドの出力をソートできる。

cat ファイルの内容を表示したい場合に使うコマンド。引数として指定されたファイルの内容を標準出力に出力する。複数のファイルを引数として指定すると、指定されたファイルの内容を連結して表示する。

head ファイル内容の先頭だけを表示したい場合に使うコマンド。指定されたファイルの内容を、先頭から指定された行数分だけ表示する。行数が指定されなかった場合は、デフォルトの 10 行分が表示される。

4　過去の卒業研究の問題点について

4.1　過去の卒業研究について

以前、同様の研究を行った井上氏[1)]は、辞書ファイルとして kakasi という漢字や仮名をひらがなに変換するソフトの付属辞書を用いて、その辞書ファイルと UNIX 上のコマンドを使用した説明方法として

- 別読みで説明
- 複数読みで説明
- 辞書ファイルからの単語の抜き出しによる説明

の 3 つを提案し、井上氏[1)]はこのうち、辞書ファイルからの単語の抜き出しによる説明について詳しく考察を行った。例えば『全快』を説明する場合、まず「全」の単漢字を含む単語を辞書ファイルから抜き出してくる。

ぜんじつ 全日
ぜんち 全治
・
・

そして、対象の単漢字が抜き出した単語の何文字目に使われているかを説明していく。「全日」を例にすると、

『全快』の一文字目は「ぜんじつ」の一文字目

のように説明する方法である。もし、一つ目で理解ができない場合は二つ目の抜き出した単語を例にしてもう一度説明する。

この方法の結果としては、熟語の多い少ないはあまり関係なく、説明する単漢字が先頭、途中、最後の三つの中では、途中に単漢字がある熟語が一番理解を得られるという結果を出している。

また、山口氏[2)]も、 kakasi や kakusu のソフトウェアまたは辞書ファイルと UNIX 上のコマンドを使用した説明方法として

- 画数で説明
- 部首で説明
- 両方を用いて説明

をそれぞれ考察した。山口氏 [2] は部首と画数を説明するために kakasi と kakusu の辞書ファイルを使用した。

部首で説明する方法は、 kakusu の辞書ファイルから対象の単漢字の画数を検索し、

厚：画数は 9 画
淳：画数は 11 画
敦：画数は 12 画

というように説明していく方法である。また、部首で説明する方法は、kakasi の辞書ファイルから対象の単漢字の部首を検索し、

厚：部首は がんだれ
淳：部首は みず
敦：部首は とまた

というように説明していく方法である。

結果として、画数だけの説明方法では単漢字の画数が前後 3 画の差がある単漢字がない場合は約 65% が説明可能という結果になった。部首を用いた方法では、説明可能な単漢字は約 69.8% となり、画数と部首の両方を使用し、画数は単漢字の画数が前後 2 画の差もない場合とすると、約 80% の単漢字が説明できるという結果を得た。

4.2 過去の卒業研究結果について

過去の卒業研究の結果について、井上氏 [1] は辞書ファイルからの単語の抜き出しによる説明の途中に単漢字がある説明方法だけで、かなり良いところまで説明できるのではないと考察している。しかし、あまり一般的ではなく熟語の数が一つや二つしかない、どこかの地名にでも使われるような単漢字にはこの方法では説明不可能としている。山口氏 [2] の画数や部首を用いて説明する方法でも、両方を用いた説明方法でも約 80% の単漢字しか説明できず、その他の約 20% の 386 個の単漢字には別の方法が必要としている。

つまり、両氏の説明方法単体だけでは十分なソフトとして動作できないということになり、その他の説明方法を考える必要があると思われる。今回の研究では、新たな説明方法

として単漢字を動詞・形容詞に変えて説明するという方法を提案し、それについて考察していくことにする。

5　説明方法の提案

5.1　動詞・形容詞での説明方法

今回の研究では説明方法として単漢字を動詞・形容詞に変えて説明するという方法を提案した。例えば『危険』を説明しようとしたときに

「危」なら、あぶない。「険」なら、けわしい。

のようにそれぞれ単漢字を動詞・形容詞に変えて

あぶない けわしい

と説明する方法である。しかし、対象の単漢字が動詞・形容詞として使われていない場合は当然この説明方法では説明不可能となる。

この方法は canna という UNIX 上で使用される日本語入力システムの辞書ファイルを使用し、そこに登録されている動詞・形容詞を抜き出して説明をしようと考えている。canna の辞書ファイルには以下のような形式で登録されている。

