

1 底本選び～作業の流れ

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、著作権の保護は継続します。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。

「著作権が生きている作品」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りです。

「著作権が消滅した作品」は、原則的に、自由に電子化して公開できます。どの作品を選ぶかは、入力にあたらうとする人の判断に委ねます。青空文庫の側で作品を割り当てることは、基本的に行いません。

底本を選ぶ

以上のことは、底本（入力の元にする本）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、「著作権が生きている作品」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、原著者の著作権に加えて、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作者の生死、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.aozora.gr.jp/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、作品の選定や配列などにあたった「编者（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。

著作権切れを控えた作家の作品は、「翌々年の 1 月 1 日までに公開できるもの」に限って、着手報告を受け入れます。どんな作家の権利切れが近いかは、「死せる作家の会」

<http://www.jca.apc.org/~earthian/aozora/dead.html> のリストを参照してください。

次に、青空文庫の「作家別本のリスト」、「作品別本のリスト」にあたって、入力したい作品が、すでに登録されていないか確認してください。青空文庫への登録を前提として、作業が進んでいる場合もあります。進行状況を示す「入力中の作品」<http://www.aozora.gr.jp/nyuryokutyu.html> にも、目を通してください。ここに記載されているものについては、原則として、新たに入力していただくことはできません。ただし「進行状態」が「入力取り消し」となっているものは、入力可能です。

旧字旧仮名で書かれた作品は、本来の姿にそった形（旧字旧仮名版）と現代表記に書き換えた形（新字新仮名版）加えて旧字のみを新字に置き換えた中間的な形（新字旧仮名版）でも登録しています。ある形式ですでに登録、着手されている作品を、別の形で、新たに入力していただくことは可能です。

「入力してみたいけれど、適当な作品に心当たりがない」という場合は、「随筆計画 2000」<http://www.jca.apc.org/~earthian/aozora/essay00.html> のページをご覧ください。このページは、作品社から刊行されている「日本の名随筆」シリーズを索引代わりに、著作権の切れている随筆作品をリストアップしたものです。青空文庫に登録済み、入力中、未入力の区別が、このページで確認できます。この中から、未入力

の作品を選ぶという方法もあります。ここに掲載されている作品はどれも短いものなので、入力に初めてという人でも、比較的楽に作業できるでしょう。

アンソロジー形式の短編集などには、長い作品の一部のみを抜き出した「抄録」形式で収められたものがあります。（「日本の名随筆」では、作品名が〔 〕で囲まれたものが、これに当たります。）また、一般には短編集やシリーズとしてひとまとまりで扱われているものの中から、1編のみを抜き出して収められている場合もあります。（「日本の名随筆」では、作品名の後に【××の一部】【××にあり】と書かれたものが、これに当たります。）青空文庫では、部分的な登録は避け、作品本来の姿での登録を心がけています。一部分のみが登録されていると、全体が登録されたときに重複する可能性もあります。入力したい作品が上記にあてはまる場合は、全体を収めた他の本を探してください。

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.aozora.gr.jp/denshitext/denshitext.html> をご覧ください。

他のサイトで公開されている作品であっても、あなたが青空文庫のファイル形式と取り扱い規準にそったものを用意したいと望まれるのなら、原則的には、取り組んでいただいてかまいません。ただし、公開済みであることが確認できている場合は、着手前に青空文庫宛にその旨を告知し、作業開始への了解を求めてください。

法令の入力は、原則として行いません。

青空文庫ではかつて、『大日本帝国憲法』『日本国憲法』を入力し、他の法律に関しても、リンク登録するなどしてきました。そんな中で、2001年4月1日、総務省による「法令データ提供システム」の運用が開始されました。その時点で施行されているすべての法令が、官報への掲載から1、2か月というわずかな遅れで、高い信頼性をもって登録され、検索の機能も充実しています。ファイルの利用にも、制限は設けられていません。こうした体制が整った以上、法令データは、同システムをはじめとするいくつかの専門サイトに委ねるのが妥当でしょう。よって、「廃

