

JIS 外漢字の電子化手法と教材開発への応用

野崎 浩成 ・ 横山 詔一
(情報科学選修) (国立国語研究所)

The Method of Making the Electric Media for Developing the Teaching Materials

Hironari NOZAKI* and Shoichi YOKOYAMA**

(Department of Information and Computer Science*, National Language Research Institute**)

【あらまし】 小中・高等学校で使用されている教科書には、古文や漢籍、地名や人名などを表記するために、JIS の規格表には掲載されていない漢字、いわゆる JIS 外漢字が数多く使用されている。しかし、JIS 外漢字をコンピュータ上で取り扱う技術は十分に確立されていないため、それらの漢字を含むテキストを電子化し、教材として利用するのは困難であった。そこで、本研究では、大規模文字フォントデータベースを活用して、JIS外漢字を電子化する手法を考案し、テキスト教材を電子メディア化する際の問題点とその解決法について検討を行った。

1. はじめに

我が国の小中学校および高等学校で使用されている文部省検定済みの教科書には、国語科にある古文や漢籍、さらには社会科で使用される地名や人名などの固有名詞を表記するために、JIS の規格表には掲載されていない漢字、いわゆる「JIS 外漢字」が数多く使用されている。しかしながら、JIS 外漢字を含むテキストを電子化する技術は十分に整備されていない。そのため、漢文のテキストなど JIS 外漢字を含む教材を教師がコンピュータで作成するのは困難であった。

そこで、本研究では、大規模文字フォントデータベース『今昔文字鏡』(文字鏡研究会 1999)⁽¹⁾を用いて、教師が持つパソコン上で、JIS 外漢字を電子化する手法を考案する。さらに、実際の教科書に掲載されている JIS 外漢字を電子化する実例を紹介した上で、テキスト教材を電子メディア化する際に生じる諸問題とその解決技法について実証的な検討を行った。さらに、その応用として、インターネットを介して、JIS 外漢字を誤りなく情報交換する方法を示した。

2. 従来の手法とその問題点

これまでに、JIS 外漢字をコンピュータ上で取り扱う方法は、いくつか提案されてきた。その最も身近な例として、ワープロの外字機能がある。これは、JIS 外漢字をユーザが自ら登録する方法であるが、これには次のような問題点が指摘されている。① JIS 外漢字の

文字フォントを自ら作成する必要があること、②文字フォントの作成は煩雑であり、かつ、多大な時間を要すること、③外字の印字品質が悪いこと、④自ら作成した文字フォントは、他者が所有するコンピュータでは利用できないため、教師同士での情報交換ができないこと、すなわち、インターネットなどのオープンなシステムでは利用できないこと、などがあげられる。

さらに、JIS 外漢字の電子化を試みた例として、『広辞苑第 5 版 CD-ROM 版』(岩波書店 1997)⁽²⁾がある。この辞書ソフトでは、専用の文字フォントを搭載することで、JIS 外漢字の電子化を実現している。しかし、これらの JIS 外漢字フォントは、辞書ソフト専用に開発されたものであるため、我々が作成する文書ファイル上で利用することができない。すなわち、この辞書で電子化されている JIS 外漢字は、他のソフトウェアとの互換性は保証されていない。また、JIS 外漢字の文字フォントの輪郭線が滑らかではないため、文字の印字品質を向上させる必要があるなど、様々な問題点が指摘されている。

3. JIS 外漢字の電子化手法

上述したように、JIS 外漢字を電子化する従来の手法には、数多くの問題点が存在する。そこで、本研究では、それらの問題点を解決するために、以下のような手法を提言する。最初に、ワープロやデータベース、表計算ソフトなど、一般的なアプリケーションソフト上でも扱える汎用性の高い文字フォントを活用することによって、JIS 外漢字を電子化できるようにする。さらに、トゥルータイプフォント (True Type Font, 以下 TTF と表記) を利用する。これにより、文字の拡大・縮小が自由にできること、字体を拡大しても輪郭線が極めて滑らかであること、高品質の表示および印字が実現できること、などのメリットが得られる。これら一連の手法を用いることにより、JIS 外漢字を含むテキスト教材の電子メディア化が実現できれば、教材開発の支援環境が整備されるものと期待できる。

