

新聞入力詳細説明書

(第1版 2002年5月1日作成)

2004年9月3日改定

1. 新聞記事テキスト化の基本となる考え方

- (1) 私どものこの事業においては、原則としていかなる記事選別の判断基準、方法をも排除することとします。従って、事業参加者個人にとって、目の前の紙面に存在する記事が実に些細であり、史料として不必要と感じられたとしても、自らの感想に抗し、漏れなく、全件、全文を忠実にテキストデータに変換していただく必要があります。

どのようにつまらぬ記事であっても、量としてまとまることで資料価値を発揮することもあります。またいずれ完成するこのデータベースから、遠い後世の史料探索者が、未来の（従って私どもには未知の）意図と価値観をもって、今は私どもが不要と思っている情報を、価値有る情報として取り出す可能性を妨げてはならないと考えます。

つまらぬ（と思う）記事を精確に時間をかけてデータに拾い上げる作業は、時として作業者にとって苦痛事となるかもしれません。

作業を始められるにあたって、自らが面白い、興味深いと思われる記事のみではなく、つまらないと思われる記事をも一律に精確に入力するのだと、ご確認下さい。

- (2) 全ての記事を網羅するといっても、ロゴ、写真、挿絵などはテキスト化することは不可能です。そこでこれらについては、紙面の中での有無区分と概要説明にのみとどめることとします。いずれかの時期に、もし静岡県立中央図書館等の公的機関による近代新聞画像データがWEB上で公開されるなどの状況となり、私どもが今後構築する新聞記事テキストデータベースとの間に有効な連携を図ることができれば、これらロゴ、写真、挿絵などを表現をなしえない私どもの欠点も、解消することになるものと考えます。
- (3) 完成するテキストデータの精度と客観性を確保するためには、不明あるいは明瞭でない文字の入力に際して、作業者個人のあいまいな推測を排除しなければなりません。入力の元となる紙面原稿上で、文字のつぶれや汚れなどで判読困難な文字がまま出現します。この場合、原則として伏字（>マークを入力。後述）の対処をし、あいまいなものはあいまいなままにしておくことを原則とします。作業者によって入力されたテキスト中の伏字部は、事務局にて再調査を行う事とします。
- (4) 原資料（紙面原稿）上に表記されている文字の中には、旧体字やJ I S第1水準、第2水準（以下「J I S内字」といいます）に無い文字が多数出現します。今回のテキスト化事業は、その完成品たる新聞記事テキストデータをインターネット上で広く公開し、NPOホームページ上で自由に記事を検索し表示していただくことを目的としています。しかるに一方、インターネット上ではJ I S内字以外は表示することは、少なくとも現在は不可能であります。このことから、私どもでは、例え

記事中に J I S 内字に無い文字が出現しても、これを J I S 外字でとすることなく、J I S 内字の中に同意、同義の簡易な字形が存在すれば、これに置きかえて入力する仕様を採用することとしました。このような対処は、原資料に対して忠実ではないと考えられる方もいらっしゃるかと思いますが、現時点でのインターネットの世界の一般的な技術水準を考え、かつ検索の利便を考えあわせた時、以上以外に方法は無いとの結論となりました。（置きかえるべき J I S 内字が存在しないケースについては後述）

(5) J I S 内字の中に、同意、同義の複数の字形（異体字）が存在する場合、上記（4）の検索の利便性の観点から、最も簡便な一つの文字に特定して入力を行う事としました。

(6) 紙面欄外の時刻表、天気予報、相場、株式市況などについては、ほとんどがマイクロフィルム収録時の切断によって複数紙面を接合しない限り完全な読み込みができないこと、また元来折れ目部あるいは読み手のめくり手がつく部分のため、破損汚れが激しいなど、著しく読み込みが困難であることをもって、残念ながら今回の収録を断念しました。このため、仮に欄外に読み取り可能な記事が有っても、原則として記事入力をしないものと致します。

2. 各項目入力の注意点

(1) 新聞名

▼ボタンを押し、民友新聞、静岡新報、重新静岡新聞、函右日報、静岡大務新聞の5紙名の名から選択して下さい。

(2) 発行年月日

発行年は西暦で入力します。新聞の上部欄外に発行日が印刷されていますが、和暦であるため、西暦は、入力フォーム内の[西暦-和暦対応表]のリンクをクリックし、換算表を表示して参照して下さい。

(3) 朝夕刊区分

明治期のものは朝夕刊の区別が無い時期のものが大半です。

第1面の新聞名の題字部分や、各紙面の上部欄外に「夕刊」とはっきり表示されていなければ、朝刊として入力を行います。

(4) 面数、段数

面数は、1発行日単位で、先頭が1面で順次追番となります。通常は紙面の上部欄外の右端、又は左端に（一）、（二）のように表示されていますが、面によっては

印刷されていなかったり、かすれで良く見えなかったりするケースがあります。また、これら原稿となっている紙面は、各地の旧家から多くの研究者の手によって集められたもので、従って1発行日の全ての面数が必ず収集されているとは限りません。1発行日の中が一部欠落していたり、一定期間まとめて欠落している場合もあります。