あぶな　危な　#KY 23

この #KY というのは canna で使われている品詞コードであり、この品詞コードを使用して辞書ファイルから動詞・形容詞を抜き出せると考えた。また、これら品詞コードは大まかに分類すると名詞、団体・会社名、人名、地名、動詞、形容詞、副詞、単漢字、数詞、連体詞、接続詞・感動詞の 11 種類に分類でき、さらに動詞は 23 種類、形容詞と副詞は 4 種類ずつ細かく分類されている。

5.2　動詞・形容詞について

動詞

　動詞とは、自立語で活用があり、動作・作用・存在を表す単語を言い、述語になる単語であり言い切りの形が『う』の段で終わるものである。また、自立語とは一つの単語で意味がわかり、文節をつくることができる単語であり、活用とは単語の語形が変化することをいう。

例：笑う (u)、書く (ku)、寝る (nu)

　活用には未然形、連用形、終止形、連体形、仮定形、命令形がある。動詞の活用の種類とその例を表 1 にまとめた。

表 1

種類	基本形の例	語幹	未然形	連用形	終止形	連体形	仮定形	命令形
五段	読む	読	ま、み	み、ん	む	む	め	め
上一段	見る		み	み	みる	みる	みれ	みろ、みよ
下一段	食べる	食	べ	べ	べる	べる	べれ	べろ、べよ
カ変	来る		こ	き	くる	くる	くれ	こい
サ変	する		し、せ、さ	し	する	する	すれ	しろ、せよ

　語幹は活用しても変らない部分を言い、語幹の欄が空白なものは活用によって変わる部分と変わらない部分の区別がつかないもので (例：見る、煮る、着る) その他は

未然形は『ない』『う・よう』につながる形
連用形は『ます』『た』『だ』につながる形
終止形は言い切りの形
連体形は『とき』につながる形
仮定形は『ば』につながる形
命令形は命令の意味になる形

となっている。

形容詞

形容詞とは、自立語で活用があり、性質・状態を表わす単語を言い、述語になる単語であり言い切りが『い』で終るものである。

例：白い、固い、広い、

形容詞の活用の種類を表 2 にまとめた。

表 **2**

基本形	語幹	未然形	連用形	終止形	連体形	仮定形
美しい	美し	かろ	かっ、く	い	い	けれ
高い	高	かろ	かっ、く	い	い	けれ

語幹の意味は動詞と同じで、その他は

未然形は『う』につながる形
連用形は『ない』『た』『なる』につながる形
終止形は言い切りの形
連体形は『とき』につながる形
仮定形は『ば』につながる形
命令形は『とき』につながる形

となっている。

5.3 問題点

canna の辞書ファイルを使用して動詞・形容詞で単漢字を説明する場合、canna の辞書ファイルには送りがながない状態で収録されているという問題がある。例えば、この説明方法で『危険』を説明しようとすると

あぶな 危な　#KY 23
あぶなっかし 危なっかし #KY 0
あやう 危う #KY 2
あやぶ 危ぶ #M5 1

けわし　険し　#KY 2

というように抜き出せるが、元々送りがなのない状態で辞書ファイルに収録されているので、送りがながないまま『あぶな、けわし』と読み上げてしまう。このまま送りがながない状態で説明しても理解が得られないと思われる。

そこで、考えられる解決方法として、各品詞コードに特定の送りがなを加えていけば良いと考えた。 canna の動詞・形容詞の品詞コードはカ行、サ行のように各行ごとに品詞コードが分類されているので、それぞれの行の『う』の段のひらがなを送りがなとして付け加える。例えばカ行五段活用の品詞コードならカ行の『う』の段、つまり『く』を加えれば良い訳である。

動詞の言い切りが各行の「う」の段、形容詞の言い切りが「い」となっており、この辞書ファイルの品詞コードは各行ごとに分類されているので、その行の「う」の段のひらがなを加えていけば良いわけである。例えば、#KY なら形容詞の品詞コードなので送りがな「い」を加える。#M5 ならマ行五段活用の品詞コードなので送りがなにマ行の「う」の段、つまり「む」を加えと、