4. 教材開発への応用

図1には、中学2年時に使用する国語科の教科書に掲載されている『古典に親しむ』より、杜甫が書いた『絶句』の白文を示す(栗原ほか1997)⁽²⁾。図1のように、JIS外漢字『逾』を含むテキスト教材を電子メディア化する手法を具体的に検討する。今回、Windows環境下で、MS-Word 97を用いたが、他のアプリケーションを用いても全く同様の手順で実現できる。

何	今	山	江	
日	春	青	碧	絶句
是	看	花	鳥	
帰	又	欲	逾	杜甫
年	過	然	白	

図1. JIS外漢字を含むテキスト教材
中学2年国語科の教科書の抜粋

4. 1 大規模文字フォントデータベースの利用

「文字鏡研究会」が提供する大規模文字フォントデータベース「今昔文字鏡」には、約9万字(そのうち漢字は約8万字)の文字フォントが納められており、世界最大規模の文字数を誇る。このデータベースに納められている文字フォントをインストールする。なお、この文字フォントはLinux等と同様にフリーで利用できる。インターネットを利用すれば誰でも自由にダウンロードでき、TTF対応のWindows版とMac版、さらに、PSプリンター用、TeX用などが用意されている。ただし、文字フォントの検索エンジンの利用については、製品版を購入する必要がある(文字鏡研究会1999)⁽¹⁾。

4. 2 「今昔文字鏡」を用いたJIS外漢字の電子化

JIS外漢字をワープロ上で扱うための操作手順は、次の通りである。①電子化したいJIS外漢字(この場合『逾』)を今昔文字鏡から検索する。②『逾』を選択し、クリップボードに転送する。③ワープロ上では、『逾』を表示したい箇所にカーソルを移動させ、[貼り付け]を実行する。

なお、本稿は、これまでに述べた方法と全く同じ手順で、JIS外漢字『逾』を含むカメラレディ原稿を作成した。

5. インターネット上でJIS外漢字を扱う方法

JIS外漢字をインターネット上で流通させる方法として、現在のところ2つのアプローチがある。1つは漢字の画像(GIFファイル)を利用する方法で、もう1つはTTFを使う方法である(横山・野崎1999)⁽⁴⁾。

現在、この2つのアプローチのいずれにおいても、

学術研究団体である「文字鏡研究会」のフォントセンターに収録されている世界最大の漢字フォントデータベースを利用する方法以外に現実的な解決策は見当たらない(<http://www.mojikyo.gr.jp/>)。

「文字鏡フォントセンター」にリンクを張れば、8万字以上の漢字GIFファイルをWebページで表示できる。また、それらのTTFもダウンロードできる(文字鏡研究会, 1999)。

5. 1 漢字GIFファイルへのリンクを活用したホームページの作成

筆者らが運営している「文字メディア研究会」のホームページでは、漢字GIFファイルへのリンクを活用して、JIS外漢字を表示している(<http://gauss.c.u-tokyo.ac.jp/~ohmori/kanji.html>)。文字メディア研究会とは、日本語における文字・語彙の使用実態調査を実施し、文字言語の認知機構に関する研究を行うことを目的として設立されたもので、国内外の心理学者、国語学者や言語学者、企業のソフトウェア開発者、大学院生などが参加する産学協同の国際的な学術研究グループである。

付録には、文字メディア研究会のホームページに掲載されている『ゲタ文字「■」化された漢字の頻度表』のソースファイルの一部を示した(<http://gauss.c.u-tokyo.ac.jp/~ohmori/kanji/mtable3.html>)。その中から、「鄧」というJIS外漢字を96dotの画像で表示する際のHTMLファイルの一部を示す(付録の該当箇所には、アンダーラインを付加)。

```
<IMG NAME="mojikyo-font96-039630"
SRC="http://www.mojikyo.gr.jp/gif96/039/039630.gif" WIDTH=48>
```

文字鏡フォントセンター(<http://www.mojikyo.gr.jp/>)へのリンクにより、「鄧」の漢字GIFファイルを表示している。この方法を用いれば、JIS外漢字が表示できること、さらには、日本語が表示できない海外のコンピュータに対しても、文字情報の交換が誤りなく実現できること、などのメリットがある。

