

青空文庫作業員 作業マニュアル

〔Windows版 Ver.1.0 1998年12月作成〕

目次

はじめに

- 1．底本と著作権
- 2．作業の流れ
- 3．入力
- 4．校正
- 5．加工

はじめに

青空文庫は、テキストの入力や校正などの作業を多数のボランティアの方々をお願いすることによって成り立っています。私たち呼びかけ人（運営メンバー）は、それらの方々の「作業員」の名称で呼んでいます。

このマニュアルは、主にこれから作業員になっていただく方々のために、底本選びから登録にいたるまでの作業の概略をご紹介しますものです。作業の具体的内容の説明については、「印刷物の制作にたずさわった経験がなく、パソコンを使い始めてまもない方」を念頭に置いています。そのため、出版社などへの勤務経験がある方、パソコンの操作に習熟なさっている方には不必要な情報もあはずです。

ただし、このマニュアルの中では「青空文庫独自の約束事」にもふれています。無用な修正作業の発生を防ぐために、どなたも一通りは目を通していただきたいと思います。

なお、底本を選ぶ際に必要となる著作権関連の情報については、ここでは詳しくは取り上げません。これについては、「青空文庫からのメッセージ 本という財産とどう向き合うか」をご参照ください。

（青空文庫・呼びかけ人一同）

1. 底本と著作権

青空文庫は、「電子化された本（テキスト）を誰もがいつでも自由に読める」ようにすることを目的とした「インターネット上の本棚」です。ここには、以下の2種類の著作物をおさめています。

- 1) すでに著作権が消滅した著作物。
- 2) 著作権所有者が金銭の授受をとまなわない公開に同意した著作物。

すべての著作物は著作権によって保護されています。著作者が亡くなったあとも、著作権の保護は続きます。現行の法律では、その期間は、「著作者の死後50年間」と定められています。より正確には、「著作権は著作者の死から50年を経過した年の12月31日をもって消滅」し、自由に公開できるのは「翌年1月1日から」です。

「著作権が活着している作品」については、「著作権所有者への公開の依頼は行わない」を原則とします。したがって、登録は「著作権所有者ご本人からの申し入れ」があった場合もしくはそれに準ずる場合に限られることとなります。

底本を選ぶ

以上のことは、底本（入力の前にする本）をどう選ぶかに直接関係してきます。いくら好きな小説だからといって、「著作権が活着している作品」を勝手に登録・公開するわけにはいきません。その小説が翻訳作品の場合には、原著者の著作権、翻訳者の著作権が存在します。

登録・公開したい作品があったら、まず原本の著者紹介欄などで著者の生死、逝去年を調べ、著作権について確認してください。文学系の著作者については、青空文庫内にある「著作権が消滅した作家一覧」<http://www.voyager.co.jp/aozora/siryo1.html>を参照してください。

短編集や論文集などの場合には、作者だけでなく、作品の選定や配

2 . 作業の流れ

列などにあった「編者（編集者・編纂者）」の著作権が存在する場合があります。著作権とは全く縁遠いように思われる古文で書かれた古典作品の中にも、注釈や校訂などに関する著作権が存在する場合があります。著作権の保護期間を終えていない筆者による注釈などは、登録・公開できません。

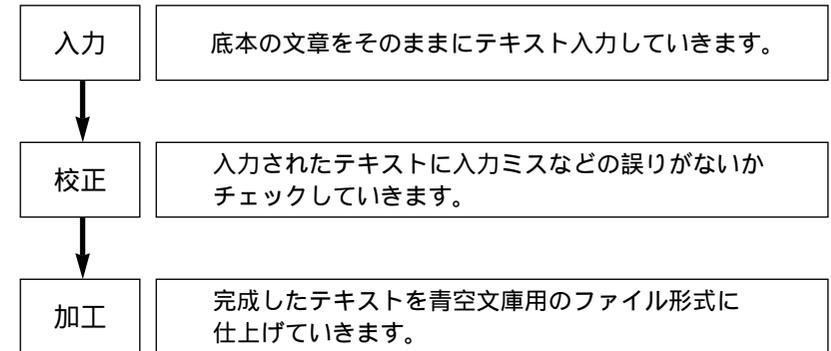
次に、青空文庫のインデックスをご覧になって、すでに登録済みの作品に含まれていないかどうかを確認してください。また、青空文庫への登録を前提として、すでに入力などの作業が進んでいる作品もあります。これについては、同じく青空文庫内の「入力中の作品」<http://www.voyager.co.jp/aozora/nyuryokutyu.html>をご覧ください。

インターネット上では、電子化されたテキストがさまざまな場所で公開されています。これについては、「電子テキストのある場所」<http://www.voyager.co.jp/aozora/denshitext/denshitext.html>をご覧ください。

ただし、以上の手続きを踏んでも、判断がつかないことがあるかもしれません。また、青空文庫の「入力中の作品」や他サイトのデータの更新が遅れている場合もあります。負担の大きい入力作業の重複を避けるために、入力したい作品のめどがついたら、まずは青空文庫aozora@voyager.co.jp宛にご連絡ください。

「著作権が消滅していて、まだどこでも公開されていず、誰も入力にとりかかっていないと思われる作品」であることが確認できたら、そのときがスタートです。

青空文庫に登録される作品は、次のような過程でつくられます。



これらの作業は、ふつう、複数の担当者が分担してあたっています。現在のところは、作業員の方々には主に入力と校正をお願いし、内容・形式の最終チェックと基本フォーマット・ファイルへの加工は私たち呼びかけ人が行っています。