あぶない 危な　#KY 23
あぶなっかしい 危なっかし #KY 0
あやうい 危う #KY 2
あやぶむ 危ぶ #M5 1

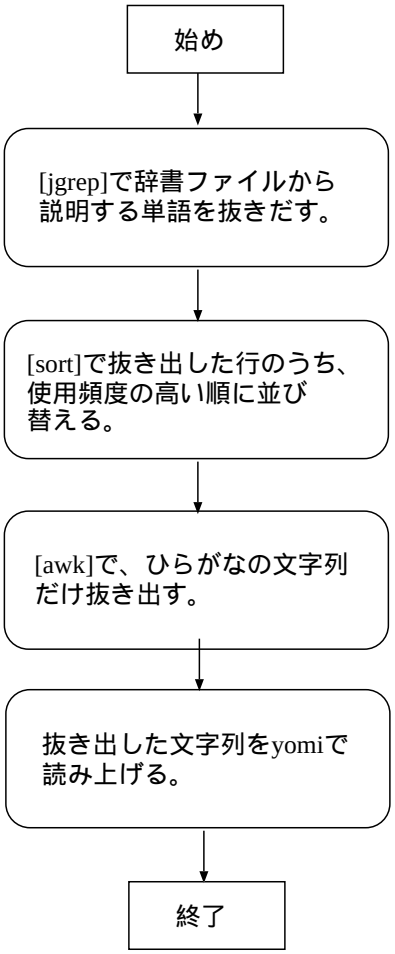
けわしい　険し　#KY 2

というように送りがなを加えていくことができる。ひらがなだけに送りがなを加えているのは、 yomi に読ませるのはひらがななのでひらがなだけに送りがなを加えている。このように各行の「う」の段や「い」をそれぞれの品詞コードに合う送りがなを加えていくプログラムを作成し、送りがなを加えた canna の辞書ファイルを使用していく。

5.4　説明ソフトについて

5.4.1　基本的な説明手順

送りがなを加えた辞書ファイルを使用し、説明方法の基本的な手順を考えていく。基本的な説明の手順としてはまず、UNIX のコマンドを使用し、辞書ファイルから説明する単漢字の含まれる行を抜き出し、それを使用頻度の高い順に並び替え、それを yomi で読みあげるという手順となる。その手順を図 1 にまとめた。



例：危

あぶない 危ない #KY 23
あぶなっかしい 危なっかしい #KY 0
あやうい 危うい #KY 2
あやぶむ 危ぶむ #M5 1

あぶない 危ない #KY 23
あやうい 危うい #KY 2
あやぶむ 危ぶむ #M5 1
あぶなっかしい 危なっかしい #KY 0

Fig. 1 基本的な説明手順

5.4.2 説明ソフトのユーザーインターフェース

上記の基本的な説明方法をもとに、ソフトを作成する場合、一つの単漢字の説明を一度に読み上げる方法と、漢字の説明文を一つずつ読ませ、理解した時点でその単漢字の説明を終了するという二つの方法を考えた。

これらを説明ソフトのユーザーインターフェースの例として、「動作」を説明した場合は以下のようになる。

うごく 動く #K5r 255
うごかす 動かす #S5 92
うごきだす 動き出す #S5 4
どうずる 動ずる #ZX 3

つくりだす 作り出す #S5 5
つくる 作る #R5 255
つくりあげる 作り上げる #KS 3

これらが抜き出され、次にこれら抜き出した動詞・形容詞を先程の方法で説明していく。

一度に説明する方法

一度に説明する方法は、抜き出した動詞・形容詞を一度に全部読み上げ、理解できれば次の単漢字へ進み、できなければもう一度全部読上げる。つまり、先程の『動作』を説明したいならまず、

うごく
うごかす
うごきだす
どうずる

と、「動」が使われている全ての動詞・形容詞を一度に読みあげ、理解が得られなければもう一度「動」が使われている動詞・形容詞を一度に読みあげる。対象の単漢字を理解した時点で次の漢字、

