

研究報告『説文解字』データベース ソフトについて⁽¹⁾

高橋 由利子

一. 『説文解字』データベース化の準備段階

1.1. 説文会と『説文解字』

お茶の水女子大学で中国文学を学んだ卒業生の有志から成る説文会⁽²⁾は、当時同大学の教官であった頼惟勤先生の指導の下に、段玉裁の『説文解字段注』の読書会を行ってきた。

それは、漢字研究のための基礎文献である『説文解字』（以下『説文』と略称）について、より深く学ぶためであったが、同時に、漢字の形と音について、より有効な研究を行うために『説文』収録漢字についての番号付けを行った。

その理由は「説文約一万字を縦横無尽に組み換え、分類考察するには、これを番号化して機械に入れることが能率的であろう」⁽³⁾と考えたためである。

その成果として「加番」（番号を加える）シリーズとも言える以下の3冊を刊行した。

1. 『加番説文解字音均表』（1985）⁽⁴⁾
2. 『加番説文解字』（1991）⁽⁵⁾
3. 『江沅説文解字音均表改正』（1994）⁽⁶⁾

当時はすでにコンピュータがあり、それにデータを入れ処理を行うには番号をつけることが最適と考えられ、これがその後すべての作業の出発点となった。

その後のIT技術と環境の進歩により、現在は色々な方法でのデータベース化が可能であるが、我々の作業は、すべてこの番号を基礎とするものである。

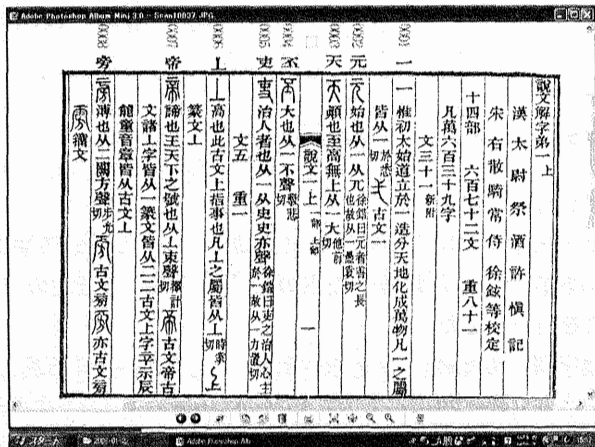
1.2. 『説文解字』データベース化の開始

頼惟勤先生が1987年に定年退官後、説文会は上記2冊の刊行と補正に加え、後任の藤山和子教官を中心として、段玉裁の『説文解字段注』についても加番作業を続け、また同時に『説文』についての多くのデータを蓄積した。

幸いにも1995年に文部科学省研究費の交付を得て、データベースソフトの

原型を作り、データ入力の本格作業を始めることができた。これは1.1.に挙げた『加番説文解字』と『江沅説文解字音均表攷正』およびその後に行った加番『説文解字段注』のデータを一つのソフトに組み込んだものであった。(7)

これに色々な改変とデータを加えたものが、今回報告するソフトである。(8)



2. の『加番説文解字』

二.『説文解字』データベースソフトのデータについて

2.1.何がデータとして入っているか

前章で述べたように、基本となるデータは、『説文解字』一篆一行本・『説文解字音均表』・『説文解字段注』から取られ、それぞれの番号でリンクする。以下に、一字の漢字ごとに、どの本のどんなデータが入っているかを示す。