3 . 入力

1冊の本はさまざまな要素から成り立っています。本文だけを見ても、例えば文章にはルビ（ふりがな）や注釈があります。表組にした文章もあります。写真や挿し絵が使われている本も多数あります。そのうち、ここで「入力」と呼ぶのは、ルビや注釈・表組を含めたテキスト（文字）の入力です。内容の理解を図っていく上で写真や挿し絵がぜひとも必要という場合には、別途ご連絡ください。その上で、著作権の扱いや入力の処理についてご相談したいと思います。さて、入力です。作業にかかる前に1つだけ注意していただきたいことがあります。それは、「底本のできるだけ忠実な再現」を目標に置き、「勝手な編集はしない」ということです。

といっても、印刷物をそのまま再現することはできません。ここで「忠実に」というのは、例えば漢字の使い方や送りがなは「底本どおりに」するということです。旧かなのテキストであれば旧かなのままに、時代によって変化の激しい送りがなについても「底本のままに入力していく」ことを大原則としてください。

入力方法

入力には、キーボードをお使いになる方が大半でしょう。その際に利用なさるアプリケーション・ソフトは、テキスト入力に使えるものであれば、どんな種類のものであってもかまいません。シンプルなテキスト・エディターであっても、ワープロ・ソフトであっても、あるいはアウトライン・プロセッサなどの特殊なエディターであっても、ご自分が使い慣れたものをお使いください。

注意していただきたいのは、ただ1つ、「誰もがファイルを開くことのできる形式で保存する」ことだけです（文書の保存）。

入力中は、できるだけこまめに保存してください。パソコンにはエラーがつきものです。また、停電などのトラブルもあります。適度な頻度でファイルを保存し、入力済みのデータが不意の事故で失わ

れないようにすることが大切です。

【OCRの活用】

書籍のページなどをスキャナーで読み込み、読み込まれた画像データを解析してテキストデータに変換するしくみを、OCR（Optical Character Reader）といいます。最近ではOCRソフトの精度が向上し、テキスト入力に活用なさる方が増えてきました。OCRの活用は負担の大きい入力作業の軽減につながります。

ただし、100%の精度で変換できるOCRソフトはありません。ソフト自体が優秀であっても、スキャニングに用いる書籍やコピーの質の問題もあります。OCRによるテキスト入力は必ずエディターなどによる修正作業をとまないとします。

基本となる書式

入力にあたり、青空文庫では下記のような独自の書式（ルール）を定めています。

- 1) 一番最初の行には「作品名」を入れる。
- 2) その次の行に「著者名」を入れる。
- 3) ファイルの最後の部分には、「使用した底本の正式名称・出版社・発行年月日、入力者名・使用機種」を入れる。

文庫本などではそれ自体に底本（全集など）が存在する場合があります。その場合には、「親本（底本の底本）の正式名称・出版社・発行年月日」もわかる範囲でつけ加えてください。

テキスト・ファイル（文書の保存）では、レイアウトに関する形式まで含めて底本をそのまま再現することはできません。そのために生ずる補足事項については、下記のような「入力者注」を活用します。

【入力者注の形式】

全体を「[]」（区点コード0146・0147、JISコード214E・214F、シフ

トJISコード816D・816E。キーボードから直接入力できます)でくり、[]内の最初には井げた記号# (区点コード0184、JISコード2174、シフトJISコード8194。キーボードから直接入力できます)を入れます。(シャープ記号 区点コード0284、JISコード2274、シフトJISコード81F2と混同しやすいので注意してください。)

底本の何ページ、何行目にあるかを示す必要がある場合は、「数(ページ)-数(行)」の形式で示します。「-」は英文のハイフンです。

本文中に「[]」が使われている作品の場合は、「[] 【】」などのように入力者注を表す記号を変更する必要があります。このときは、ファイルの最初の部分(本文の前)にも

[#入力者注を表す[]は本文中に使われているので【】に変更]

と入力者注を入れます。

入力の実際

入力にあたっては、以下のようなルールを守ってください。

(1) レイアウト関連

【行頭の字下げ】

ほとんどの書籍では、新しい段落の始まりの部分が1文字下げであります。これはワープロ・ソフトなどでは、ふつう、「インデント設定」を用い、ルーラーや数値で指定します。しかしながら、青空文庫用のファイルではインデント設定は使わずに、段落の最初に全角のスペースを入れてください。

【引用文】

書籍によっては、引用文が頻繁に出てくることがあります。しかも、引用文の扱いは書籍によってまちまちで、通常の段落との区別の仕方が「1字下げ」のものもあれば、「2字下げ」のものもあります。

テキスト入力では、それをそのままに再現することはできません。そこで、かわりに引用文のはじめには[#ここから引用文、 字下げ、本文とは 行アキ] 引用文の終わりには[#ここで引用文終わり]という形の入力者注を入れてください。

【行の長さ】

写真やイラストレーションの入った書籍では、その部分だけ行の長さが変更になっていることがよくあります。上記の引用文と同じく、このようなレイアウトは再現できませんので、無視してください。ただし、写真やイラストレーションの著作権が切れていて、それらを入れた形で登録したいという場合には[#写真入る]などの注記を入れてください。

【改行】

底本(印刷物)は、それぞれ、固有の行の長さを基準にレイアウトが固定されています。入力の際には、「底本に合わせて、1行1行の行末に改行(リターン)を入れる」ことは絶対に避けてください。「改行は段落から段落への移り目だけに入れる」が原則です。