6. ま と め

本研究では、JIS外漢字を電子化する際に生じる諸問題を指摘し、その解決方法について検討した。さらに、中学校国語科の教科書に掲載されている漢文を具体例として取り上げ、大規模文字フォントデータベースを活用して、JIS外漢字を電子化する手法を考察した。この手法を用いれば、JIS外漢字を含むすべてのテキスト教材を電子メディア化し、教育資源として有効に活用できるものと考えられる。さらに、文字鏡フォントセンターを利用して、インターネット上で、JIS外漢字を表示させる手法を示した。

本研究での成果を利用すれば、古書や仏典などJIS

外漢字を数多く含む文献のデータベース化、図書館の電子メディア化、WWW を利用した電子化出版の実現など、幅広い分野での応用が期待できる。

参 考 文 献

本文中直接引用した文献

- (1) 文字鏡研究会編 (1999) パソコン悠悠漢字術—今昔文字鏡徹底活用—, 紀伊國屋書店, 東京
- (2) 栗原一登ほか (1997) 国語2 (文部省検定済み教科書), 光村図書出版(株), 東京
- (3) 岩波書店 (1997) 広辞苑第5版 CD-ROM 版, 東京
- (4) 横山詔一, 野崎浩成 (1999) インターネット上で JIS 外漢字を処理する技法と教科書電子化への応用, 日本教育工学会第15回全国大会講演論文集 (課題研究)

日本語における漢字使用に関する分析

野崎浩成, 横山詔一, 磯本征雄, 米田純子 (1996) 文字使用に関する計量的研究—日本語教育支援の観点から—, 日本教育工学雑誌, Vol. 20, No. 3, 141-149

*1997年度日本教育工学会学会論文賞受賞論文

野崎浩成, 横山詔一, 近松暢子 (1997) 新聞と雑誌における漢字使用頻度の分析, 日本教育工学会誌, Vol. 21, Suppl. 21-24

野崎浩成, 横山詔一 (1998) テキストデータベースに対応した KWIC 作成システムの開発, 日本教育工学会誌, 22(Suppl), 57-60

Nobuko Chikamatsu, Shoichi Yokoyama, Hironari Nozaki, Eric Long, Hiroyuki Sasahara, Sachio Fukuda, Development of a Japanese Kanji Character Frequency List, Twelfth International Unicode/ISO 10646 Conference, PART 1, 1-10, Tokyo, Japan (1998. 4)

JIS 外字を含めた漢字使用実態調査

横山詔一, 笹原宏之, 野崎浩成, エリク・ロング編著 (1998) 新聞電子メディアの漢字—朝日新聞 CD-ROM による漢字頻度表—, 国立国語研究所プロジェクト選書No.1, 三省堂, 東京
*JIS 外漢字を含めた文字使用に関する分析。JIS 外漢字を分析の対象とした我が国で最初の大規模調査。

笹原宏之, 横山詔一, 野崎浩成, 米田純子 (1998) 「朝日新聞」の CD-ROM と紙面における幽霊文字と辞書非掲載字—「JIS X 0208」の漢字を中心に—, 計量国語学, Vol. 21, No. 4, 145-161

横山詔一, 笹原宏之, エリク・ロング, 野崎浩成 (1999) 新聞記事データベースにおける「槇」の消失現象, No.20, 57-63, 人文学と情報処理, 勉誠出版

漢字教育への応用

野崎浩成, 市川伸一 (1997) 漢字学習支援システムの開発—漢字の構造理解と筋運動感覚の獲得—, 日本教育工学雑誌, Vol. 21, No. 1, 25-35

Stephen Wood Ryner, Jr., Nobuko Chikamatsu, Hironari Nozaki, Shoichi Yokoyama, Sachio Fukuda, Java Kanji Flashcard 500: Kanji, Java, and the World Wide Web, Twelfth International Unicode/ISO 10646 Conference, PART 2, 1-7, Tokyo, Japan (1998. 4)

附 記

文字鏡研究会の古家時雄氏 (株)エーアイ・ネット代表) には, 今

昔文字鏡の利用許諾をいただき, 感謝致します。

(平成11年9月10日受理)

付録:

『ゲタ文字「≡」化された漢字の頻度表』の HTML ファイルの一部を示す。これは, JIS 外漢字を示すために, 漢字 GIF ファイルへのリンクを利用したものである。