つくりだす
つくる
つくりあげる

という風に「作」が使われている全ての動詞・形容詞を一度に読みあげていく方法である。最終的には、読み上げている途中でも理解できたら終了できるようにしたい。

長所 動く、動かす、動き出すといった風に、一度に説明文を全部読み上げるので、その単漢字を連想しやすいのではないかと思う。

短所 読み上げている途中で理解しても、その時点で説明を終了できない。また、3、4 個しかない単語なら良いが、中には 10 以上の説明文を含む単語もあるので、その場合途中で中断できないなら読み上げが終了するまで待たなければならない。

一つずつ説明する方法

一つずつ説明する方法は、一つだけ動詞・形容詞の説明文を読み上げ、理解した時点で説明は終了し次の単漢字へ、理解できなければ次の動詞・形容詞の説明文を読み上げるという方法である。『動作』を例にあげると、まず「動」が使われている動詞・形容詞で一番使用頻度が高いものを読みあげる。

うごく

この一つ目の動詞・形容詞で対象の単漢字が理解できなければ、二番目に使用頻度の高い動詞・形容詞、

うごかす

を読みあげる。このように理解が得られるまで一つずつ動詞・形容詞を読みあげる方法である。対象の単漢字「動」だと理解が得られたならば、次の単漢字「作」を、

つくりだす

というようにまた一つずつその対象の動詞・形容詞を読みあげていく。

長所 自分のペースで説明を聞けて、考えながら対象の単漢字を想像できるので、理解しやすいのではないかと思う。

短所 次の説明を受るためにはユーザーがいちいち操作しなくてはいけない。これは視覚障害者の人達にはなるべくそのようなわずらわしいキーボード上の操作を行なわせない方法の方が良いと思うからである。

その二つの方法を図 2 にまとめてみる。

6 判別可能な漢字とアンケート

6.1 説明可能な単漢字

Version3.7 の canna の辞書ファイルには 42011 個の単語が収録されている。このうち、動詞・形容詞は 3972 個含まれている。この 3972 個の動詞・形容詞でどの程度の単漢字が説明可能なのかを調べていくことにする。