【表組】

テキスト・ファイルで表組を正確に再現することはできません。簡単な表組の場合には、項目と項目との間でTABキーを押し、TABで区切って入力してください。また、複雑な表組の場合は、別途ご連絡ください。

(2) 外国語

【アルファベット】

底本中に外国語がそのまま使われている場合は、全角のアルファベットは使わずに、英文モードで入力してください。

J RやN A S Aなどのように、縦書き表示されることの多い略語の場合は、和文モード(全角のアルファベット)で入力してください。

【英語以外の外国語】

Windowsでは、「テキスト・ファイル」には文書形式などの情報がないため、言語に関する情報が保存できません。そのため、ワープロ・ソフトで言語を切り換えて、アクセント（´）が付いたフランス語のアルファベット、ウムラウト（¨）が付いたドイツ語のアルファベットなどを入力しても、「テキスト形式」で保存すると言語の情報が失われ、文字化けしてしまいます。そこで、英語以外の外国語で使用される特殊なアルファベットは、

ae ao , ae ao , eo , aeo eo ! [# この行の「e」はすべてアクセント（´）付き]

のように入力者注で処理します。

(3) 数字と記号

【数字】

1文字のアラビア数字（算数字）は全角で、2文字以上の場合は半角で（10キーを使って）入力してください。

底本の表記が「一九九八年」のような漢数字の場合は、そのまま漢数字で入力してください。

【記号】

「(、」)、(、)」、「、,」、「、,」、「・」などは、すべて和文モードで入力してください。これらはすべてキーボードから直接入力できます。

【特殊記号】

MSIMEやATOKの文字パレットにある「修飾数字」や「ローマ数字」などは異機種間の互換性がないので、使用しないでください。入力時には、普通の数字などへの置き換えが必要になります。（異機種間の互換性）

【単位】

「cm」や「mg」などの単位は英語の小文字として扱い、英文モード

で入力してください。

MSIMEやATOKの文字パレットには「cm」や「mg」などが全角1文字の記号として用意されています。これらは異機種間の互換性がないので、使用しないでください。（異機種間の互換性）

【ハイフンとダッシュ】

記号類の多くは、英文から和文に移植されたものです。そのうち、英単語の複合語や半角数字の電話番号で使われる「-」は、原則として英文のハイフンを使用してください。

ハイフンを少し長くした形のダッシュは、ダッシュ記号「-」（区点コード0129、JISコード213D、シフトJISコード815C）を使用してください。（キーボードから直接入力できるのは、線が少し短いマイナス記号です。混同しないよう注意してください。）

印刷物でよく使われる2倍の長さ（全角2文字分）のダッシュは、上記のダッシュ記号を2つ続け、「- -」と入力してください。実線についても同様です。

ダッシュが二重になった形の記号は、等号「=」（区点コード0165、JISコード2161、シフトJISコード8181。キーボードから直接入力できます）を使用してください。

【点線】

底本が「.....」のようにになっている場合は、3点リーダー「...」（区点コード0136、JISコード2144、シフトJISコード8163）を使用してください。

底本が「- - -」のようにになっている場合は、2点リーダー「- -」（区点コード0137、JISコード2145、シフトJISコード8164）を使用してください。

底本が「・・・」のようにになっている場合は、中黒「・」（区点コード0106、JISコード2126、シフトJISコード8145。キーボードから直接入力できます）を使用してください。

(4) 特殊な表記

【ルビ】

ルビは該当する熟語のあとに全角の「《》」(区点コード0152・0153、JISコード2154・2155、シフトJISコード8173・8174、)でくくって入れます。(学術記号の区点コード0267・0268、JISコード2263・2264、シフトJISコード81E1・81E2と混同しやすいので注意してください。)

例えば、底本が

耳まで^{ほて}火照って来る

の場合には、

耳まで火照《ほて》って来る

のように入力します。ルビ自体がカタカナの場合は、カタカナで入力します。「半角カタカナ」はコンピュータ上だけで使用される特殊な文字種です。青空文庫のための入力では使用しないでください。)印刷物では、ふつう、ルビには拗促音(「っ」などの小さなかな)は使いませんが、これは活版印刷時代のなごりです。青空文庫の入力作業では、ルビにも拗促音を使用してください。ただし、拗促音を使用しない旧かなのテキストは例外とします。

漢字の熟語が2つ以上連続し、そのあとに「《ルビ文字》」を入れてただけでは、熟語とルビとの対応関係が不明確になることがあります。ルビのつかない熟語のあとにルビのつく熟語が続く場合は、

一応 何時《いつ》もの

のように「|」(区点コード0135、JISコード2143、シフトJISコード8162。キーボードから直接入力できます)の区切りを入れてください。

また、ルビ付きの熟語が連続する場合には、

一応《いちおう》何時《いつ》もの

のようにルビを分割します。ただし、複合語として定着している語

の場合は、

一瀉千里《いっしゃせんり》

のように1つにまとめます。

【傍点】

強調のために語句のわきに打つ点を、傍点といいます。傍点は、入力者注として処理してください。例えば、底本が

胡麻塩^{おやじ}

の場合には、

胡麻塩おやじ[# 「おやじ」に傍点]

と入力します。(入力者注の活用)

傍点には「、」の形のものだけでなく「」の形のものも多く使われています。この場合は、「丸傍点」と記述して区別してください。また、稀には「」の形のものもあり、これについては「白丸傍点」「白三角傍点」と記述してください。印刷物によっては大きめの丸傍点が使われている場合がありますが、大きさは無視してください。