以下は, 文字メディア研究会のホームページ (<http://gauss.c.u-tokyo.ac.jp/~ohmori/kanji/mta-ble3.html>) から一部分を抜粋し, 本論文用に加筆したものである。

```
<HTML> <HEAD> <TITLE> Table3 </TITLE> </HEAD>
<BODY BGCOLOR="#FFFFFF">
<A HREF="maki.html"> 本文に戻る </A> <HR> <P>

<MALTICOL COLS="2">
<TABLE BORDER WIDTH=300>
<CAPTION ALIGN="top"> 表3 ゲタ文字「≡」化された漢字の頻度表 </CAPTION>
<TR ALIGN=center> <TD> 漢字 <TD> 頻度 <TD> 順位 <TD> 属性
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-039630">
SRC = "http://www.mojikyo.gr.jp/gif96/039/039630.gif"
WIDTH=48>
<TD> 426 <TD> 1 <TD> IBM & 補助
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-054374">
SRC = "http://www.mojikyo.gr.jp/gif96/054/054374.gif"
WIDTH=48>
<TD> 230 <TD> 2 <TD> 83JIS
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-054082">
SRC = "http://www.mojikyo.gr.jp/gif96/054/054082.gif"
WIDTH=48>
<TD> 72 <TD> 3 <TD> 83JIS
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-054408">
SRC = "http://www.mojikyo.gr.jp/gif96/054/054408.gif"
WIDTH=48>
<TD> 31 <TD> 4 <TD> 83JIS
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-077620">
SRC = "http://www.mojikyo.gr.jp/gif96/077/077620.gif"
WIDTH=48>
<TD> 25 <TD> 5 <TD> IBM
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-054613">
SRC = "http://www.mojikyo.gr.jp/gif96/054/054613.gif"
WIDTH=48>
<TD> 13 <TD> 6 <TD> 83JIS
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-010995">
SRC = "http://www.mojikyo.gr.jp/gif96/010/010995.gif"
```

```

WIDTH=48>
<TD> 6 <TD> 7 <TD> IBM &amp; 補助
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-
019174"
SRC ="http://www.mojikyo.gr.jp/gif96/019/019174.gif"
WIDTH=48>
<TD> 4 <TD> 8 <TD> IBM &amp; 補助
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-
018319"
SRC ="http://www.mojikyo.gr.jp/gif96/018/018319.gif"
WIDTH=48>
<TD> 4 <TD> <TD> IBM &amp; 補助
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-
054902"
SRC ="http://www.mojikyo.gr.jp/gif96/054/054902.gif"
WIDTH=48>
<TD> 3 <TD> 10 <TD> IBM
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-
050014"
SRC ="http://www.mojikyo.gr.jp/gif96/050/050014.gif"
WIDTH=48>
<TD> 2 <TD> 11 <TD> IBM &amp; 補助
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-
050040"
SRC ="http://www.mojikyo.gr.jp/gif96/050/050040.gif"
WIDTH=48>
<TD> 2 <TD> <TD> IBM &amp; 補助

```

```

<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-
017695"
SRC ="http://www.mojikyo.gr.jp/gif96/017/017695.gif"
WIDTH=48>
<TD> 2 <TD> <TD> IBM
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-
021062"
SRC ="http://www.mojikyo.gr.jp/gif96/021/021062.gif"
WIDTH=48>
<TD> 2 <TD> <TD> IBM &amp; 補助
<TR ALIGN=center> <TD> <IMG NAME="mojikyo-font96-
021717"
SRC ="http://www.mojikyo.gr.jp/gif96/021/021717.gif"
WIDTH=48>
<TD> 2 <TD> <TD> IBM
</TABLE> <P> <BR>
<font size=3 color="#FF3300"> 注1) </font>
「IBM」はIBM特殊漢字, 「IBM*」はIBM特殊漢字で包括でき
るもの,
「補助」は補助漢字 <BR> (JIS X0212-1990)を示す。
<BR>
<P> <HR> <A HREF="maki.html"> 本文に戻る </A> <P>
(C)Copyright 1998-1999 横山詔一・笹原宏之・エリク=ロング・
野崎浩成
</BODY>
</HTML>

```