一行本：親字（小篆）：コード番号、楷書化画像。

説明（説解）：字音と字形の部分（六書パターン）。意味は含まず。

注釈（徐鉉）：字音に関するもの：反切。

追加（説文会）：反切の所属（声母、韻母）。（広韻・嚴学寤⁽⁹⁾）

段注本：親字：コード番号。

注釈（段玉裁）：古音に関するもの。（反切、古音十七部）

音均表：親字：コード番号。段玉裁に基づく古音十七部。

このソフトは『説文解字』の全文を載せたものではなく、音に関する情報を字ごとにまとめたものである。上記の下線を引いたものが主なデータである。

詳細は2.2. 以下で実際の画面に基づき、述べることとする。

2.2.漢字情報画面

↓ 1

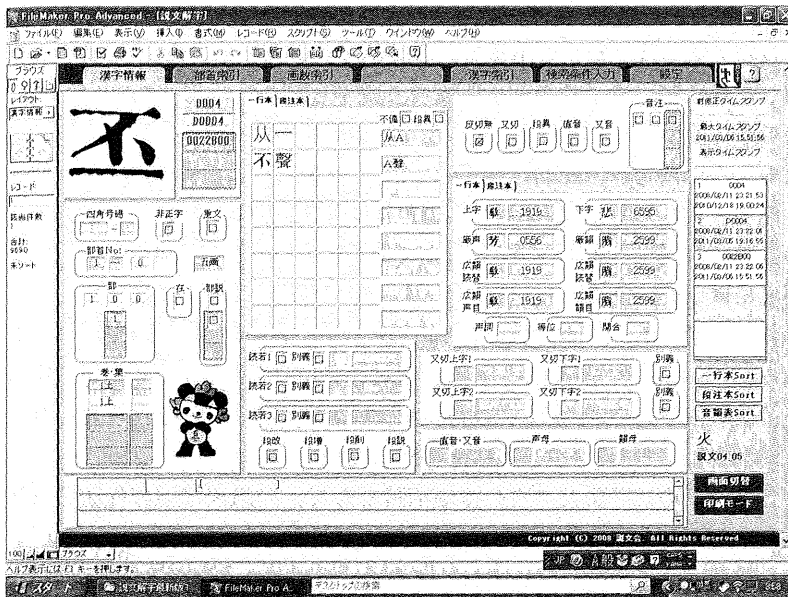
↓ 2

↓ 3

↓ 4

↓ 5

↓ 6



以下、矢印で示された箇所を画面の下に向けて見てゆくこととする。

2.3.親字

↓ 1 : この下には漢字のフォントがある。説文解字の親字には小篆が使われているが、それを楷書化した中華書局の『説文解字』（一篆一行本）の巻末索引の字をイメージスキャナで取り込み、その字をフォント作成ソフトでフォント化し、番号とリンクさせた。(10)

番号だけでなく、部首や画数からも親字を呼び出せるよう、巻末の部首索引、総画索引をソフトに組み込んだ。

2.4.文字番号（コード番号）

↓ 2 : 漢字の横の3つの番号が説文会の加番した番号である。

上から順に、一行本・段注本・音均表の同じ「丕」という漢字の番号で、それぞれ0004, D0004, 0022B00である。

字の配列は一行本と段注本は部首別であり、最初の方の字の番号は、配

列に違いがないので一致しているが、後の方では、段玉裁が字を増減したり、字の配列を変えたりして、違う番号となる。そのため先頭にDをつけて区別し、一行本番号を0004、段注本番号をD0004とした。

音均表は、江沅が、段玉裁の『説文解字』の漢字の古音を17の部に分類した説に基づき、その部に所属する漢字を部ごとに第1部から第17部まで順に並び変えた書物である。部の中の漢字の配列は漢字の諧声符順である。

最初の4ケタの数字が同種類の諧声符に付けた番号、次のアルファベットは諧声符が同じ部分の字体を含んで発展していくものに付けた記号で、多く発展していけば、A・B・C・D・……と増えていく。

例えば「丕」字は「不」から始まる「不」系統の2番目の諧声符である。「不」系統の諧声符は0022、最初の諧声符である「不」という字の番号は0022A00となり、2番目の諧声符の「丕」は0022B00となり、3番目の諧声符の「否」は0022C00となる。

また「不」、「丕」、「否」は諧声符として、同諧声符字の先頭字(初文)であるので最後の2ケタの番号はいずれも00とした。

このように加番していくと、例えば「丕」という諧声符を持つ字の5番目の字である「駟」字は0022B04となる。

2.5.部首、総画、古音17部番号、巻数

漢字のフォントと3つの番号の下には以下のボックスがある：

一段目：四角号碼(未入力)のボックス、正字でない場合のチェックボックス(非正字)、異体字・別字(重文)がある場合のチェックボックス。

二段目：『説文解字』の部首番号(全540部の通し番号)、総画数。

三段目：横ボックス：段玉裁の古音17部の番号；複数部に所属する字があるため3つ作った。その右には段玉裁が部の前に「在」を付けたもののチェックボックス、また段玉裁が論説を展開している場合のチェックボックスを作った。

