

視覚障害者用漢字詳細読みの単語親密度及び構成要素の分析

渡辺 哲也^{*1} 山口 俊光^{*1} 渡辺 文治^{*2} 岡田 伸一^{*3}Analysis of Word Familiarity and Components of Conventional *Shosaiyomi* with the Change of Difficulty of Kanji CharactersTetsuya Watanabe,^{*1} Toshimitsu Yamaguchi,^{*1} Bunji Watanabe,^{*2} and Shinichi Okada^{*3}

Abstract - We analyzed the word familiarity and components of *shosaiyomi* of screen reader products. The word familiarity and components were compared among three groups of Kanji difficulty; educational Kanji group, common use Kanji group, and JIS 1 level Kanji group. It was found that as the degree of difficulty increases, (1) word familiarity lowers, (2) the number of Kanji characters which can be explained simply with the pronunciation and paired Kanji characters decreases, and accordingly (3) the number of Kanji characters which are explained with elements of Kanji or the names of famous persons or places increases.

Keywords : Blind Persons, Screen Reader, Kanji Characters, Vocabulary, Familiarity

1. はじめに

視覚障害者に音声のみで漢字を判別させるための表現形式である「詳細読み」について、その理解度に影響を及ぼす要因の解明と改善を進めている。これまでの一連の研究で、理解度に影響する要因として大きいのは、詳細読みで使われる説明語が利用者の語彙範疇にあるかどうか、どれくらい馴染みがあるか(単語親密度が高いか)、教育漢字ならば配当学年で習う読みかどうか、といった点であることを明らかにした^{[1][3]}。ただしこの知見の適用は、教育漢字の詳細読みを児童に聞かせた場合に限られる。

教育漢字以外の常用漢字や、常用漢字以外の JIS 第 1 水準漢字の詳細読みでは、語彙・親密度の問題のほかに、難易度の高い漢字に特有な問題 - 当該漢字を含む熟語の少なさや、異体字の存在など - も予測される。このことを統計的に検討するため、スクリーンリーダー 5 製品の JIS 第 1 水準漢字の詳細読みを対象に、詳細読みで説明に使われている単語の親密度と、詳細読みの構成要素という二つの観点から分析した。分析結果から難易度の高い漢字特有の表現の特性と課題について考え、より分かりやすい詳細読み策定のために留意すべき点を考察した。

2. 詳細読みの分析

2.1 対象

5 種の Windows 用スクリーンリーダー (95Reader 6.0, PC-Talker XP, VDMW300-PC-Talker, JAWS 4.5, JAWS 6.2)

について、JIS 第 1 水準の漢字 2,965 字の詳細読みを書き起こし、これを分析対象とした。JAWS のみ 2 種類のバージョンを取り上げたのは、両バージョンで全く異なる詳細読みが採用されているためである。なお、視覚障害者にも利用しやすいとされる携帯電話 (FOMA らくらくホン III (F882iES)) に搭載されている詳細読みは JAWS 6.2 のものと同じという情報を得ている。

2.2 分析方法

まず、詳細読みに用いられている単語を抽出した。詳細読みの構成要素のうち訓読みは単独で意味が通るため単語として扱った。他方で音読みは単独で単語となる場合が少ないため、単語として抽出しなかった。性質による説明の中に現れる単語 (例えば、「～を意味する」)、及び字形による説明の中に現れる単語 (偏や傍の名称、位置を表す語「上」「下」など) も抽出しないこととした。「意味する」や「上」「下」などは主たる説明語ではないこと、偏や傍の多くは辞書の見出し語となっていないことが理由である。

抽出した単語は NTT のデータベース^[4]にあたって親密度の数値を求めた。このデータベースで調べられるのは『新明解国語辞典第 4 版』(三省堂)の見出し語に限られる。単語親密度には、単語を音声で聞いた場合の音声単語親密度、文字で見た場合の文字単語親密度、文字と音声の両方を提示された文字音声単語親密度の 3 種類がある。音声で提示される詳細読みに適用するには音声単語親密度をみるのが妥当だが、日本語で 3~4 割にも上る同音異義語を音声単語親密度では区別できないため、今回は文字音声単語親密度を親密度データとして用いる。単語親密度は、1 から 7 の間の数値で示され、数値が高いほど親密度が高い。