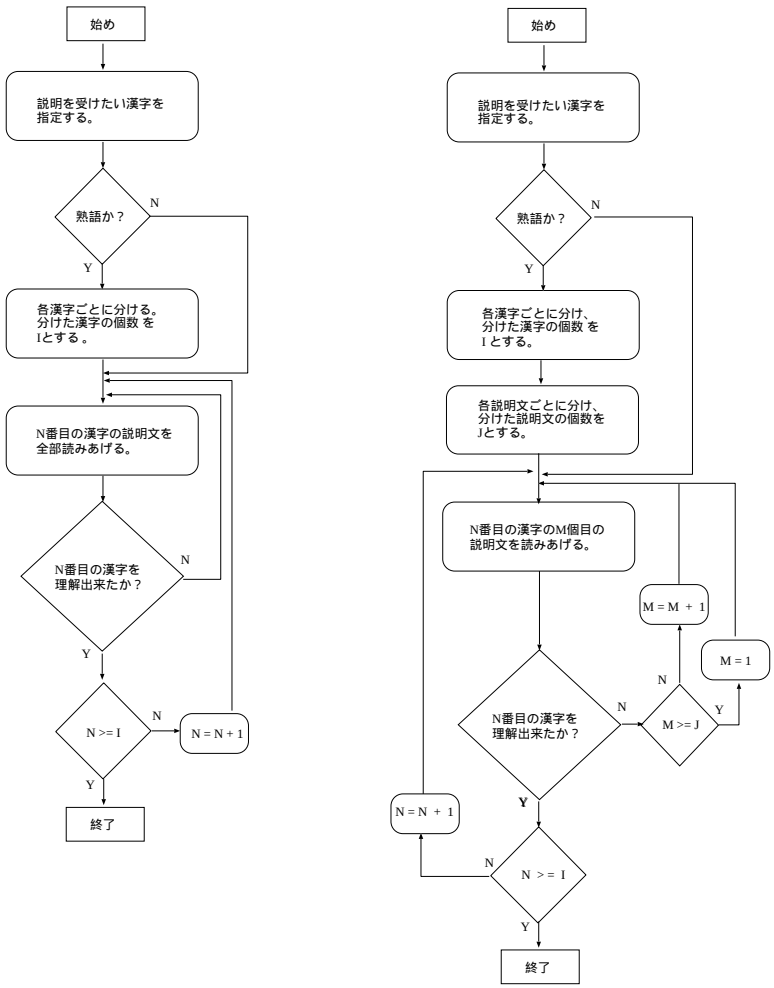


Fig. 2 説明ソフトのユーザーインターフェース

6.1.1 対象の単漢字

説明可能な単漢字を調べるにあたり、その対象として JIS 第一水準の常用漢字 2965 個と JIS 第二水準に収録されている単漢字 3388 個を対象とした。

JIS 第一水準

JIS 第一水準とは、 JIS (日本工業規格) が定めたもので、情報交換用符号系で規定された文字セットのうち、常用漢字を中心に単漢字 2965 字、ひらがな 83 字、カタカナ 86 字、数字 10 字、英字 52 字、特殊記号 147 字、ギリシア 48 字、ロシア文字 66 字、罫線文字 32 字の計 3489 字からなる文字コードであり、漢字は 50 音順で並んでいる。

JIS 第二水準

JIS 第二水準とは、JIS (日本工業規格) が定めたもので、情報交換用符号系で規定された文字セットのうち、多少の部首字形を含めた計 3388 個の単漢字からなる文字コード系である。 JIS 第一水準とは違い、漢字は部首別で並んでいる。

6.2 判別方法

この、 JIS 第一水準に含まれる常用漢字 2965 個と JIS 第二水準に含まれる単漢字 3388 個をそれぞれ判別していく。判別方法は csh スクリプトによるプログラムを使用し、判別を行った。

csh を使った判別方法

判別の基準として、対象の単漢字が辞書ファイルの中に動詞・形容詞として収録されていれば説明可能とし、動詞・形容詞が一つもなければ説明不可能とする。

また、動詞・形容詞として抜き出しても対象の単漢字が先頭にないものは除外する。例えば、『下』を辞書ファイルから動詞・形容詞を抜き出すと

おりる 下りる #KS 7
おろす 下ろす #S5 3
かきおろす 書き下ろす #S5r 0
かきくだす 書き下す #S5 1
きりさげる 切り下げる #KSr 1
・
・
・

というように抜き出されていく。ここで、『書き下ろす』、『書き下す』、『切り下げる』のような動詞・形容詞は、『下』という単漢字が含まれているが補助的な意味合いで使われていて、『書き下ろす』で『下』を連想することはできないと考えられる。よって、このような先頭にない動詞・形容詞は除外することにする。しかし、『書き下ろす』や『書き下す』等は、先頭の単漢字である『書』の説明には使用できるものとする。

6.3　健常者に対してのアンケート

単漢字を動詞・形容詞に変えて説明する方法だけで、どの程度理解が得られるのかアンケートを取った。アンケートの内容は以下の通りである。

説明ソフトの手順

説明ソフトの手順として、一度に説明する方法と、一つずつ説明する方法の二種類を提案した。そこで、アンケート対象者にそれぞれの手順について具体例をあげ、どちらの方の説明ソフトの手順が良いか答えてもらった。

同音異義語の判別

動詞・形容詞で同音異義語を判別できるかアンケートをとった。今回使用した『こうしょう』は同音異義語が多く、広辞苑 [7] には 49 個の熟語が記載されている。その熟語のうち、動詞・形容詞で説明できる 10 個の熟語を選び、その説明をする。そしてその説明した熟語だと思うものを 49 個の中から選んでもらうというアンケートをとった。表 3 は、広辞苑 [7] に記載されていた 49 個の『こうしょう』の熟語をまとめたものである。

表 3

黄鐘	交床	交渉	交鈔	公傷	公娼	公称
公証	公相	厚相	厚賞	口承	口証	口誦
好尚	工匠	工商	工廠	巧匠	巧笑	康尚
康正	校章	紅晶	洪昇	洪鐘	甲匠	綱掌
考証	行省	行賞	行粧	行障	後章	後証
講頌	鉱床	降将	高商	高唱	高尚	高昌
高笑	高声	高翔	高姓	鴻鐘	咬傷	哄笑