止されたもので、他では参照することが困難」といった特別な理由がある場合をのぞいて、青空文庫で法令を新規入力、公開することは行いません。

以上の手続きを踏んでも、見逃している要素があるかもしれません。また、青空文庫の「入力中の作品」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、着手前に必ず、青空文庫宛にご連絡ください。

編者や校訂者などの権利の扱いは、微妙です。これらの作業に当たった人が、底本に明記されているときは、連絡の際その旨を必ず書き添えて、進められるか否か、確認を求めてください。

同一作品の異なったバージョン（旧字旧仮名版、新字新仮名版、もしくは新字旧仮名版）に取り組む際は、どれが登録済みであり、どの形で入力を希望されるかを書き添えてください。

「著作権が消滅していて、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。

2．作業の流れ

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。

1.入力	底本の文章をそのままにテキスト入力していきます。
2.校正	入力されたテキストに入力ミスなどの誤りがないかチェックしていきます。
3.加工	完成したテキストを青空文庫用のファイル形式に仕上げていきます。

これらの作業は、ふつつ、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。

応募して入力や校正に取りかかったものの、作業を完了できそうになり見通しとなったら、遠慮なく info@aozora.gr.jp に申し出てください。「入力中の作品」の「進行状態」を「入力中」から「入力取り消し」へ、「校正中」から「校正待ち」へと戻し、新たな作業協力の申し入れを待ちます。

作業着手後、メールアドレスが変わった場合は、必ず info@aozora.gr.jp に連絡してください。

2 入力-1

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。青空文庫では、「文字のみを含むファイル」を作業の基本にしています。みなさんに入力していただくのは、この形式のファイルです。

「文字のみを含むファイル」のことを、一般に「テキスト・ファイル」と呼びます。Windows、Macintosh、UNIX、MS-DOS など、ほとんど全てのパソコンには、「テキスト・ファイル」を扱うことのできるソフトが組み込まれています。Windows ではメモ帳、Macintosh では SimpleText が代表的なものです。

それに対し、ワープロソフト等の独自の形式で保存したファイルは、そのソフトを使って開いたときに文字のみで表されていても、「テキスト・ファイル」ではありません。これらのファイルの多くには、人間が文字として読める情報の他に、そのソフトのみが解読できる情報が書き込まれています。従って、そのソフトがインストールされていないパソコンでは、正しく表示することができません。

ワープロソフト等の独自形式、画像、音声など、文字以外の情報を含むファイルを総称して、一般に「バイナリ・ファイル」と呼び、「テキスト・ファイル」と区別しています。ここでは詳しい説明は省きますが、「青空文庫では、どのパソコンでも扱えるテキスト・ファイルのみを入力作業の対象とする」ということを覚えておいてください。

青空文庫では、みなさんに作業していただいたテキスト・ファイルと共に、読みやすく加工したファイルを公開しています。「テキスト・ファイル」という呼び名が、これらのファイルと区別するために使われる場

合もありますので、注意してください。

写真や挿し絵などは、画像ファイルとして扱い、基本的に、入力の対象には含めません。加えて、本文の著作権とは別個に、写真撮影者や画家の著作権が存在する場合があります。

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。青空文庫は原則として、「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」という点です。

漢字の使い方や送りがなは、「底本どおり」とします。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も、底本にならって入力してください。

ただし旧漢字、旧かなのテキストが、私たちの多くにとって読みにくいこともまた事実です。現代表記にあらためたものがあれば、それを底本とすることで読みやすくできます。けれど、すべての作品で書き換え版が用意されているわけではありません。

そこで、旧字、旧かなの書きあらために関しては、作業を進める上での目安を定めた上で、例外的にこれを認めることとします。自分自身で書き換えを行おうと考える人は、「旧字、旧仮名で書かれた作品を、現代表記にあらためる際の作業指針」

<http://www.aozora.gr.jp/KOSAKU/genndaihyouki.html> にそって、対処してください。

「底本どおり」を目指しても、テキスト・ファイルではレイアウトを忠実に表現することはできません。字下げなどの組みに関する情報は、適宜、入力者注で表します。

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（WIN 版文書の保存、MAC 版文書の保存）。

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文字データに変換し、テキスト・ファイルとして保存するしくみを、OCR（Optical Character Reader）といいます。最近は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らなければなりません。