次に、詳細読みをその構成要素によって分類した。構成要素には、音読み、訓読み、各読みを含む熟語、その熟語を含む用例 (複数の文節からなるもの)、人名・地名、

*1: 国立特別支援教育総合研究所

*2: 神奈川県総合リハビリテーションセンター
七沢ライトホーム

*3: 障害者職業総合センター

*1: National Institute of Special Needs Education

*2: Nanasawa Lighthouse, Kanagawa Rehabilitation Center

*3: National Institute of Vocational Rehabilitation

性質（字義）による説明，漢字の字形の説明，漢字の種類（旧字/新字，異体字など）などがある。表1に各要素の例を示す。出現頻度が10程度と少ない要素は省略した。詳細読みは，これらの要素が単独あるいは組み合わせによって構成される。

JIS 第1水準漢字2,965字を，教育漢字1,006字，常用漢字から教育漢字を除いた939字，JIS 第1水準漢字から常用漢字を除いた1,020字の3群に分け，それぞれを教育漢字群，常用漢字群，JIS 第1水準漢字群と本稿では呼び，この順序で漢字の難易度が上がると見なす（図1）。

以上のように調べた単語親密度と構成要素の割合について，スクリーンリーダ製品ごとに漢字の難易度による変化を見る。

3. 分析結果

3.1 単語親密度分布の変化

図2に，単語親密度分布の漢字難易度による変化をスクリーンリーダごとに示した。図2は箱ひげ図になっており，ひげの上端と下端は最大値と最小値，箱の上端と下端は第3及び第1四分位数，箱の中央の横棒は中央値を表す。いずれのスクリーンリーダにおいても，教育漢字群の詳細読みの単語親密度の中央値は6前後，第1四分位数も5.6前後と高い値であった。常用漢字群になると単語親密度は全体的に下がるが，それでも中央値が5.5前後，第1四分位数は5前後を保っている。単語親密度5以上の単語は成人の95%くらいの人を知っているとされるので^[4]，常用漢字までの詳細読みの説明語の約75%には，ほとんどの成人が知っていると言われているといえる。

JIS 第1水準漢字群では単語親密度は大幅に下がり，スクリーンリーダ4製品で中央値が5前後，第1四分位数が3.7~3.9となった。1製品はこれより低く，中央値が4.7，第1四分位数が3.4であった。単語親密度4の単語を知っている成人の割合は50%程度なので，説明語を知っている人の割合が半分以下の単語が詳細読みの4分の1を占める計算となる。

最小値に注目すると，教育漢字群，及び常用漢字群の中にも単語親密度が1.5から2となるものが含まれていた。単語親密度2はほぼ誰も知らない単語であり，数は少なくともこのような単語が含まれていたことが，特に児童にとっては詳細読みからもの漢字を想起する際の問題となっていたことが推察される^[1]。

3.2 構成要素の割合の変化

詳細読みを構成する要素の割合が漢字難易度によって変化する様子を図3に示す。構成要素の割合にはスクリーンリーダごとの特徴が見られる。教育漢字群を見ると，95Readerでは音読みと訓読みが50%以上の詳細読みで使われ，また音読みを含む熟語が40%を占める（図3(a),(b),(c)）。PC-Talkerでは音読みを含む熟語が90%以上，

表1 詳細読みの構成要素の分類と例。括弧内は説明される漢字。

Table 1 Examples of components of <i>shosaiyomi</i> .	
構成要素	例（括弧内は説明対象の漢字）
音読み	ガ（牙）
訓読み	ハコ（箱）
音読みを含む熟語	エイキュウノエイ（永）
訓読みを含む熟語	イドミズノイ（井）
音読みを含む用例の用例	ヨインヲノコスノイン（韻）
訓読みを含む熟語の用例	イヌガホエルノホエ（吠）
人名・地名	シガナオヤノヤ（哉）
性質による説明	ヒャクブンノイチノリン（厘）
漢字の字形	カワヘンニバケルノクツ（靴）
漢字の種類	カルノイタイジノカル（疋）