【ママ】

作品中に「ママ」というルビがふってある文字がときどきあります。

^{ママ}
吹喋

これは形はルビであっても意味はルビではなく、「親本(原稿)の誤植(誤字)と思われるがそのまま(ママ)にする」という注釈です。

これについては、ルビではなく入力者注を用い、

吹喋[# 「喋」に「ママ」の注記]

のように入力してください。

【踊り字】

「人々」の「々」のように同じ文字の繰り返しを表す文字を、踊り

字といいます。踊り字には、右の例のように2倍の長さ（全角2文字分）のものもあります。2倍の踊り字は、JIS（JISと外字）には含まれていません。代わりに「ノヽ」（区点コード0131・0132、JISコード213F・2140、シフトJISコード815E・815F）を入力してください。

2倍の踊り字には濁点付きのものがあります。この場合は、間に「」（区点コード0177、JISコード216D、シフトJISコード818D）を入れて「ノヽ」を入力します。

印刷物は行の長さが固定です。そこで、「々」「ゝ」などの踊り字が行頭にくる場合には本来の漢字に置き換えて、「人々 人人」のようにされていることがよくあります。しかしながら、青空文庫の入力作業ではこの処理は無視し、「々」などに置き換えてください。

【返り点】

引用文として漢文が用いられている場合があります。漢文につきものの「レ点」などの返り点は、入力者注を使って、

居無レ [# 「レ」は返り点] 幾拜二 [# 「二」は返り点] 中納言一 [# 「一」は返り点]

のように入力してください。なお、「レ」はカタカナで代用します。

【上付き文字、下付き文字】

数式や元素記号などで使われる行の上部や下部に小さく入る文字は、入力者注として処理してください。例えば、「2の2乗」を表す数式は、

22 [# 2つめの「2」は上付き小文字]

と入力します。また、水の元素記号は、

H2 [# 「2」は下付き小文字] O

と入力します。

【伏せ字】

第二次大戦前～戦中など、国家による検閲が盛んに行われた時代に

達雄はいらくした。

発行された書籍には、伏せ字が使われているものがあります。伏せ字を表す記号は、原則として全角のアスタリスク「*」（区点コード0186、JISコード2176、シフトJISコード8196）を使ってください。

伏せ字の使用については、たいていの場合、底本にその旨の注釈があります。もしも何もふれられていない場合には、ファイルの最初の部分（本文の前）に、

[#本文中、伏せ字は「*」で表した。]

と、入力者注を入れてください。

どう処理すべきか迷ったとき

世に出ている書籍の数は膨大、しかも内容は多岐にわたります。入力途中でどう処理していいか迷う場面が、きっと出てくるはず。また、底本の文字がかけていて判読できないなどといったトラブルもありえます。処理に困ったときは、原則として「入力者注」で処理してください。（入力者注の形式）

【誤植・誤記・脱字】

本にはしばしば誤植があります。著者の誤記と思われる表現にもぶつかります。明らかに誤植、誤記、脱字だと判断できる文字は、入力者注で処理してください。迷う場合には、原文通りに入力してください。

例えば、「宮沢憲治」のように明らかな誤植・誤記と思われる場合は、

宮沢賢治 [# 「賢治」は底本では「憲治」と誤記]

のように、正しい文字に置き換えた上で入力者注を付けてください。

【判読不能などの場合】

文字が判読できないなどの場合は、その文字のかわりに「」（区点コード0208、JISコード2228、シフトJISコード81A6）を入れ、

[#判読不可、30-16]

のように入力者注で処理してください。

漢字を探す

テキストを入力していて、必ず悩まされる問題があります。それは目的の漢字が、かな漢字変換ソフト（フロント・エンド・プロセス、略してFEPともいいます。Windowsでは標準装備の「MSIME」、ジャストシステム社の「ATOK」が代表的です）の候補に出てこないことです。例えば、「聚落」を入力したいときに「じゅらく」とひらがなで入力して変換すると、ATOK11で表示されるのは「入洛」だけです。これは「聚」と「落」という2つの漢字がコンピュータ上で使える文字であるにもかかわらず、「聚落という熟語」が「ATOK11の辞書には登録されていない」ことを示します。

変換しても目的の漢字や熟語が表示されない場合には、

- 1) 漢字辞典などでJISで定められたコードを調べ、コード入力もしくは文字パレットを使って入力する。
- 2) FEPの文字パレットを使って同じ部首や読みなどの文字群の中から該当文字を探す。

などの方法で単漢字入力をします。

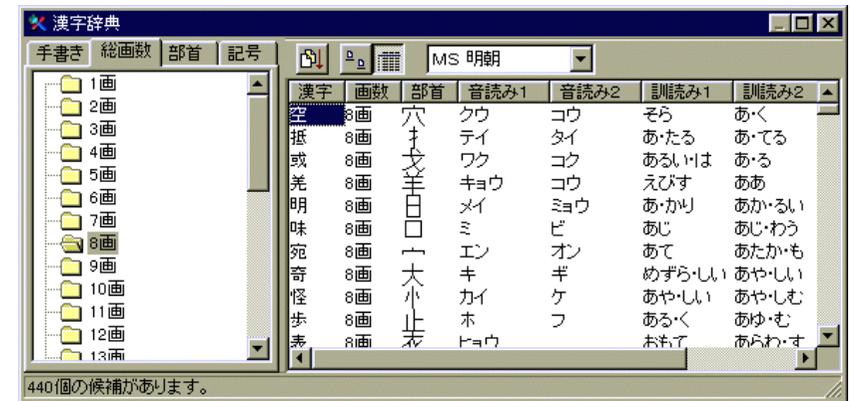
【必要な装備】

テキスト入力にあたっては、JISで定められたコードが掲載された漢字辞典が必須です。少々高価ですが『JIS漢字字典』（日本規格協会）または複数の出版社から刊行されている『ワープロ漢字辞典』『パソコン・ワープロ漢字辞典』の部類をぜひ用意してください。