縦ボックス：音均表の部の番号、複数部所属字のため3つある。

ただし、現段階では、段玉裁の部には、音均表の部を加番番号から自動入力した。いずれは段注を調べて違う箇所があれば訂正する予定である。

四段目：巻数と葉数。上から順に、一行本・段注本(葉数は未入力)。

2.6. 文字の構成要素

- ↓ 3 : 説解の字体・字音に関わるデータ、(いわゆる六書)。この丕という字は、構成要素として「一」、諧声符に「不」を持つが、説文解字の記述通り「从一不聲」と入力。

2.7. 構成要素のパターン、讀若

- ↓ 4 : この六書の入力には「从A」、「A聲」などのパターン⁽⁴⁾を作り、まずパターンを選んでから、そのAに実際の漢字を一行本番号で入力するという方法を取った。当然のことながら、フォントは親字と同じである。

ここでは使用したパターンを表示し、ここからデータの修正ができるようにした。複数のパターンを持つ字があるため、8個のボックスを作った。

また、パターン表示の上のチェックボックスは説解の六書の叙述が不完全な場合や段玉裁が修正を加えていることを示すためのものである。

これらの下には讀若音に使われている漢字を一行本番号で入力するためのボックスを作ったが、まだ入力途中である。複数の讀若を持つ親字があるため、3つのボックスを作り、意味による使い分けがある場合の別義チェックボックスを作った。また、段玉裁が修正・増減・論説を加えていることの有無を示す、段改・段増・段削・段説チェックボックスを一番下に作った。

2.8. 反切データ

- ↓ 5 : 一行本 (大徐) の反切についてのボックス。

一段目：反切が無い場合、複数の反切がある場合、段玉裁が修正している場合、直音の場合、複数の直音の場合のチェックボックス。

二段目：反切の詳細で、上字・下字の漢字を一行本番号で入力。

説文解字に無い字は作字し、別番号を付けた。正字以外のは、作字し、もとの番号の前にCやEを付けて正字との関係を示した。例えば「於」はC2384となる。

また、すべての反切のボックスで、右側に一行本の番号を入れれば漢字のフォントに変換されて番号の左側に出てくる。

三段目：嚴学寤 (注9を参照) が分類した反切の所属。聲類と韻類。

四段目：それを『広韻』の分類に読み替えたもの。

五段目：『広韻』の分類による反切の所属。聲目は『広韻聲系』に拠る。この下の声調・等位・開合も『広韻聲系』に拠って入力予定。これから下の2段は、2つ以上の反切がある時にその反切を入力するボックスと、意味による違いがある場合の別義チェックボックス。最後の段は、反切が無く直音や又音がある場合の、その字と所属を示す。これらの反切のデータは、未入力のものもあり、検討中である。

2.9.その他

↓6：各自のログイン履歴が出る。データを変更した場合、最終変更の記録を残すためである。

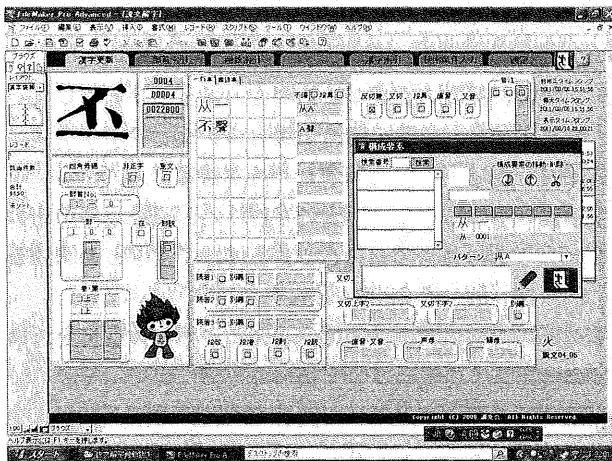
その下の各本のsortボックスはすべての本のデータ入力後のためのもの。その下はソフトのフォントファイルの場所で、丕字のフォントはJIS漢字の火字に貼り付けられていることを示す。⁽¹²⁾