このようにさまざまな原因で入力しようとしている紙面の面数が不明であったり欠落があったりした場合は、まず当日1発行日の紙面原稿の下部に印字させている画像ファイル番号（例：F0010001.TIFF）が若い順になっていることを確認した上で、以下の対処を行います。

①面番号の不明な面の前後の面数がそれぞれ3、5と表示されていれば4とします。

②面番号の不明な面の前後の面数がそれぞれ3、6であれば、4とします。

③最終の紙面の面数が不明の場合は、前の紙面の面数が4であれば5とします。

同一発行日で、同一面数の紙面が複数枚出現した場合は、相互に記事内容が完全に一致しているか確認した上で、事務局に至急連絡を下さい。単純に事務局にて印刷をダブってしまったという原因のほか、早刷り版、遅刷り版あるいは地域版の相違で、紙面の一部のみが異なっているというケースであることも考えられます。

段数は、入力しようとしている記事の先頭部分が位置している段を入力します。

たとえば、その記事が複数段にまたがっていても、あくまで記事先頭がある段を入力します。また1つの記事が別枠となっていて通常の複数段を合併しているような場合（商品広告や登記広告などの場合に見うけられます）でも、同様にその先頭（枠の左上角）位置が属する段数を入力して下さい。

（5）内容区分

新聞紙面に掲載されているものを、その内容を読み込み以下の基準で判断して区分してください。

①一般記事

記者によって書かれた政治、経済、生活文化一般に関する報道記述をこの区分とします。いわゆる三面記事もこの区分に入ります。「社説、論説」との区別が困難な場合は、この区分に入れてください。明治期の紙面には、記事冒頭に「電報」と書かれたものが頻出しますが、これは東京にいる記者（特派員）が、中央政治や国際外交に関する記事を、速報として電報で送っていたもので、電話やファクシミリも

無く、東京／静岡間の鉄道の便も良くなかった時代が偲ばれます。（以上余談）

②社説、論説

新聞社の論説委員名あるいは新聞社の署名入りで時の政治や経済に対して意見を論述しているものです。明治大正期の紙面では、現在のように「社説」あるいは「論説」明記されていることは少なく、論述者が新聞社に属する個人又は新聞社であることがはっきりしている場合のみ「社説、論説」の区分とし、あいまいなものは「一般記事」として結構です。

③時事コラム

現代で言えば朝日新聞の「天声人語」のような、論説や一般記事とは別に、長期間継続の定まった標題と枠（コラム）で、時々々の世相、事件を記述者の雑感の形で掲載したもの。明治期のものは、コラムの形式（一定期間同一標題）でいわゆる三面記事の内の小事件を簡条書きで多く載せているなど、内容的には一般記事の延長とも言えるものであつたりしますが、これらも時事コラムとして扱ってください。

④小説

連載、単発を問わず掲載されているものが小説であればこの区分になります。小説の場合は、タイトルに小説名、サブタイトルに作者名を入力し、挿絵画家名が有る場合は小説作者名の後に（ ）書きで入れてください。

⑤短歌、俳句

投稿短歌、俳句又はプロの作者のものが掲載されている欄についてこの区分を使用します。漢詩、現代詩の定期掲載欄はまず出現しないと思いますが、もし出現すれば、この区分に入れてください。

⑥市況、相場（株式、商品・茶況など）

現在の新聞にも、商品の市況、相場、あるいは株式市況が毎日掲載されています。この種のものは全て「市況、相場（株式、商品・茶況など）」に区分します。市況や相場は、通常表形式で掲載されています。

⑦戦時公報

日清戦争、日露戦争、支那事変、太平洋戦争と、新聞記事のそこそこに、戦争に関する記事が出現します。戦争が行われている期間、新聞には戦死者及び戦傷者の名簿が地区単位で毎日のように掲載されます。この名簿を「戦時公報」と区分します。

⑧被災死者公報

戦争と同様に、明治～大正、昭和初期には、大きな天災事故が頻繁に起こっています。特に水害や地震、大規模な漁民事故などで多数の死者や怪我人が出ました。このような災害死傷者の名簿が掲載されている場合は、被災死者とします。