最近では、『新版漢語林』（大修館書店）など、コンピュータ用のコードを併記した漢和辞典も増えています。『ワープロ漢字辞典』とは違って、JIS外の文字を調べることができますから、これも必須の装備です。

【文字パレットの表示】

MSIME97、98の場合は、MSIMEツールバーの「鉛筆と虫眼鏡（漢字辞典）」、MSIME95の場合は、「虫眼鏡（単漢字変換）」のアイコン



MSIMEの文字パレット



ATOKの文字パレット

をクリックすると、文字パレットが出てきます。総画数、部首などのタブをクリックし、さらに、それぞれに含まれるフォルダのアイコンをクリックすると、文字の一覧が表示されます。一覧から必要な文字をクリックすると、カーソルのある位置に入力されます。

ATOK11、ATOK12の場合は、ATOKツールバーの「鉛筆とパレット（文字パレット）」のアイコンをクリックすると、文字パレットが出てきます。「漢字検索」のタブをクリックすると、部首、読みで漢

字を検索することができます。(ATOK12の場合は、総画数でも検索できます。)
「検索結果」のボックスに表示された文字をクリックすると、カーソルのある位置に入力されます。ATOK9、ATOK10にも、文字パレットは用意されています。

【コード入力】

コードには、「区点コード」「JISコード」「シフトJISコード」の3種があります。これらは「JIS漢字コードをどう表記し、JIS漢字コード中に含まれる文字に対してどんな符号を割り当てるか」の違いであり、元になるJIS漢字コードそのものは共通です。詳しくは、専門書などをご覧ください。

【コード入力の実際】

MSIMEの場合

1) 日本語入力をオンにして、右図のように入力モードを「全角英数」か「半角英数」にします。



2) 目的の漢字のコード (JISコードまたはシフトJISコード) を入力します。

3) キーボードのF5キーまたはCtrl+Yキーを押します。これで、目的の文字が現れます。クリックするか、Enterキーを押すと、カーソルのある位置に入力されます。(右図)



ATOKの場合

1) 日本語入力をオンにすると、右図のようなツールバーが表示されます。



2) キーボードのF10キーを押すと、半角モードになります。さらにもう一度押すと、上図のようにシフトJISコードの入力モードになります。「シフトJIS」と書かれているボタンをクリックすると、「区点」「JIS」と切り替わります。

3) 目的の文字のシフトJISコード (区点コード、JISコード) を入力します。これで、目的の文字が現れます。現れた文字をクリックすれば、カーソルのある位置に入力されます。

この他に、ATOKの場合は、F10を押してコード順の一覧を表示することもできます。F10を1回押すと半角モードに、もう一度押すとコード入力モードになり、さらにもう1度押すと記号モードになって、該当するコード表の先頭から10文字ずつ一覧が出てきます。上下矢印でページ移動が、左右矢印で文字選択ができます。文字をクリックするか、選択してEnterキーを押すと、カーソルのある位置に入力されます。「記号モード」と名付けられていますが、Windowsで扱える文字コードの文字は、すべてこの方法で入力できます。

MSIMEの場合も、環境設定で「テンプレート」を「ATOK」にしておけば、F10キーを2回押すと該当するコードの先頭から一覧が出てきます。ただし、この設定では、前述したF5キーまたはCtrl+Yキーによる変換はできません。

【JISと外字】

コンピュータ上で共通に使える和文文字は、基本的には、JIS (日本工業規格。現在使用されているのは、JIS X0208) で定められた「第一水準漢字 (2965字)」「第二水準漢字 (3390字)」「非漢字」「記号」だけです。これらに含まれない文字を、外字といいます。

外字の一部は、JISの空き領域に埋め込まれています。しかしながら、

このような処理をした外字は異機種間の互換性がないので、使用しないでください。(異機種間の互換性)

入力にワープロ・ソフトを使う場合は、JISの空き領域やASCIIコードを割り当てた市販の外字専用フォント(書体)にも注意が必要です。この種のフォントには異機種間の互換性だけでなくWindowsユーザー間の互換性までが失われる懸念がありますので、使用しないでください。

【外字の処理】

外字は、すべて入力者注で処理してください。例えば、

喉を搔き撈って

の場合には、

喉を搔き [# 「 」は「てへん+劣」、読みは「むし」、30-16]
って

のように、代わりに「 」を入れ、さらに元の文字の形についての説明と読みを加えます。そのあとには、「ページ-行」の形式で底本の何ページ・何行目にあるかを記述します。

外字は、多くの場合、底本でもルビ付きで使われています。上記の例にルビがついて、

喉を搔き^{むし}撈って

となっている場合には、

喉を搔き [# 「 」は「てへん+劣」、30-16] 《むし》
のように入力します。

【異機種間の互換性】

JISの空き領域にどんな外字が埋め込まれているかは、メーカー(OS)によって異なります。例えば、Windowsで入力した 付きの修飾数字(MSIMEの特殊文字、ATOKの丸数字)「①」(区点コー