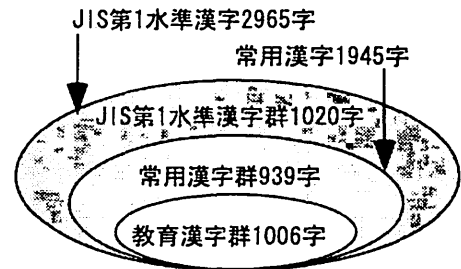


図1 3種の規格による漢字の分類

Fig. 1 Classification of Kanji characters by three standards.

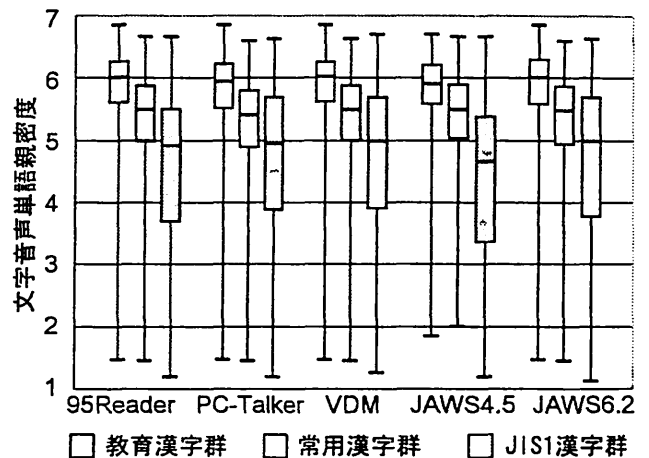


図2 単語親密度の漢字難易度による変化 箱ひげ図のひげの上端と下端は最大値と最小値，箱の上端と下端は第3及び第1四分位数，箱の中央の横棒は中央値を表す。各スクリーンリーダの三つのグラフのうち，左側白い箱は教育漢字群，中央薄いグレーの箱は常用漢字群，右側濃いグレーの箱はJIS 第1水準漢字群を表す。

Fig. 2 Change in word familiarity with the change in difficulty of Kanji characters.

訓読みが70%であり，両者の組み合わせが主に利用されていることがわかる（同(b),(c)）。VDMでは音読み，訓読み，音読みを含む熟語がある程度の割合ずつ使われている（同(a),(b),(c)）。JAWS 4.5は音読みを含む熟語が90%以上を占める（同(c)）。JAWS 6.2では，音読みと訓読みが多く（60~70%），音読みを含む熟語も30%弱含まれる（同(a),(b),(c)）。教育漢字群については各要素の提示順序も含めて文献[1]で詳細に記述しているので，参照されたい。