また太平洋戦争時の空襲や、大火災による死傷者名簿についても、「被災死者公報」として区分します。

⑨一般死亡広告

亡くなられた方の遺族や所属の会社、団体が、掲載料を新聞社に支払って死亡広告（公告）を掲載しているものです。現在の新聞にも同様の死亡広告（公告）が有ります。政治家、高級官僚、著名人が亡くなられた場合、掲載料無しに、新聞独自の判断で死亡記事を掲載する場合があります、このようなケースでは本来「一般記事」に区分されるべきです。ただし、一般の死亡広告と区別がつかない場合は、「一般死亡広告」に区分しても支障ありません。

⑩一般商業、商品広告

企業、商店が自社又は自社製品、自社販売品の宣伝、広告を行っているものです。企業が有償で掲載するものとしては、登記公告や、決算公告が存在しますが、これらは一般的な広告、宣伝とは異なりますので、下記⑪に区分されることになります。注意して下さい。

⑪その他の公告

⑦～⑩に該当しない公告、広告の類をこの区分とします。企業、団体が行う登記公告、登記変更公告、決算公告など、商法で義務付けられている公告のほか、国、県、市区町村官庁、裁判所などから出される公告（⑦、⑧を除く）などがこれに当たります。

記事冒頭のタイトルに「公告」、「〇〇について公告す」などと書かれている記事が出現することがありますが、これらも「その他の公告」に入れてください。これらは、人の意見開陳である場合もありますが、当時の政党人が個人名で掲載しているケースもあります。記事内容によっては、「一般記事」と区分する方が適切と思われる場合は、「一般記事」、「その他の公告」のいずれとしても可といたします。

⑫天気予報（説明省略）

⑬写真、絵（独立掲載）

記事や広告などに附属する写真や絵ではなく、そのみが独立して存在する写真、絵があればこの区分とします。一応区分を置いてありますが、このケースに該当するケースはまづあり得ないと思います。

(6) 記事タイトル (7) 記事サブタイトル

記事タイトルとサブタイトルは、記事の冒頭に、通常は記事本文より大きな文字であったり、太い文字であったり、あるいは行が別行になっていたりするもので、一つの記事全体の見出し文をいいます。

特に明治期の新聞は、江戸期の瓦版（かわらばん）の延長と言ってよいかと思いますが、現在の新聞のようにはっきりとした記事作成のルールが無く、従ってタイトル（見出し）の付け方も執筆者個々人次第といった状態であったようです。（印刷技術の問題もあったでしょう）このため、冒頭に述べたような特徴がはっきりしないタイトルや、時としてタイトルらしいタイトルが存在しないものも出現します。

以下はそのようなケースについてのタイトル、サブタイトル入力の方法に関する説明です。

①タイトルに該当するものが見当たらず、突然本文が始まっている場合は、タイトル欄には以下のように入力します。この項目は必須項目です。

タイトル欄： (タイトル無し)

②サブタイトルに該当する記載が無い場合は、サブタイトル欄はブランクのまま何も入力しなくて結構です。

③タイトル欄、サブタイトル欄のその他の入力例は、サンプル集のSample-01～10を参照して下さい。

(8) 記事内容

ここには記事本文を入力します。

上記(5)の内容区分ごとの、記事内容欄への入力方法の違いと注意は、第4章以降の本文及びサンプル集に表示しますので確認して下さい。

また文字や記号などの各種変換規則は第5章にて説明します。

(9) 挿絵、写真の有無

ひとまとまりの記事中に挿絵（商品ロゴ、商標含む）、写真が有る場合は、有りのラジオボックスにチェックマークを入れます。

上記(5)で、⑬の区分を選択した場合も、有りでチェックマークを入れます。

(10) 挿絵、写真の説明

上記(7)が有りでチェックマークされた場合、該当の挿絵や写真の概要を説明する文章を入力します。紙面原稿の記事中に挿絵や写真の説明が入っている場合はこれを入れますが、説明が無いケースが大半です。このような場合は、入力者自身で説明文を考え、入力して下さい。

(11) 画像ファイル番号

紙面原稿の下部に、当該紙面の元の画像ファイルが印刷出力されています。

印刷例： F0010001.TIFF

この番号は静岡県が保有する画像磁気データのファイル名であり、このファイル名の前8桁は固有番号となっています。

この前8桁(上記の例でいうとF0010001)のみを、この欄に入力して下さい。

入力は半角で行います。下5桁(.TIFF)は不要ですので入力しないでください。

3. 入力必須項目

上記2の(1)～(11)の内必須項目は以下の通り。必ず何かを入力しなければなりません。これら必須項目への入力無しに送信すると、未入力項目がある旨の警告メッセージが表示されます。以下の必須項目に入れるべき内容が記事中に無い場合は、(記事タイトル無し)、(記事内容なし)のように()書きで入力すべき内容が無いことを書き込んでください。