ド1301、JISコード2D21、シフトJISコード8740)は、Macintoshでは「(田)」と表示されます。

このような「異機種間の互換性のない文字」は、

1 [# 「 1 」は底本では 付き数字] インターネット
のように、JIS漢字コードに含まれる文字に置き換えた上で、入力者注を付けてください。

区点コード0833から1594まで、JISコード2841から2F7Eまで、シフトJISコード84BFから889Eまでに割り振ってある文字や記号は、異機種間の互換性がありません。

また、区点コード8901~9294、11501~11994の拡張文字はWindows独自の割り当てなので異機種間の互換性がありません。

以上のほか、9301~11494のユーザ登録外字は、同一機種間でも互換性がありません。

【包摂文字】

JISには、「包摂」という考え方が取り入れられています。包摂とは「1つの文字から派生した異形字については同一の文字として扱う場合もありうる」を意味し、実際に数多くの漢字に適用されています。その結果、「コードは同じなのにフォントによっては字形が異なる」ということがしばしば起こります。例えば、「祈祷」の「禱」のフォントに大日本印刷の秀英太明朝体を指定すると、右

禱

図のように表示されます。これら包摂文字については、現在のところは入力時は無視していただくしかありません。

文書の保存

入力データの保存は、「テキスト・ファイル(プレーン・テキスト・ファイル)」と呼ばれるファイル形式で行います。テキスト・ファイルとは、「フォント(書体)や文字サイズ、行間、字送りなどの書式

指定」をしない、「JIS漢字コードで定められている文字と、ASCII規格で定められている欧文文字だけで構成された」最も単純な文書形式を指します。

ソフトによっては、保存用に複数のファイル・フォーマットが用意されています。例えば、Windows95に付属のWordpadでは 下図のような画面で「テキスト文書」を選択します。その他のソフトの場合には、「テキスト」「TEXT」などを選択してください。



Wordpadの保存画面

保存の際には、ファイル名は小文字のアルファベットを使用します。ファイルフォーマットで「テキスト」を選択して保存すると、大半のソフトでは.txtという拡張子が自動的に付きます。

入力者校正

プロの入力者、プロの校正者、プロの編集者の共同作業でつくられる市販の書籍でさえ誤植があります。アマチュアである私たちが入力したテキストには「確実にミス入力が含まれている」と見なすべきでしょう。

そこで、入力が完了したら、必ず「入力者校正」を行い、入力データを修正しておいてください。入力者校正は、後述する「原稿つきあわせ」と「素読み」の二段階に分けて行っていただくのがベストです。(校正の実際)

ファイルの精度を高めるためには、「できるだけ早い段階でミスを修

正しておく」ことが鉄則です。ぜひご協力ください。

文書の送信

入力者校正が済んだファイルは、aozora@voyager.co.jp宛にインターネット・メールで送信してください。その際、ファイルには「圧縮」をかけ、「添付ファイル」として処理するようにしてください。

【圧縮】

インターネットは、デスクトップでお使いのパソコンに比べ、処理速度が極度に遅いメディアです。ファイルを圧縮しておけば、送信時間が短縮でき、電話料金などのコストも軽減できます。

ファイルの圧縮には、専用のソフトが必要です。また、圧縮されたファイルの解凍（復元）にも専用のソフトが必要です。WindowsではWinZip（シェアウェア）を使って「ZIP形式」に圧縮するか、LHA（フリーウェア）を使って「LHA形式」に圧縮するのが一般的です（ZIP形式、LHA形式に圧縮するソフトは、この他にも多くの種類があります）。青空文庫の呼びかけ人にはMacintoshユーザーもおりますが、Macintoshではどちらの方式にも対応できます（ただし、LHAは「lh5」形式まで）ので、使い慣れた圧縮方式を使ってください。

【ファイル添付】

インターネット・メールで送信できるのは、「テキスト・データ」だけです。そこで、画像などのバイナリー・データを送信する際には、ファイルを「特殊なテキスト・ファイルに変換して、メールに添付する」必要があります。このような変換を「エンコード」、変換されたファイルを元に戻すことを「デコード」といいます。

入力済みのテキスト・ファイルは、エンコードなしでそのまま添付できますし、メールの本文にコピー＆ペーストで組み込むこともできます。しかしながら、メール本文に組み込まれたテキストには、1行1行に強制的に改行コードが付加されてしまいます。また、添

4 . 校正

付されたテキスト・ファイルは、メール・ソフトによっては、受信の際に自動的に本文に組み込まれてしまいます。圧縮 添付という方法で処理すれば、このようなトラブルを防ぐことができます。

エンコードは、Windowsの標準であるBase64を指定してください。(多くのメール・ソフトでは、特に何も指定しなくてもBase64が選択されます。)

この項でご紹介したフリーウェアやシェアウェアは、パソコン雑誌の付録CD-ROMで入手できます。また、最新版は下記のウェブサイトからダウンロードできます。

WinZip = <http://www.winzip.com/>

LHAおよび各種 = <http://www.forest.impress.co.jp/> (窓の杜)
<http://www.vector.co.jp/> (Vector)

【底本の用意】

送信していただいたテキスト・ファイルは、校正担当の呼びかけ人もしくは工作人員の方のところへとまわされ、校正作業に移ります。

このとき、入力に使用された底本が必要になります。

現在は、入手が容易であれば、同じ書籍の一番新しい版を購入するようにしています。しかしながら、ときには絶版・品切れのため、古書店などをまわっても入手が不可能な場合があります。そのため、底本もしくはそのコピーを送付していただく必要が出るかもしれません。ご協力をお願いします。