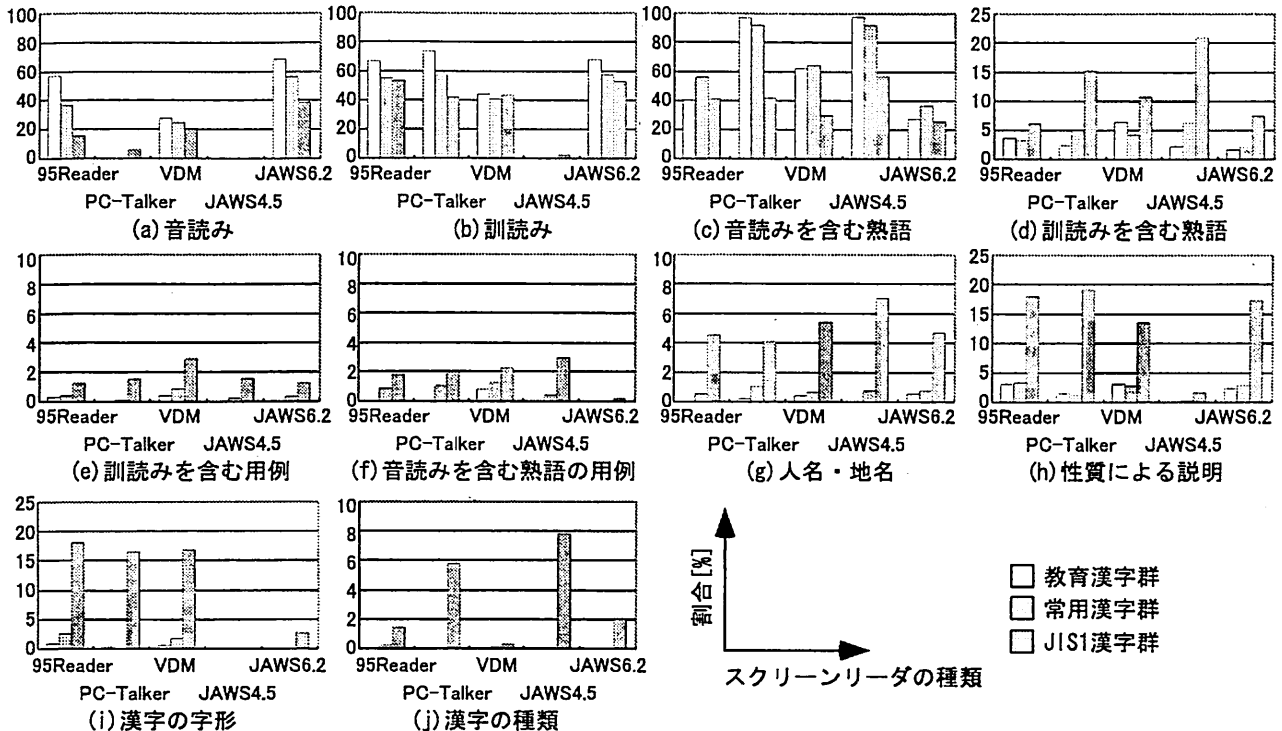


図3 詳細読み構成要素の割合の漢字難易度による変化。割合は、その要素の出現数を各漢字群の総数で割り100を掛けた数値。各スクリーンリーダーの三つのグラフのうち、左側白い棒は教育漢字群、中央薄いグレーの棒は常用漢字群、右側濃いグレーの棒はJIS第1水準漢字群を表す。

Fig. 3 Change in the proportion of shosaiyomi components with the change in difficulty of Kanji characters.

常用漢字群になると、教育漢字群で主流であった要素の割合が10~20%程度ずつ減った。その分、逆に増えた要素は、音読みを含む熟語の用例(図3(f))、人名・地名(同(g))などである。

JIS第1水準漢字群では構成要素の割合の変化が大きかった。全スクリーンリーダー製品において、音読み、訓読み、音読みを含む熟語といった教育漢字群で主だった要素は10%から最大では50%も減った(図3(a),(b),(c))。その代わりに、訓読みを含む熟語(同(d))、人名・地名(同(g))、性質による説明(同(h))、漢字の字形(同(i))、漢字の種類(同(j))、訓読みを含む用例(同(e))、音読みを含む熟語の用例(同(f))の割合が増えた。性質や字形などは、スクリーンリーダー製品によっては15~20%まで割合が増えた。使われる構成要素の種類が増えたこと自体も特徴的である。漢字の難易度が上がるにつれ、一定の方式だけでは説明しきれなくなったものと思われる。

4. 考察

4.1 単語親密度の観点から

説明に用いる単語の親密度を理論的にどの程度まで上げられるかを三つの漢字群ごとに検討してみた。その手順は以下の通りである。

- 1) 各漢字を含む単語を辞書(『新明解国語辞典第4版』,三省堂)の見出し語より抽出する。
- 2) 同じ辞典の見出し語間で同音異義語のある単語を除外する。

表2 各漢字を含んだ単語の理論的最高親密度の分布
Table 2 Distribution of theoretical highest word familiarity of each Kanji character.