入力必須項目

- (1) 新聞名
- (2) 発行年月日
- (3) 朝夕刊区分
- (4) 面数、段数
- (5) 内容区分
- (6) 記事タイトル
- (8) 記事内容

4. 記事の一まとまりとは

今回の記事テキスト化では、一まとまりの記事を1データとして入力、送信していただきます。WEBサーバに格納する際にも、また完成したデータを検索して画面に表示する場合にも、最初に入力・送信を行ったこの一まとまりの記事が1単位となります。ここで言う一まとまりの記事とは、以下のように定義されます。

「1紙面の中のまとまった位置に掲載され、同一内容について説明又は主張している記事の1まとまり。」

以下、サンプル集を実際に見ながら、1つの記事として処理する基準について説明します。

①Sample－0 1

冒頭のタイトル部と思われる箇所に「商業登記公告」とあって、以下二つの企業の登記記事が中仕切りを入れ掲載されています。

これら全体を一まとまりの記事と見なして入力をして、上記の定義から言えば誤りとはいえません。しかしこれをそのままひとまとまりの1データとして入力を行った場合、完成データがWEBサーバに格納され公開されたとき、例えば企業名を検索語にして検索したとすると、ヒットはするものの、画面に表示された記事には二つの企業の登記公告が画面に表示されることになります。このサンプルのケースでは2社だけですが、もっと多くなった場合には、だらだらといくつもの企業の記事が並び、非常に見にくい画面になってしまうこともありえます。このようなケースでは、これら複数社をひとまとまりとせず、5つの記事に分けて処理することとし、冒頭の「登記公告」を各々のタイトルとして個別に登記記事を入力するのが良いのではないかと思います。

Sample－0 1のケースは2社ですから、それほど見にくくはならないという判断で、一つの記事として入力を行う方法を採用することも、この場合は可能かと思えます。そうした場合には、WEB検索ヒット画面上の1つの記事枠で、複数の会社の記事を（WEB検索結果の）画面上でどのように区別して見せることができるかを考えた上での入力が必要となります。

前の会社の記事が終わったところで改行し、次ぎの会社の記事の先頭が必ずWEB画面の行の先頭に表示されるよう配慮したり、あるいはもう1行余分に改行して、1行分の空白行を作るとか・・・という入力時の工夫をしなければなりません。

繰り返しますが、**Sample－0 1**の事例の記事を、一まとまりの記事としても誤りではないのです。要は、最終的にWEBに公開され多数の人が検索し、画面に表示したとき、見る側（検索者、閲覧者）にとってどのように見えればわかりやすいかを頭において記事のひとまとまりを考えて頂くということです。

④Sample－0 2

「政界」というタイトルがあつて、その中に▲の記号で区別した複数の異なった内容の政治記事が掲載されています。この記事の場合も、**Sample-01**のときと同様に、全体を一つの記事とすべきか否かについて考えていただかなければなりません。結論を言えば、全体を6つの記事として6回に分けて入力、送信を行う方法が望ましいと考えます。まず一つの記事とするには長文過ぎて、WEB検索して表示した場合、入力時に区別して見せる工夫をするにしても、結果的にきわめてわかり難い画面となるというのが最大の理由です。

各々を別の記事として処理した場合のタイトル、サブタイトルの取り方は以下の通りとなるかと思ひます。

タイトル	政界	サブタイトル	政府の致照
記事内容	布哇政府より仲裁裁判を申込みたるに付ては他の独立国に仲裁を依頼する前諸般の決定すべき要件・・・・・・・・		

*以下残る5つの記事のタイトルは全て共通で **政界** とします。

5. 入力規則

ここでは、個々の記事を入力するにあたっての入力規則について説明します。

入力フォームでの入力時に、一連の文章が終了するまでは「改行」はしないでください。実際の新聞紙面の行変えや、入力フォームの枠にとらわれず、原則としてあくまでも一つの文章が終わった位置で改行するものとします。

紙面原稿と入力フォーム上での改行の例

実紙面上の記事

右の実紙面で行変えされている位置で、入力時に「改行」(Enter)して行を変えてはいけません。

(例-A)

一 社名 産業株式会社
 二 営業所 静岡県駿東郡清水村新宿10番地
 三 変更ノ理由
 増資本新株式 300 振込未済金 187
 50円ノ内明治30年7月16日第三回金
 3750円振込テ滿了ス
 右明治30年7月19日当区裁判所二備フル原
 登記簿中変更シタリ依テ此旨公告ス
 但原登記ハ明治17年3月28日株式會
 社登記簿第一冊3丁3号ニ登記シタルモノナ
 リ
 明治30年 沼津区裁判所
 7月19日

WEB入力フォーム上の入力と表示

(□は全角Space)