この漢字情報画面の最下段には各自がコメントを書き入れることができる。

2.10.データの修正・更新

データの修正や更新は、各ボックスをクリックするとポップアップされる更新画面で行う。下に例として構成要素のデータ(从A、A聲)の更新画面を示す。

2.11.データ更新画面



三.『説文解字』データベースソフトの検索について

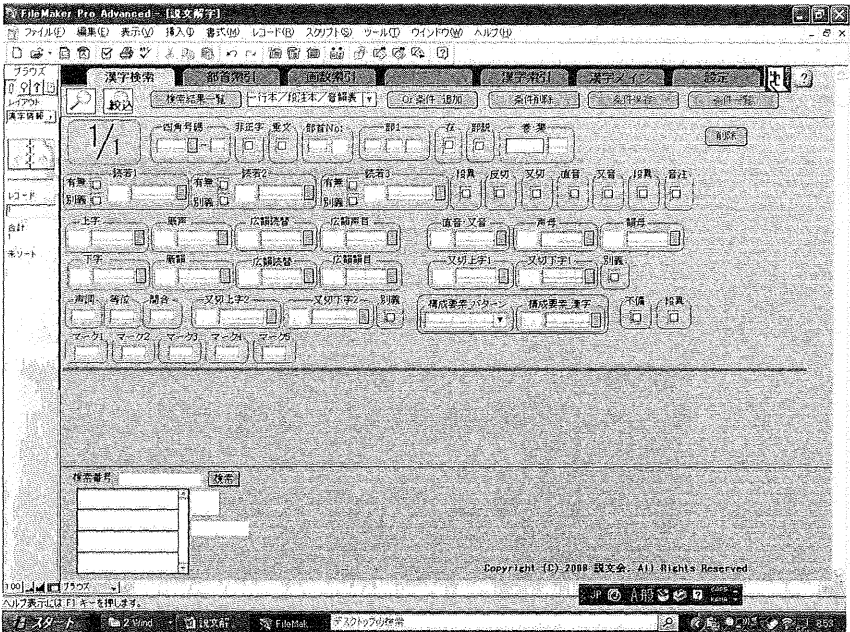
3.1. 何が検索できるか。

前章で、どのようなデータが入っているかについて述べた。これらのデータすべてについて、いろいろな条件を設定し、検索することができる。

一つのデータについて検索することもできるし、データ間の and 検索と or 検索が可能である。

3.2. 漢字検索画面

2.2の漢字情報画面の↓5と↓6の間に、検索条件入力というタブがある。ここをクリックすると、以下の漢字検索画面に切り替わりここから検索を行う。



空欄になっているボックスの検索したい部分にデータを入れて、左上虫眼鏡マークをクリックすると検索結果が出る。

データの入れ方は数字または文字番号で行う。チェックボックスはそこにチェックを入れて行う。

画面下のボックスは、文字番号によって漢字を検索するツールで、ここで漢字と番号を出しておいてから、検索したいボックスにカーソルをあわせ、もう

一度その番号をクリックすると、そのボックスにデータが入る。

and条件検索は、同じ画面の複数のボックスにデータを入れて検索する。

or条件検索は、一つの画面にデータを入れて、画面上なかほどの or 条件追加をクリックすると、次の画面が現れ、2/2となる。そこに検索データを追加していく。順次追加が終わったら最後に虫眼鏡マークをクリックして検索する。