	第3 4		第1 4		最小値
	最大値	分位	中央値	分位	
教育漢字群	6.844	6.312	6.125	5.906	2.188
常用漢字群	6.940	5.875	5.531	5.125	1.469
JIS1 漢字群	6.940	5.156	4.281	2.938	1.156

- 3) 2)で残った単語の文字音声単語親密度^[4]を求める。
- 4) 各漢字ごとに最も高い単語親密度の値をその漢字の理論的最高親密度とする。
- 5) 4)を三つの漢字群ごとに集計し、最大値、第3四分位、中央値、第1四分位、最小値を求める。

計算結果を表2に示す。なお、単語親密度を求めることのできた漢字の割合は、教育漢字群は100% (1,060/1,060)、常用漢字群は99.7% (936/939)、JIS第1水準漢字群は87.9% (897/1,020)であった。

表2の値をスクリーンリーダー製品の詳細読みの単語親密度(図2)と比較すると、教育漢字群の最も低い単語親密度は、4種の製品の約1.5から約2.2まで引き上げられる可能性が示された。教育漢字群と常用漢字群の中央値は、製品群の値と理論値の間に大差はない。JIS第1水準漢字群では逆に、理論値の4.281は製品群の5前後よりも低い。これは、製品群の詳細読みでは、親密度が低いと思われた単語を避けて、字形や性質などの構成要素による説明方式に切り替えたためと思われる。詳細読みの構成の変化に関しては次節でも考察する。

4.2 詳細読みの構成要素の観点から

漢字は、字形、読み、意味の三つの要素を有する。視覚障害者に漢字を判別させるには、視覚情報である字形を除いて、読みと意味、そして意味から派生した熟語や用例を使うのが適切である。既存スクリーンリーダーの詳細読みを分析したところ、常用漢字までは読みと熟語・用例で大部分の漢字の説明ができていた。しかし、常用漢字以外の JIS 第 1 水準漢字となると、字形や漢字の種類で説明することが避けられない場合があった。それらには以下の 2 種類がある。

- 異体字があるもの。異体字同士は読みと意味が同じなため、字形の違いでしか説明できない。例：「鈎」(カギバリノカギ)と「鈎」「カネヘンニハイクノク」(PC-Talker)。
- 常用漢字外のため、常用漢字で書き換えられることが多く、読み・意味・用例とも、対応する常用漢字と一致するもの。例：「稀」(常用漢字外)と「希」(常用漢字)。「廓」(常用漢字外)と「郭」(常用漢字)。その漢字の条件から字形で説明するのやむを得ないとしたが、漢字を見た経験のない視覚障害者にとって字形による説明は一般的には理解しづらいので^[5]、その使用は最低限にとどめるべきであろう。ただし、学習により視覚障害者の漢字の知識は変化するとともにされており、今後長期的な研究が必要である。

JIS 第 1 水準漢字でもう一つ特徴的なのは、人名・地名を使った説明表現の増加である。人名・地名以外に用いない漢字では、これらによる説明もやむを得ない。しかし、どの範囲の人名・地名まで使ってよいかという基準が必要であろう。例えば地名であれば県名と県庁所在地までに限るのか、市区町村でもよいか、人名ならどの資料に載っているものまでか、などの基準である。

これまで明らかにしてきたとおり、詳細読みを聞いて漢字を判別できるかどうかは利用者の知識に大きく依存する^[6]。常用漢字であれば、平均的な成人の語彙を前提に詳細読みを策定できるが、常用漢字以外の JIS 第 1 水準漢字となると、比較的高い知識水準の利用者を想定しなければ、たとえどのような説明表現をしても判別は難しいものと思われる。

5. まとめ

JIS 第 1 水準漢字の既存詳細読みについて、説明単語の親密度と、詳細読みの構成要素という二つの観点から分析した。その結果、常用漢字以外の JIS 第 1 水準漢字において単語親密度、構成要素の割合ともに大きな変化が観察された。この分析結果は、既存のものより理解しやすい詳細読みの策定基準の考案に役立てたい。