ここで改行してはいけない

1□社名□産業株式会社	(例-B)
2□営業所□静岡県駿東郡清水町新宿10番地	
3□変更の理由□増資本新株式300 振込未済金18750円の内明治30年7月16日第三回 金3750円振込で満了す□右明治30年7月19日当区裁判所に備ふる原登記簿中変更したり 依て此旨公告す□但原登記は明治17年3月28日株式会社登記簿第1冊3丁3号に登記したる ものなり□明治30年7月19日□沼津区裁判所(入力END)	

入力フォーム上で、1行を順に入力して行って右端にくと自動的に次行の先頭に（あたかも改行したように）次ぎの文字が表示されます。しかしこれは画面表示の枠に規定されて一見改行されているように見えているだけで、改行のように見える箇所に「改行情報」が入力されている訳ではないのです。

キーボード上で改行の為にEnterキーを押すという行為は、実は文字と文字の間に、目に見えない「改行情報ク」を入力するという事にほかなりません。

本来の改行は、WEB検索結果画面上で、左端文頭に持って来たい文字の直前で行なうのが原則です。このことを頭に入れておいてください。

以下に、念のため、入力フォームを使用した入力で起こりやすい改行誤りのパターンを記します。

- ①入力しているフォームの枠の右端がくると、ついEnterキーを押して（目に見えない）本文の改行情報を入力してしまった。
- ②入力するための原稿の上で行が変わるたびに、つい入力画面でもEnterキーを押して（目に見えない）本文の改行情報を入力してしまった。

(2) 改行とブランク

例-A、例-Bの中で、「右」から始まる文章と、「但」から始まる文章が、各々文章が一旦終わっているにもかかわらず、その位置で改行せず、代わりに全角スペースを一つ入れ、一つの文章としてあります。

必ずしも改行をしなくても、スペースを空ければ、WEB画面上で見難くならないと考えられるケースでは、この例の入力方法をとって頂いて結構です。

ただし、例－A、例－Bのケースでは、もし「右」、「但」の文字の前で改行するという判断をしたとしても誤りでは無いものとします。

(3) 箇条書き及び箇条書き文頭の記号の処理

紙面上の箇条書きの場合は、1箇条に属する一連の文章の末尾で改行します。

ただし、1箇条の中で、文意形式が異なる文が存在していて、改行したほうが（WEB画面上で）わかりやすい、見易いと思われるケースでは改行をしていただいて結構です。

箇条書きの文頭に、▲や□、◎などの記号が置かれている場合、その記号がJIS標準の記号であれば使用して支障ありません。

標準の記号に無い記号である場合は、良く似た記号に置き換えて入力してください。

(4) 投稿短歌、俳句、川柳などの取扱い

記事欄へは、以下の例のように入力して下さい。

金賞 めでたさも中くらいなりおらが春□小林一茶／銀賞 我と来て遊べや親のない雀□
山田侘助／銅賞□名月を取ってくれろと泣く子かな□田中真樹子
選評 小林氏秀逸 ○×△×女史 □○□○氏も秀なれど及ばず

*一句（首）ごとの改行は行わず／で区切ります。 作品と作者名の間は全角スペースを一つ空けます。

(5) 決算公告、市況など表形式記事の処理

法人の決算公告、市況などは、罫線こそありませんが、全ていわゆる表の形式で掲載されています。今回のテキスト化では、これを表の形式のまま入力し、かつ表示させることは不可能のため、これを箇条書き風、あるいは羅列式に収録せざるを得ません。

次頁のサンプルC，Dにならって入力してください。

(サンプル-C / 三段の表形式記事)

長野	東京	静岡	地名	上半期	明治三十年
二五、三二一、六二八	二五、三二一、六二八	二五、三二一、六二八	当期預り高	静岡貯蓄銀行各地取扱高一覧表	
三二一、六二八	三二一、六二八	三二一、六二八	当期払戻高		
二五、〇〇〇、〇〇〇	二五、〇〇〇、〇〇〇	二五、〇〇〇、〇〇〇	現預り高		

(サンプル-D / 入力見本)

- タイトル 静岡貯蓄銀行各地取扱高一覧
- サブタイトル 明治30年上半期
- 記事内容 (以下入力フォームでの記事入力見本)

(地名) 静岡□ (当期預り高) 25、321、628□ (当期戻高) 321、628□ (現預り高) 25、000、000
 / 東京□ 25、321、628□ 321、628□ 25、000、000 / 長野□ 25、321、628□ 321、628□ 25、000、000

- * □は、全角スペースを表します。
- * 第列のみ括弧書きで列のタイトルを入力し、次列以降は省略して可とします。
- * サブタイトルに、「明治30年上半期」と入力していますが、タイトルを「明治30年上半期□静岡貯蓄銀行各地取扱高一覧」とし、サブタイトルを空白としても可とします。