以下に最も単純な検索を行い、例とする。

3.3. 「又切」を持つ親字の検索

『説文解字』一行本には各親字に大徐による反切が付いているが、その反切が2つ以上のものがある。「又AB切」と書かれているので、「又切」と言う。

ここでは、その「又切」のチェックボックスにチェックを入れ、一行本の字で「又切」を持つ字すべてを検索し、出していく。

3.4. 又切検索結果帳票画面

5 ↓ 6 ↓ 7 ↓

親字	反切	部	上字	下字	声
贛	0588B01	部	上字古	1392	从聲
	0588C01	部	上字古	1392	从聲
	0588D01	部	下字退	1099	
	0588E01	部	下字退	9798	
脩	0178B01	部	上字徒	E1054	从聲
	0178C01	部	上字徒	E1054	从聲
	0178D01	部	下字聊	7439	
	0178E01	部	下字聊	0425	
苗	0224A01	部	上字徒	E1054	从聲
	0224B01	部	上字徒	E1054	由聲
	0224C01	部	下字遷	1026	
	0224D01	部	下字遷	1026	
齋	1004A02	部	上字疾	4903	从聲
	1004B02	部	上字疾	4073	从聲
	1004C02	部	下字咨	0807	
	1004D02	部	下字咨	2599	
蔣	0682E01	部	上字子	9321	从聲
	0682F01	部	上字子	9321	从聲
	0682G01	部	下字麗	3206	
	0682H01	部	下字麗	3206	
菱	0027A01	部	上字古	1392	从聲
	0027B01	部	上字古	1392	从聲
	0027C01	部	下字夜	0902	
	0027D01	部	下字夜	0902	

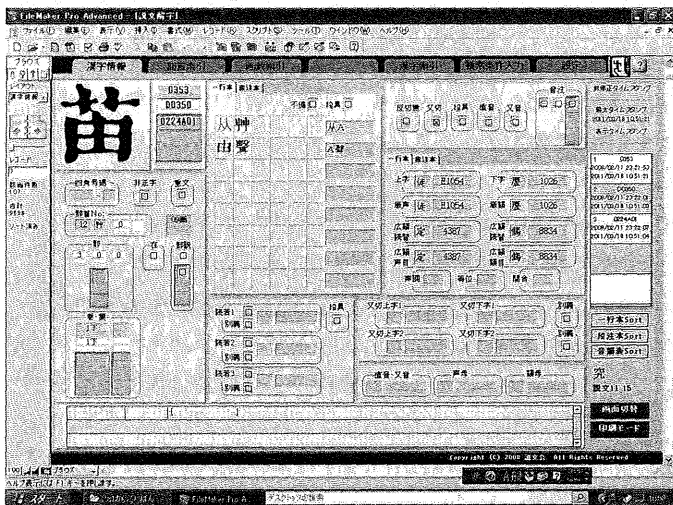
以下に、それぞれの矢印に従って詳細を述べる。

1 → : 「レイアウト」と書かれた下のタブに「検索結果」と出ている。これは

検索の結果の全てのデータを帳票（別にソフトの中に組み込まれている）
の形で一覧表として出すというモードであることを示す。

- 2→：本の絵の右側のページをクリックしたり、その下のインジケータを右にしたり、この画面では切れているが、右端の送る▼マークをスクロールダウンすると、次の検索結果帳票が順次出てくる。
- 3→：「レコード」と書かれた下のタブに「3」と出ている。これは左線でマークされているのが全データの3番目の「苗」であることを示す。
- 4→：「該当件数」「107」は説文解字の中に「又切」を持つ親字が107字あることを示す。なお、その下にある「合計」「9698」は、全部で9698の親字についてチェックしたということを示す。このソフトには前章2.8.で述べたように、反切を表示するために『説文』に無い字を作字し、それも親字として組み込まれているので、実際の『説文』収録字9431字よりも多い。
- ↓5：親字のフォント、その右に一行本番号と段注本番号、音均表番号（複数の部に属するので2つ番号がある）。
- ↓6：古音17部の所属、最初の字は8部と9部に所属している。その右に反切上字・下字のデータ。
- ↓7：構成要素。

3.5. 又切検索結果漢字情報画面



3.4. 又切検索結果帳票画面の1→で示した「検索結果」をプルダウンメニューで「漢字情報」に変えると、上記のような一文字ずつの漢字情報画面がでる。

同じく3.4. 又切検索結果帳票画面の2→で示した本の絵のページ、インジケータ等のスクロールダウン機能を使うと次の文字の漢字情報画面が出て、各文字のより詳しい情報を見ることができる。(13)