単漢字の判別

ある特定の単漢字の動詞・形容詞の説明文を聞いてもらい、それで連想した単漢字を書いてもらうというアンケートを行った。ここで、動詞・形容詞が多い、少ない単漢字では理解度に差が生じるかを調べるため、説明文が多い単漢字と少ない単漢字をそれぞれ 5 個ずつ計 10 個選んだ。それの説明をそれぞれ聞いてもらい、理解したらその漢字を書いてもらうというアンケートをとった。ここで、今回使用した単漢字を表 4 にまとめた。ちなみに 1 ~ 5 番目は説明文の多い単漢字で、 6 ~ 10 番目は説明文の少ない単漢字である。

表 4

単漢字	説明文の個数	説明文の例
引	45	ひきだす
割	12	わりふる
困	2	こまる
繰	12	くりかえす
見	83	みる
暗	1	くらい
叶	1	かなう
久	1	ひさしい
告	1	つげる
保	1	たもつ

7　判別結果とアンケート結果

7.1　判別結果

JIS 第一水準の常用漢字 2965 個のうち、 canna の辞書ファイルに動詞・形容詞として収録されている単漢字は 1138 個であった。つまり、 canna の辞書ファイルだけで動詞・形容詞として収録されていない説明不可能となる単漢字は 1829 個である。また、 JIS 第二水準の単漢字 3388 個のうち、動詞・形容詞として収録されている単漢字は 50 個しかなかった。それを以下の表にまとめた。

JIS 第一水準			
総数	説明可能数	説明不可能数	割合
2965 個	1138 個	1829 個	約 62%

JIS 第二水準			
総数	説明可能数	説明不可能数	割合
3388 個	50 個	3338 個	約 1.5%

また、JIS 第一水準と JIS 第二水準の説明可能な単漢字と不可能な単漢字の例を以下にあげる。

JIS 第一水準		JIS 第二水準	
判別可能例	判別不可能例	判別可能例	判別不可能例
哀	亜	儚	寃
引	烏	眩	帛
曲	伽	翔	斫
取	皆	躾	穿
追	査	貶	驅
痛	野	咎	佳

説明可能・不可能の単漢字の例

この説明方法では JIS 第一水準に含まれる常用漢字 2865 個のうち、約 62% は説明できないという結果となった。 JIS 第二水準にいたっては、約 1.5% しか説明できないという結果になっている。漢字はもともと音読みが多く、動詞・形容詞等の訓読みが少ないので音読みしかない単漢字が説明できなかったと推測する。

このままでは過半数の単漢字は説明できないままなので、何らかの改善策、例えば他の辞書ファイルを併用すれる等の改善や改良を加えなければこの説明方法だけで説明ソフトを作成することは不可能といえる。

7.2 アンケート結果

先程の内容のアンケートを当大学の大学生七人を対象に行った。

まず、説明ソフトの手順については、アンケート対象者が単漢字を理解するとき、一つずつ説明文を聞いた方が自分のペースで考えられると思い、一つずつ説明する方法が良いと考えていた。しかし、結果は対象者七人全員が一度に説明した方法が良いと答えた。理由としては、説明文が複数あるならば一度に聞いた方が対象の単漢字が連想しやすいということだった。

以下の表はそれぞれのアンケート結果をまとめたものである。表 5 は同音異義語の判別によるアンケートの結果であり、次に表 6 は単漢字の判別に関するアンケートの結果をまとめた。アンケート対象者が正解したならば「○ 」で表わし、不正解ならば「× 」で表わした。そして、最後に正答者と誤答者の比率を表わした。

表 5

熟語	一人目	二人目	三人目	四人目	五人目	六人目	七人目	正答者	誤答者
行省	○	○	×	×	×	○	○	4	3
考証	○	×	×	×	×	○	×	2	5
高商	○	○	○	○	×	○	○	6	1
厚相	×	○	○	×	×	×	×	2	5
行障	×	○	×	×	×	×	○	2	5
高翔	×	×	×	×	×	×	×	0	7
高承	×	○	×	×	×	○	×	2	5
高唱	○	○	×	×	○	○	×	4	3
口証	○	×	×	×	○	×	○	3	4
高笑	○	○	×	○	○	○	○	6	1