元原稿と入力データのプリント、あるいは一度チェックのすんだプリントと新しいプリントを比較対照しながら間違いを探していく作業を、出版の世界では「校正」と呼んでいます。

校正という作業

校正は、「入力したデータ」が正しいかどうかをチェックする作業です。作業にあたって尊重しなければならないのは、「文章」ではなく「文字」。ふつうに本を読むときは違って「読書の楽しみ」は捨て、「文字面だけを追う」のが基本です。極論すれば、文章の意味は理解できなくとも、正しい文字さえ把握できればよいのです。

【校正作業の流れ】

青空文庫で校正を担当していただく場合は、入力済みのテキスト・ファイルをプリントしたものと底本(たいていの場合はコピー)を郵送します。赤字(校正によって生じた修正)は、そのプリントに記入していただくこととなります。

なお、プリント校正だけでなくデータの修正まで担当していただける場合には、テキスト・ファイルをメール送信またはフロッピー送付します。

校正の実際

出版の世界には、長い歴史を通じて培われてきた「校正の基本」とでもいうべきものがあります。以下、青空文庫の校正作業で必要と思われるものを列記します。

【記入する色】

プリントへの記入は、必ず赤のボールペンなどで行うようにしてください。赤を使うのは目立たせるためです。

記入にあたっては、あとでご紹介する「校正記号表」(28ページ)に載っている記号をできるだけ使うようにしてください。同じ記号を

使うことによって、校正後のプリントを受け取った側が記入内容をより理解しやすくなるからです。

【書き込み方】

修正内容が赤で記入してあっても、行間に小さな文字で書かれていたりすると、受け取った側が見落としてしまう場合があります。間違っていてタイプされている箇所に消し線を引き、そこから欄外まで線を引いて、できるだけ大きく目立つように記入してください。

【原稿尊重が基本】

校正は「原稿（底本のコピーなど）尊重」が基本。入力結果が原稿に合っているかどうかをチェックするのであり、その逆ではありません。ただし、底本も印刷物ですからそれ自体に誤植や間違いが含まれている可能性があります。「これは原稿自体の間違いではないか」と思われるものについては、赤ではなく他の色や普通の鉛筆などで記入するようにしてください。これによって、「明らかな間違い」と「疑問点」が一目で区別できます。

【校正は繰り返し行う】

校正が一度だけで完結することはまずありません。以下のような異なった方法で繰り返し行くと、校正ミスの減少が期待できます。

- 1) 原稿つきあわせ = 原稿（底本）とプリントとを1字1句比較しながら、ミスタイプの箇所を指摘し、修正内容を記入していきます。
- 2) 素読み = 校正が済んだプリントを通し読みします。誤字脱字などの疑問が出た場合は、底本を参照して確認します。
- 3) 特定の観点からの素読み = 見出しの付け方に不揃いはないか、人名の表記に不揃いはないかなど、特定の観点からもう一度素読みします。特に、OCR入力をしたファイルの場合、「タ（カタカナ）」と「タ（漢字）」などよく似た文字の読みとりミスが頻繁に生じますので、特定の文字だけを探しながら見ていく必要が出てくる場合があります。

「観点読み」にあたっては、1ページごとにいくつかの観点を同時に見ていくことはやめ、「タとタ」なら全ページそれだけをチェックしていくことが大切です。複数の観点からの素読みが必要な場合には、それを繰り返します。

【再校の仕方】

入力者校正で多量の修正が出た場合は、上記の3度の校正作業のあとに修正作業を行い、再度プリントして、最初に赤字を記入したものの（初校といいます）と新たにプリントしたものの（再校といいます）との比較・確認が必要になります。

このとき、左に初校を、右に再校を置き、左手に青（赤以外の色）、右手に赤の筆記具を持ってチェック作業をし、チェックが終わったものは初校の赤字の上に「修正確認」を表す線を青（赤以外の色）で引いていくとよいでしょう。あとからざっと見直すだけで、チェック洩れを容易に発見することができます。

【振り返りが大切】

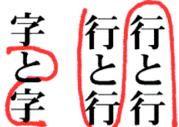
素読みの途中で、原著者の熟語の表記の勘違いなど、「パターンとして繰り返されやすいミス」が見つかることがあります。そのときには、そのまま先へ進むのではなく、いったん最初のページまでもどり、その漢字や用語について最初からもう一度チェックし直すようにしましょう。

【完了後は1日寝かす】

集中力には限りがあるもの。つきあわせ 素読み 観点読みを連続して行くと、かえって見落としが増える場合があります。校正が済んだプリントは、最低でも一晩寝かし、翌日最後の素読みを行いましょう。思わぬ発見があるはずです。

5 . 加工

主な校正記号

 <p>文字を削除して詰めるときは「トルツメ」と記入。削除したあとをそのまま空けるときは「トルアキ」。</p>	 <p>行の特定の部分から前を前行に送る。</p>
 <p>間違っただ修正を取り消し、原文のままにもどすときは、記入した赤字を消して「イキ(モトイキ)」。</p>	 <p>改行する。</p>
 <p>新たに語句を挿入するときに用いる。</p>	 <p>字と字、行と行を入れ替え。</p>
 <p>ひらがなやカタカナを拗促音に変更。</p>	 <p>あとの行を前に送って前行とつなげる。</p>
 <p>拗促音文字を普通の文字に変更。</p>	 <p>指定の位置まで文字を移動。引っ込める記号と位置指定の記号を併用。</p>
 <p>行の特定の部分からあとを次行に送る。</p>	 <p>は1字空にする、>は空きを詰める。位置指定記号を併用。</p>