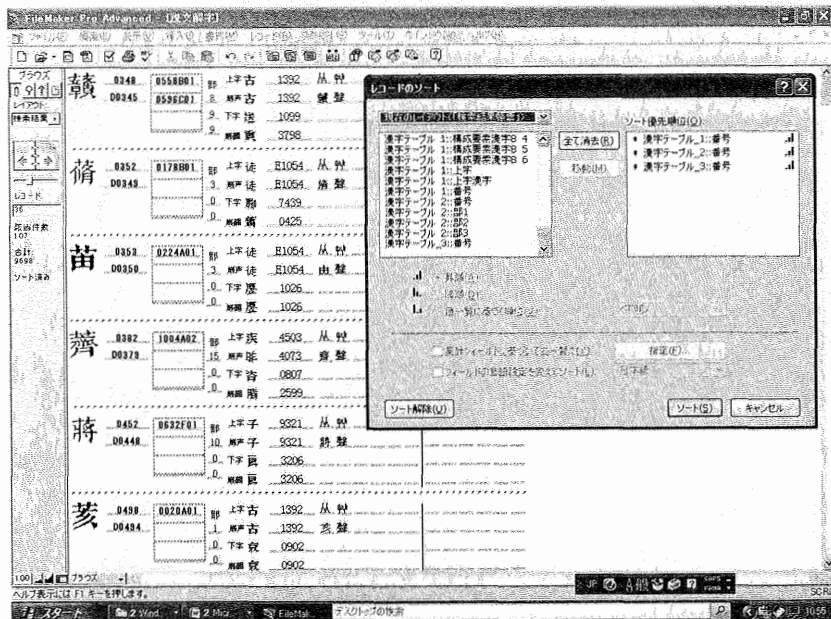
以上、ごく簡単なデータの検索を例にとりて、検索の概要を述べた。

この他に、FileMakerの「ソート機能」を利用して、検索結果帳票にリストアップされた文字の順番を検索したいデータの順に並べ替えることができる。以下に画面に従って詳細を示す。

3.6. レコードのソート画面

3.3. 又切検索結果帳票画面の6↓と7↓の間の上部タスクバーに「レコード(R)」という項目がある。ここをクリックし、レコードのソートを選ぶと、以下の画面が出る。ソート優先順位を「部」に変えると「部」の順に変わる。

その次の画面が該当字が部の順にソートされた帳票画面である。



漢字Hokori Pro-Advanced (国文辞書)			
ファイル(F)	編集(E)	表示(V)	辞書(D)
開	閉	表示	辞書
開	閉	表示	辞書
フリガナ	5312	0018405	部 上平岡 4337 从 欠
ローマ字	DS002		1 声 戸 許 1415 欠 聲
レイアウト			0 下平 戒 1035
辞書検索			0 声 戸 拜 57450
全	8331	0018009	部 上平 屹 7979 从 欠
上	DS025		1 声 戸 余 0657 有 聲
レコード			0 下平 容 4805
辞書検索			0 声 戸 考 9431
全	0499	0027601	部 上平 古 1352 从 艸
上	DS434		1 声 戸 古 1392 欠 聲
レコード			0 下平 倪 0502
辞書検索			0 声 戸 夜 0502
全	5262	0158A00	部 上平 奠 0670 从 見 冂
上	DS253		2 声 戸 奠 0670
レコード			0 下平 紅 8234
辞書検索			0 声 戸 紅 8234
全	5428	0111A18	部 上平 丁 0009 从 心
上	DS420		2 声 戸 胡 2571 欠 聲
レコード			0 下平 交 6318
辞書検索			0 声 戸 交 6318
全	8855	0100A18	部 上平 匹 8025 从 水
上	DS847		2 声 戸 匹 8025 異 聲
レコード			0 下平 辨 7017
辞書検索			0 声 戸 辨 0425

菱という字が6番目から3番目が変わっており、その他の字も入れ替わっている。これは、前の画面では一行本の番号順であったデータ一覧表が、17部の順にソートされているのである（ここでは1部から2部の字が示されている）。このように、ソート機能を使って、より有効で、見やすい検索結果を得ることもできる。

以上、このデータベースソフトの検索機能について述べた。もちろん、このような簡単なものだけでなく、もっと複雑な検索を各人の興味に応じてすることができる。また、今まで検証されてきた結果を再確認することもできるし、新しいものを発見することができるかもしれない。

例えば、構成要素のパターンと声母・韻母、反切上字・下字との関係、古音17部と声母・韻母、反切上字・下字との関係、讀若の原音と讀若音、古音17部との関係等々。(14)

このデータベースソフトは、まだまだ未完成で、不備な点も多いが、色々な使用方法を試しながら、これからも、より有効なものにしていくために協力し、努力していく所存である。