なお、実用的な目的のためには、JIS 第 2 水準漢字、及び非漢字についても、理解しやすい説明表現に関する研究が必要である。

謝辞

詳細読みの分析には、千葉大学の学部生/大学院生(当時)の福田慧人氏、大須賀智子氏、矢島直人氏、神谷佐武郎氏、及び当研究所で非常勤勤務している村上真理氏に御協力頂いた。本研究は、科学研究費補助金(基盤研究(B)課題番号:16300191)により実施した。

参考文献

[1] 渡辺哲也, 渡辺文治, 藤沼輝好, 大杉成喜, 澤田真弓, 鎌田一雄: “スクリーンリーダーの詳細読みの理解に影響する要因の検討—構成の分類と児童を対象とした漢字想起実験—,” 電子情報通信学会論文誌 D-I, Vol.J88-D-I, No.4, pp.891-899 (2005).

[2] 渡辺哲也, 大杉成喜, 澤田真弓, 山口俊光, 渡辺文治, 岡田伸一: “スクリーンリーダーの漢字詳細読みに関する研究—児童を対象とした言葉の親密度調査—,” 電子情報通信学会技術研究報告, WIT2005-04 (2005).

[3] 渡辺哲也, 渡辺文治, 岡田伸一, 山口俊光, 大杉成喜, 澤田真弓: “スクリーンリーダーの漢字詳細読みに関する研究—試作した詳細読みによる漢字書取り調査—,” 電子情報通信学会技術研究報告, WIT2005-47 (2005).

[4] 天野成昭, 近藤公久(編著): “NTT データベースシリーズ 日本語の語彙特性 第 1 期 CD-ROM 版 単語親密度,” 三省堂, 東京 (2003).

[5] 海保博之, 佐々木正人: “先天盲の漢字存在感と漢字検索過程,” 特殊教育学研究, Vol.21, No.4, pp.7-16 (1984).

[6] 渡辺哲也, 渡辺文治, 山口俊光, 大杉成喜, 澤田真弓, 岡田伸一: “スクリーンリーダーの詳細読みによる漢字想起実験—成人を対象とした場合—,” 電子情報通信学会論文誌 D, Vol.J89-D, No.3, pp.602-605 (2006).

(2006 年 11 月 1 日受付, 2007 年 3 月 9 日再受付)

著者紹介

渡辺 哲也 (正会員)



1993 年北海道大学大学院工学研究科修了。同年水産庁水産工学研究所研究員, 1994 年障害者職業総合センター研究員, 2001 年国立特殊教育総合研究所研究員, 2005 年主任研究官, 2007 年国立特別支援教育総合研究所主任研究員。この間 2004 年 3 月~8 月米国ウィスコンシン大学工学部客員研究員。音声・触覚情報を用いた視覚障害補償技術の研究開発に従事。電子情報通信学会, 日本音響学会, ヒューマンインタフェース学会, 日本特殊教育学会, 視覚障害リハビリテーション協会等各会員。博士(工学)。

山口 俊光 (非会員)



2004 年神奈川工科大学大学院工学研究科情報工学専攻修了。同年 4 月同大工学部福祉システム工学科研究生。2004 年 11 月より国立特殊教育総合研究所科学研究支援員。視覚障害者の情報補償に関わる研究開発に従事。情報処理学会, 視覚障害リハビリテーション協会各会員。

渡辺 文治 (非会員)



1978年東北大学大学院教育学研究科博士課程前期終了。同年神奈川県総合リハビリテーションセンター七沢ライトホーム職員，現在，副技幹。視覚障害者の福祉・リハビリテーション・教育・レクリエーション等についての調査・研究に従事。日本特殊教育学会，日本ロービジョン学会，日本レジャー・レクリエーション学会等会員。

岡田 伸一 (非会員)



1974年慶應義塾大学大学院経済学研究科修了。国立職業リハビリテーションセンター研究員を経て，現在，障害者職業総合センター主任研究員。日本ロービジョン学会評議員。視覚障害者用感覚代行機器の開発に従事。日本職業リハビリテーション学会，日本特殊教育学会等会員。