(6) 紙面に表記されている文字 (記号) 形と入力すべき文字形

ア. 全角文字、記号使用の原則

① 全角入力の原則

WEB入力画面中では、▼マークからリスト選択する項目を除き、原則として全て全角で入力を行うこととします。たとえ、紙面原稿の上で半角あるいは小さな文字として表示されている場合でも、入力はいくまでも全角で行います。

文中に空白を入力する場合についても、必ず全角空白となっているかを確認して下さい。洋数字や「、。」などの句読点についても例外無く全角を使用しますので注意が必要です。

② 記号使用について

紙面原稿上に表示されている記号は、原則としてあるがままに入力するものとします。ただし、その範囲はいくまでもJIS第1水準、第2水準の中の記号の範

囲に限るものとします。従って、紙面中にJ I Sに無い記号が有った場合、これを代わりの記号に置き換えて入力をする必要のあるケースがでてきます。

置き換える場合は、J I S内の記号の中から、より近似の形状の記号を選択します。

出現することは無いと思いますが、その記号自体が記事中で意味を持つもので、かつJ I S内記号に置き換えができないケースが生じた際は、事務局までご相談ください。

新聞紙面に出現頻度の高い記号で、①、②、③・・・などの丸付き数字の場合、洋数字の(1)、(2)、(3)、あるいは単順に1、2、3に置きかえることが考えられますが、同一記事内に、1、2、3や(1)、(2)、(3)が使用されている場合は、これらが相互に矛盾の無い様に置きかえるべき記号を選択しなければなりません。

*使用できない記号の一覧は、「詳細入力説明PDFマニュアル」のQA3-6に掲載しています。WEB新聞記事入力ページ～詳細入力説明のページに入り、PDFファイルを取得し、印刷してご覧ください。

イ. 旧字体の処理とJ I S内字使用の原則

新聞記事の中には多くの旧字やJ I Sに収載の無い漢字が出てきます。今回の入力では、原則としてJ I Sの第一水準、第二水準の文字(=J I S内字/約6千字強)のみを使用することとします。

個々のパソコン上ではJ I S内字以外の文字について、個別にご自分で外字を作成して画面表示したり、印刷することも可能ですが、インターネット上ではJ I S内字以外は表示が出来ない為です。今後の技術としてXMLのような文字の処理が可能となるとは思いますが、あくまで現状を踏まえ上記の対処をとることとしました。

①旧体字は現在の通用字体(常用漢字、当用漢字、相互に異体字関係にある複数のJ I S内字の内、いちばん簡単な文字)に置き換えて入力することを原則とします。澤、沢や邊、邊、辺のように、新旧字体がともに日常的に用いられている文字の場合でも、原則として沢、辺で変換して入力します。

理由は以下の2点です。

■複数の異体字がデータ内に共存してしまうと、記事検索で問題が生じます。「渡邊」という文字列を検索語として記事検索すると、「渡邊」や「渡辺」を含む記事は検索できないこととなります。渡邊、渡邊、渡辺のいずれで入力されているかわからない人物氏名を検索語として検索する際に、渡邊、渡邊、渡辺と、3回同じ検索をしないと、必要な情報にヒットしないということです。

■旧字体でもかなりの文字がJ I Sの第一水準、第二水準の中に存在します。従ってあるがままに入力することは不可能ではないのですが、この場合皆さんのパソコンに格納されているF E P（日本語変換機能、一般的にはI M EやA T O Kが主流です。）では、かな（カナ）漢字変換をしても、旧字体の多くはJ I S内字ではあっても、候補文字として表示されません。そうすると、コード入力を行うしかなくJ I Sの辞書（ワープロ辞書）を開いて正しいコードを探すという多大の手間をかけなければならなくなってしまいます。

ちょっと一服

実際に記事入力を始めると、まずは虫眼鏡、ルーペが必需品になると思います。次いで難解な旧字体漢字を探す辞書も手放せなくなるのではないかと思います。

そこでみなさんに、漢字検索可能なWebとソフトを紹介します。一度各サイトを訪問してみてください。

ハイパー漢字検索 <http://www2.theta.co.jp/kanji/> 無償サイトです。検索方法がいくつか用意されていて、その中でも、「ハイパー検索」は役に立ちます。例えば、「巖」という文字を検索するのに、「山」「耳」「口」と文字を部分に分解してキーワードにします。読みがわからない、部首名もよくわからないなど、自分にとって未知の漢字を探す時に最適のツールです。無償ということもあり、事務局からの一番のお奨めです。一度お試しを。

今昔文字鏡 <http://www.mojikyo.com/> 私どもNPOエイジングブライght倶楽部と提携、協力関係にあるNPO今昔文字鏡研究会（漢字を専門とする世界の学者の組織です）が作成した世界最大（漢字12万字）の漢字検索ソフトです。