登録への最後のステップは、ファイルの加工です。

青空文庫の基本フォーマット

青空文庫の登録ファイルは、「テキスト・ファイル、HTMLファイル、エキスパンドブック・ファイル」の3つのフォーマットを基本とします。これらのファイル形式へのとりまとめ作業は、原則として私たち呼びかけ人が担当します。

しかしながら、「HTMLファイルの知識なら自信がある」「エキスパンドブック制作のスキルなら誰にも負けない」という方もいらっしゃるでしょう。そう思われる方は、どうぞ積極的に名乗り出てください。

以下、3つのフォーマットについての処理の原則を簡単に説明しておきます。

フォーマットによる処理の違い

3つのフォーマットで処理が大きく異なるのは、「ルビ」と「外字」です。

【ルビ】

テキスト・ファイル=入力されたテキストどおりに「漢字《ルビ文字》」の形式で処理します。

HTMLファイル=ルビをくくる記号は全角の「()」を使用し、テキスト・ファイルから「《》 ()」に変更します。また、連続する漢字熟語の后者にのみルビがつく場合に用いた区切り線「|」は削除します。

エキスパンドブック・ファイル=エキスパンドブックのルビ機能を使って、印刷物と同じような形式の処理をします。

【外字】

テキスト・ファイル=入力されたテキストどおりに入力者注の形式

で処理します。

HTMLファイル=外字部分の入力者注は削除。一方、ファイルの末尾には、グラフィック・ツールで作成した外字画像の一覧表を置きます。そして、「 」から一覧表のタイトル部分へアンカーを使ったファイル内リンクを張ります。

エキスパンドブック・ファイル=エキスパンドブックのフォントサイズと同一サイズの画像をグラフィック・ツールで作成し、直接本文に埋め込みます。

【改行コードの確認】

Macintoshで入力されたテキスト・ファイルについては、改行コードが「CR+LF」(MS-DOS仕様)になっているかどうかを確認します。

【T-Timeへの対応】

(株)ポイジャーから発売されているテキスト・リーダーT-Timeには、HTMLファイルをルビ付きで表示する機能があります。ルビ表示したほうが読みやすい、理解がスムーズになると思われる作品については、今後は「T-Time対応」を原則としていきます。

その場合の書式は、タグ<!R>を使用し、

<!R>抽出(ひきだ)し

のようにします。

ファイル冒頭の注記

以上の処理原則については、読者の側で混乱が生じないようにするため、各ファイルの冒頭(作品名・著者名のすぐあと)に下記のような注記を入れます。

テキスト・ファイル

[表記について]本文中、底本のルビは「《ルビ》」の形式で処理した。また、JISにない外字は「 」で代用し、そのあとに[#注釈]の形式で字形などを示した。

HTMLファイル

[表記について]本文中、底本のルビは「(ルビ)」の形式で処理した。また、JISにない外字は「 」で代用し、字形についてはファイル末尾の外字一覧表中のグラフィックで示した。

エキスパンドブック・ファイル(目次もしくは奥付ページに入れる)

[表記について]本文中、JISにない外字は、すべてグラフィック化した文字で代用した。グラフィックを埋め込んだページは以下のとおり= 、 ページ。

PDFの扱い

異機種間に横たわる互換性の壁を越え、画像などのマルチメディア素材を含むレイアウト文書を共通して閲覧できるファイル形式として、Adobe社が提唱するPDF(Portable Document Format)が広く注目を集めています。

登録希望者・工作人員の方々からの要望があれば、同じAdobe社のAcrobat日本語版によるPDF化にも積極的に対応していきます。

TTZの扱い

テキスト・ファイルやHTMLファイルを縦組みやルビ付きで読むことができるT-Timeの機能は、製品版を購入したユーザーにしか活用できません。これに対し、同じ機能がより広い範囲で活用できるよう、フリーウェアT-Time Liteが開発・販売元の(株)ポイジャーからリリースされています。

T-Time Liteで製品版のT-Timeと同じ機能を活用するためには、対象となるファイルが専用の「TTZ形式」である必要があります。また、TTZ形式への加工には、(株)ポイジャーからライセンス提供される専用のツールが必要です。今後、青空文庫ではTTZ形式にも徐々に対応していきます。

本文中のURLおよびメールアドレスの変更について

本文中に使用されているURL、およびメールアドレスが、マニュアル作成時点から変更になっています。作業上の事情により、本文への反映は当面行いませんので、お手数ですが、次のとおり訂正の上、ご利用下さいますようお願いいたします。

3 ページ下から 3 行目「著作権が消滅した作家一覧」

<http://www.voyager.co.jp/aozora/siryoy1.html> (旧)

<http://www.aozora.gr.jp/siryoy1.html> (新)

4 ページ10行目「入力中の作品」

<http://www.voyager.co.jp/aozora/nyuryokutyu.html> (旧)

<http://www.aozora.gr.jp/nyuryokutyu.html> (新)

4 ページ13行目「電子テキストのある場所」

<http://www.voyager.co.jp/aozora/denshitext/denshitext.html> (旧)

<http://www.aozora.gr.jp/denshitext/denshitext.html> (新)

4 ページ19行目及び23ページ 4 行目 青空文庫のメールアドレス

aozora@voyager.co.jp (旧)

info@aozora.gr.jp (新)

(1999年8月 青空文庫)