注

- (1) 本稿は以下の2つの口頭発表をまとめたものである。
 1. 『説文解字』データベースソフトについて：2008.1.26(麗澤大学言語研究センター第36回研究セミナー)
 2. 『説文解字』データベースソフト改訂版1について：2010.6.26(中国語学会2010年度大会)
- (2) 説文読書会ということもある。筆者がそれに参加したのはお茶の水女子大学大学院に進学した1971年からである。その後、多数の参加者が出入りし、データカードの作成や加番作業に加わった。現時点での参加者は、藤山和子・南谷葉子・高橋由利子・臼田真佐子・根岸政子・陶山知穂・塚越千史の7人。またシステムエンジニアの牧野英夫氏が賛助会員として参加、本会の活動を支えている。
- (3) 『江沅説文解字音均表攷正』(1994)の頼惟勤序文。
- (4) 実際は2つの部分からなる：第一部分は江沅の『説文解字音均表』の収録字を諧聲符順に分類加番したもの。第二部分は『説文解字』の一篆一行本の収録字を部首ごとにまとめて加番したもの。一字ごとではない。
- (5) 『説文解字』一篆一行本の収録字を一字ごとに加番したもの。注(4)で言う第二部分がその基礎である。中華書局影印本(1963,北京)が底本。P.28の画像参照。
- (6) 江沅の『説文解字音均表』の収録字に加番し、一字ごとに『加番説文解字』の番号を加えたもの。注(4)で言う第一部分がその基礎となっている。復興書局影印本(1972,台北)が底本。
- (7) 平成7年度～8年度科学研究費補助金(基盤研究< B > < 2 >)、総合的漢字研究のための『説文解字』のデータベース化とそれを利用しての基礎的研究(研究代表者：藤山和子お茶の水女子大学文教育学部教授、研究分担者：高橋由利子上智大学比較文化学部助教授(所属等はいずれも1995年当時)、研究成果報告書(全3冊)：『説文解字』データベース化のための検索システムコード番号表Ⅰ一篆一行本篇、同番号表Ⅱ段注本篇、同番号表Ⅲ段注本篇。
- (8) 最初の頃のデータの入力にはマイクロソフトのAccessを使用した。現在は画面のレイアウトや検索条件の設定が容易なFileMakerを使用している。
- (9) 厳学齋が「大徐本説文反切的音系」(国学季刊第6-1 民国25 pp45-143)で大徐の反切に使用されている最多字を分類項目にしたもの。
- (10) 段注本や音均表には段玉裁や江沅が『説文解字』の収録字を削ったり加えたりしている。加増した字については別にフォントを作りソフトに入れた。これらの字には一行本の番号は無く、段注本の番号や音均表の番号だけがある。また、その他の

反切などの情報は入っていない。削ったものについては一行本の番号はあるが、段注本・音均表の番号は無い。

- (11) その他、色々なパターンがあり、ソフトではプルダウンメニューで選べるようになっている。
- (12) 2.8. 反切データで述べたように、「右側に一行本の番号を入れれば、漢字のフォントに変換されて番号の左側に出てくる」のは、その番号からJIS漢字とフォントファイル名を取得し、そのJIS漢字をそのフォントで表示すると、目的とする『説文』の字が表示できるためである。
- (13) なお、「又切」の詳しいデータはまだ未入力のものが多いが、先にこのように該当字をすべて出してから、点検・入力することができ、より効率的である。また、データ入力の誤りも、同様にして発見・訂正できる。
- (14) 『江沅説文解字音均表攷正』(1994)の頼惟勤序文では、以下のようなことも述べられている：

「本書は江沅の諧声符番号と大徐の番号とを対照させている。そこで発生した問題は、①大徐と段氏との違い、②段氏自身の学説の変化（分類のゆらぎ）、江氏自身の分類のゆらぎ、以上の三者にそれぞれに由来するものがある。この点は将来会員によって考察が加えられて行くであろう。

また根本に遡って、江沅（本来は段玉裁）のように諧声符によって文字を古韻ごとに分類しきることができるかどうかということも、本書によって始めて実質的に検証されることとなるに違いない。

その際、①聲類（頭子音）の共通性による諧声符、②正字と重文との間で諧声符所属が違うとき、③假借の諧声、というような項目が前もって考えられる。」

(上智大学)