「解字木検索機能」が秀逸。私（渡邊）も使っていて重宝していますが、28,000円とちょっと高いのが玉にキズです。しかし、「大漢和辞典」（大修館書店／日本最大の漢和辞典／全17巻）が5万字で20万円以上の価格であることを考えれば、大漢和の5万字を全て含む12万字でこの値段というのは安いとも言える。こちらでも有償ですがWEB上で検索する契約も可能。

②紙面上に記載されている漢字形がJ I Sの内字の中に有るが、同じくJ I S内字の中に、同意、同義、同源の他の文字形が複数存在する場合があります。このようなケースで、紙面に記載されている文字が旧体字形であれば、この文字を入力せず、通用字体（最も簡単な字体）で入力する事とします。下の例を参照して下さい。

（紙面の文字）

釵

（J I S内の異体字形）

釵、釵、釵、釵、釵

（入力する文字）

釵

* J I S 第 1、第 2 水準の文字の中で、相互に異体字関係にある文字を集めた「異体字一覧表」を資料として添付します。利用して下さい。

③とは言え、漢和辞典などに字源を遡れば相互に同字とされている文字の中には、下例のように、現在では異なった用方をされ、違う文字と一般に考えられている文字もあります。このような場合は、別の文字として置き換えを行わずに入力して下さい。

《異体字の判断》

《置き換え不可の事例》 楳⇒楳（梅とはしない）寅⇒寅（虎とはしない）

（帛⇒虎は、置き換えを行う）

④紙面に表記の字形は J I S 内字に存在しない字形だが、その文字の異体字形が J I S 内字に一文字存在する場合の処理

紙面に記載されている字形が J I S の内字には存在しない場合は、必ず漢和辞典でその文字を調べ、辞書に記載されている異体字（□の同字、□の誤字、□の譌字、□の俗字、□の元字、□の本字、□の旧字などと辞書に書かれていますが、これらの□に当てはまる文字が、辞書で探した文字の異体字にあたります）の全てを一文字ずつ、J I S の辞書（ワープロ辞書など）を引き、存在しないかどうか調査して下さい。

調査の結果、J I S 内字に存在すれば、この J I S 内字形に置き換えて入力します。また調査の結果、異体字として複数の J I S 内字が有った場合は、上記②の規則に

準じ、そのうちの通用字体（常用漢字、当用漢字、より簡単な字形）を選択して入力して下さい。

⑤紙面に表記の字形は J I S 内字に存在しない字形で、かつその異体字形も J I S 内字に存在しない文字の処理

上記④で漢和辞典の調査をし、その異体字について J I S の辞書も調べた結果、J I S 内字の中に、該当する異体字形が存在しなかったケースでは、当座その文字 1 文字の代わりに、全角の記号 > を文字数（一文字であれば 1 個）だけ入力するものとします。

このケースでは、当該文字の字形と、収載されていた画像番号、段数、タイトルを

事務局宛てにファクシミリ等の方法（MS ペイントで文字形を書き、ビットマップファイルの形でメールに添付して頂いても可）で必ずお知らせください。事務局にて再度調査を行います。

ウ. 和数字の処理

和数字は、原則として全て洋数字に置き換えて入力を行います。箇条書きの一、二あるいは一、一の場合も洋数字に置き換えます。

ただし、氏名の文字、地名、その他の固有名詞の中の文字として和数字が使われている下例の類の場合は、そのまま和数字を入力するものとします。

《和数字のまま入力すべき例》

一郎、二郎、三軒茶屋、三味線、三保、三番町

エ. 変体仮名の処理

紙面には、主として女性の名として「変体仮名」が使用されている場合があります。これらの変体仮名は、J I S 内字には存在しませんので、すべて現代かなで入力を行うものとします。変体仮名は、現代ひらがなに較べ、より漢字に近い崩し字形といえます。従って、変体仮名には必ず元となった漢字形が存在しますし、種類によっては元漢字形をすこし崩した程度の形状のものもあります。しかし、元漢字がわかったからといって、その漢字を代わりに入力するのは誤りです。ご注意ください。

今回の資料の中に、変体仮名表を入れておきます。紙面に出現した場合にお使いください。ただし、この表には無い元漢字の変体仮名も時として出現します。この場合は事務局にファクシミリにて字形をお送り頂ければお答えします。