表 6

熟語	一人目	二人目	三人目	四人目	五人目	六人目	七人目	正答者	誤答者
引	×	○	○	○	○	○	○	6	1
割	○	○	○	○	○	○	○	7	0
困	○	○	○	○	○	○	×	6	1
繰	○	○	○	○	○	○	×	6	1
見	○	○	○	○	○	○	○	7	0
暗	○	○	○	○	○	○	×	6	1
叶	○	×	×	×	×	○	○	3	4
久	×	○	○	×	○	○	○	5	2
告	○	○	×	×	×	×	○	3	4
保	○	○	○	○	○	○	×	6	1

それぞれのアンケートの結果はこのようになった。同音異義語の判別では、半分以上の対象者が漢字を理解できなかったという結果になり、特に『高翔』は正答者はいなく、『厚相』、『行障』、『高承』も悪かった。

『高翔』は、『翔』の単漢字の説明文は「かける」しかなく、この説明だけで『翔』の単漢字を連想できなかったようである。また、『厚相』、『行障』、『高承』はそれぞれ『厚』は『暑』、『障』は『触』、『承』は『受』の単漢字と間違いていることが多かった。共に説明文が一つか二つしかない単漢字である。説明に出てきた熟語の先頭の単漢字に『行』と

『高』が多くみられたため、一番目の単漢字は連想できたが、熟語の二番目の単漢字が連想できなかったようである。

単漢字の判別では、説明文が多い単漢字は正解率はかなり高いという結果になり、逆に説明文が少ない単漢字、特に『叶』と『告』の正解率が良くなかった。これは、『叶』は「かなう」という説明文だけで、この単漢字で「かなう」とは普段使わない動詞・形容詞なので連想しづらいという意見があった。また、『告』も「つげる」しか説明文がなく、『次げる』、『継げる』、『注げる』等の他の単漢字を連想し、結局どれの単漢字なのか絞れず、回答しなかった人もいたようである。

また、アンケートをとった人々にそれぞれ感想を聞いてみた結果、以下のような意見があがった。

同音異義語の判別

- 分かりやすい単漢字もあったが、説明文が一つしかない単漢字は分かりにくい。
- 訓読みとして普段使わない単漢字は動詞・形容詞で説明されても分かりくい。逆に普段動詞・形容詞として使っている単漢字は分かりやすい。
- 同じ単漢字、『高』等が頻繁に出てきたので、同じ単漢字の説明文を何回も聞いているとその単漢字の一つ目の説明文で理解できるようになった。
- 動詞・形容詞だと連想しづらい。

単漢字の判別

- 説明文を読み上げただけではその単漢字を思い出せないときがあった。
- 説明文が一つだけで、それが同じ動詞・形容詞の単漢字が複数あるようなもがあるので、その場合はどれなのか理解できない。
- 一つ目である程度の漢字は連想できたが、確信を持つため何回か説明を聞いたが、一つしか説明文がないものは理解できなかった。

7.3 判別可能な単漢字とアンケートの考察

今回 JIS 第一、第二水準の単漢字がどの程度説明できるか判別を行い、動詞・形容詞で説明する方法でどの程度理解できるかアンケートを行った。それについてそれぞれ考察する。

判別可能な単漢字について

今回は、 canna の辞書ファイルの中で対象の単漢字が使われている動詞・形容詞が一つでもあれば説明可能ということにした。その結果、約 38% の単漢字しか説明できないという結論になった。つまり、約 68% 、過半数以上の単漢字は説明できないということになる。この時点で、このままこの説明方法だけでソフトを作成することは不可能といえる。

解決方法としては、複数の辞書ファイルを使用し、その中から抜き出しせば多少は改善させるのではないかと考えている。しかし、辞書ファイルの中には品詞コードがないものがある。もし、品詞コードがあったとしても、 canna の品詞コードとは異なる可能性もある。品詞コードを加える又は品詞コードを改良するという大々的な辞書ファイルの変更作業はできない。

また、説明可能という単漢字でも説明文が一つしかなく、それが同じ動詞・形容詞の読み方だった場合の問題もある。例えば『准』や『殉』は説明文を「じゅんずる」しか持たない。この場合『准』を説明したくても『殉』なのかユーザーは判断できない。これらの問題をどのように解決していくのかが今後の課題といえる。