オ. 重複音記号の処理

紙面にはさまざまな重複音記号が出現します。原則として重複記号は漢字形に近い一部以外は使用しないものとします。出現した場合の置き換えは、以下の①～

④の要領で行ってください。

①かゝさま、かㇿさま、かゝさま ⇒ かかさま で入力します。

②たゞ同然、たゞ同然 ⇒ ただ同然 で入力します。

③なんと ⇒ なんと なんと で入力します。

④全箇所 ⇒ 全箇所 でそのまま置き換えず入力します。

カ. 句読点の処理

原則として、句読点は、紙面に有ればそのまま入力します。あるべき位置に「、」や「。」が無い場合はそのまま無い状態で入力して下さい。「。」などは、紙面に無くてもつい習慣でキーボードをたたいて入れがちですので、十分に注意して下さい。また、下例のように、二つの文が行変えなしに連続して表記されていて、前の文末に「。」が無いというケースでは、やはり「。」は入力せず、代わりに文の区切りを示すために1ブランク（全角1スペース）を入力する事とします。

《前文末に「。」無し、後文続く入力例》 以上がことの顛末也口さてその後

キ. カタカナ表記について

文中の送りがながカタカナ表記されている場合がまま出現します。この場合、必ずひらがなに置き換えて入力をしてください。

ただし、送りがな以外の氏名、固有名詞、地名、擬音としてカタカナが使われている場合は、そのままカタカナで入力して下さい。

(7) 文字使用に関するその他の注意点

ア. ふりがなの処理

紙面には、漢字の右に小さな活字でルビがうってある場合がありますが、今回のテキスト化では、再現のしようが無いため、ふりがな、ルビは一切収録を行いません。

イ. 旧かな使い

てふてふ=ちょうちょう、かふ思い=こう思い、けふ=きょう・・・などに代表される旧かな使いが紙面には頻出します。

これらは全てあるがままに入力を行うものとします。てふてふと有ればそのままてふてふと入力して下さい。

ウ. 旧送りがな

紙面には、現在の正しいとされる送りがなとかなりずれた用法が頻繁に出てきます。旧送りがなと言うよりも、むしろ明治期における漢文体表記から漢字かな混じり表記への変化の過程で起きている不統一、無基準と考えるべきかと思います。これらについても、下例の通り、原文のままの入力を行うものとします。

依て 知べからざる 振込て

エ. 濁点の有無

紙面には、本来濁点のあるべきかな部分に濁点が無い場合がまま頭れますが、これについてもあくまで原文どおりとして入力を行うものとします。

(8) 不明文字の処理

ア. 紙面原稿上の文字がカスレや、にじみ、ツブレなどで判別不可能な場合の処理
上記(6)ーイー⑤のケースと同様に、判別不可能な文字数だけ全角記号 > を入力して下さい。

紙面の一区画が失われている場合や、一定面積が汚れなどで見えない場合には、> (紙面2段左部分欠落) > のように入力して下さい。

イ. 紙面原稿上の文字の点画の判別はできるが、文字そのものが難字で判断できず、漢字変換して入力できない場合の対処

辞書をしらべてもどうしても判らない文字で、従って入力すべき文字もわからないというケースでは、(6)ーイー⑤の対処を行いません。

繰り返し申し上げますが、>で入力をして良いのは、徹底した辞書調べを行ったにもかかわらず、入力すべき文字が見つからないという場合だけです。

調査を充分行わないで簡単に諦めてしまい、>を入力する頻度が高くなると、結果としてWEB読者の信頼を失うことになります。

■このようなケースで>を入力した時には、その字形を大きく書いて、当該文字のある紙面の画像番号と段数、タイトルを書いた紙を事務局宛てにご送付ください。事務局にて再度調査を行いたいと思います。

(9) 外国名、外国都市名の漢字表記

布哇（ハワイ）、倫敦（ロンドン）、華盛頓（ワシントン）、桑港（サンフランシスコ）、巴里（パリ）、伯林（ベルリン）、羅馬（ローマ）、伯刺西爾（ブラジル）、亜爾然丁（アルゼンチン）、墨西哥（メキシコ）、玖馬（キューバ）、紐育（ニューヨーク）、亜米利加（アメリカ）など、国名や都市名を漢字で表記している紙面が多く出現します。これらは、原則としてそのまま漢字にて表記を行います。（カタカナに置き換えは不可。ただし紙面にカタカナで表記されていればそのままカタカナで入力します。）

また、都市名、国名に使用されている漢字が旧字体である場合には、一般の記事中の文字の扱いと同じく、(6)のイのルールを適用します。

6. 補記

(1) タイトルのみで、いわゆる記事に該当するものがない場合記事内容欄に（記事無し）と入力して下さい。

(2) 商品広告中に良く出てくる商標については、収録するにいかんともしがたい為、これらを絵、写真ととらえ、「添付写真、絵」の欄を有りでチェックし、写真、絵の内容欄に、「商標」と入力してください。

(3) 記事中の文字の大きさの変化を収録することも不可能ですので、一文中に大きな文字で強調している部分があっても、同じ大きさの文字で入力してください。

太字や文字色を変えるなどをしないでください。