同音異義語の判別

同音異義語を判別するアンケートは、全員の正解率が悪かったといえる。これは、説明文を聞いてもその単漢字が出てこないことが原因であると推測する。普段、動詞・形容詞で使用している単漢字等はすぐに想像できるが、普段動詞・形容詞として使わない、または説明文が少ないという単漢字は理解されにくいという結果になった。

二字熟語くらいならば、例えどちらかの単漢字が説明不可能でも、熟語をある程度ならば理解できるのではないと考えていた。しかし、アンケートの熟語を見てもらえば分かるが、一文字目が同じ、又は二文字目が同じという熟語はかなりある。つまり、一文字目だけや二文字目だけ説明を聞いても対象の単漢字にたどり着かないのではないと思われる。

単漢字の判別

単漢字に関するアンケートは、説明文の多い、少ない漢字でそれぞれアンケートをとった。予想では、説明文の少ない漢字は理解度が低いと予想していた。説明文が少ない単漢字は数種類に絞られるかもしれないが、説明文が一つや二つだけではどの単漢字なのかは特定しずらいと考えたからである。

実際、アンケート結果は、やはり説明文の多い単漢字の正解率は良かったが、少ない単漢字の正解率は悪かった。特に『叶』と『告』の正解率が悪い。理由としては、それぞれ

の単漢字の説明文が「かなう」と「つげる」であり、それらは普段よく使わない動詞・形容詞である。そのため、「かなう」と説明しても『叶う』と連想できなかったと考えられる。実際、アンケート対象者からそのような意見が聞かれた。

8　まとめ

本研究では、単漢字や熟語にそれぞれ固有の説明がある辞書ファイルを用いずに、フリーなものを用いて視覚障害者の人達に上手に説明できる方法はないかというものであった。今回は単漢字を動詞・形容詞に変えて説明する方法の基本的なシミュレーションやアンケートを行っただけなので、基礎段階の途中であり、今後もさらなる研究が必要だろうと思われる。

今回、単漢字を動詞・形容詞に変えて説明する方法について調べてきた。結果としてはJIS 第一水準の単漢字 2965 個のうち 1138 個、つまり約 38% しか説明できないという結果が得られた。 JIS 第二水準にいたっては 50 個、約 1.5% しか説明できないという結果になっている。結論としては、この説明方法だけで単漢字説明ソフトを作成しようとすることは難しく、もしもこの説明方法で説明ソフトを作成するならば、何か別の説明方法と組み合わせるか、または別の説明方法の補助的な役割で使われることになると思う。

この説明方法は、『危険』を「あぶない、けわしい」のようにそれぞれ動詞・形容詞に変えると、動詞・形容詞がその熟語の意味的になるような場合があるので理解が得られやすいと考えていた。しかし実際には、多くの単漢字は動詞・形容詞に変えることができず、熟語の意味的にとらえるものは少なかった。今後の課題としては、今回は canna の辞書ファイルだけで判別を行ったのでもし、他の辞書ファイルと併用した場合はどの程度改善されるかを調べていきたい。しかし、辞書ファイルの中には品詞コードのないものもあり、そのような場合の解決策も考えなければいけない。

今回は説明方法の判別とアンケートで終了した。結果として、この説明方法は単体で説明ソフトに使用するには不十分だという結論になった。それならば、以前の研究で提案された井上氏 [1)] や山口氏 [2)] の説明方法と組み合わせたとき、それら単体の説明よりは理解度が高まるのではないかと思い、井上氏 [1)] や山口氏 [2)] の説明方法単体よりこの説明方法を組み合わせた説明の方が理解度が高くなるのか調べるなどの課題が出てくる。今後はそのような組み合わせの判別が行えたら良いと思う。

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” , 新潟工科大学情報電子工学科卒業論文 (2001)

[2] 山口 圭: “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” , 新潟工科大学情報電子工学科卒業論文 (2005)

[3] Paul DuBois: “入門 csh & tcsh”, 鈴鹿倫之 福澤 裕 訳, オライリー・ジャパン (2002)

[4] 深沢 千尋: “すぐわかる Perl” , 技術評論社 (2004)

[5] IDEA・C: “UNIX コマンド ポケットリファレンス ビギナー編”, 技術評論社 (2001)

[6] 舟本 奨: “実用 UINX ハンドブック”, ナツメ社 (1997)

[7] 新村 出: “広辞苑”, 岩波書店 (1998)