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。（底本の表題にルビが付いていたとしても、1行目に入れる作品名には、ルビはいりません。）
- 2）その次の行に「著者名」を入れる。
- 3）1行開けて、ルビなどを表す記号の使い方を説明した、【テキスト中に現れる記号について】を入れる。（記号を使わないですむ場合は、入れ

る必要はありません。)

4) 1行開けて、本文を入れる。

5) 本文の末尾から3行開けて、「使用した底本の正式名称・出版社・発行年月日、入力者名」等を「青空文庫収録ファイルへの記載事項」

<http://www.aozora.gr.jp/guide/kisai.html> の形式に沿って入れる。

6) 現代表記に書き改めた場合の作業履歴、その他の注記事項などがある場合は、「青空文庫収録ファイルへの記載事項」の形式に沿って入れる。

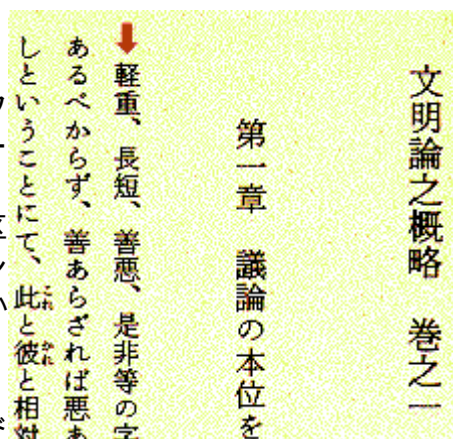
文庫本などではそれ自体に底本(全集など)が存在する場合があります。その場合には、「親本(底本の底本)の正式名称・出版社・発行年月日」もわかる範囲でつけ加えてください。

テキスト・ファイル(WIN 版文書の保存、 MAC 版文書の保存)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。数字と「-」は、英文モードで入力できる、半角のものを使ってください。



行数を数える際には、行アキはカウントしないでください。右の例では、赤矢印の行は「3」と数えます。

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) 使う文字

【パソコンで使える文字】

一般にパソコンでは、JIS X 0201 (「7ビット及び8ビットの情報交換用符号化文字集合」) で定義された、ラテン文字、数字、カタカナ、記号と、JIS X 0208 (「7ビット及び8ビットの2バイト情報交換用符号化漢字集合」) で定義された漢字、かな、ラテン文字、ギリシア文字、キリール文字、数字、記号が使えます。

JIS X 0201 の文字はすべて、JIS X 0208 に組み込まれています。

共通する部分がダブってしまっているわけですが、JIS X 0201 のものが「半角」で表示されるのに対し、JIS X 0208 の文字は「全角」で表示されるという形で、使い分けられています。

加えて、各社のパソコンでは、OS メーカー、パソコン・メーカーが独自に拡張した、いわゆる「機種依存文字」が使えるようになっています。

【青空文庫で使う文字】

青空文庫の入力では、JIS X 0201 の「半角カタカナ」と、各パソコンに組み込まれた「機種依存文字」は使いません。

「半角カタカナ」をのぞく JIS X 0201 の文字 (いわゆる半角英数と半角の記号) 加えて JIS X 0208 のすべての文字を使って、入力を進めていきます。

【包摂規準】

入力作業に取り組むと、底本の漢字の字体と、画面やプリントアウトのものの細部が微妙に違って、「このコードで入れて良いのか?」と迷うことがあります。(WIN 版微妙な字体差と包摂規準、 MAC 版

微妙な字体差と包摂規準)

青空文庫の入力ですべての文字を使うことのできる JIS X 0208 は、第 1 水準第 2 水準の文字を定義した上に、細かな字体差によって漢字のコードを区別するかしないかの約束事を、「包摂規準」という名前で定めています。

第 1 水準第 2 水準に追加して使う、第 3 水準第 4 水準の漢字などを定めた JIS X 0213 (「7 ビット及び 8 ビットの 2 バイト情報交換用符号化拡張漢字集合」) も、JIS X 0208 のものに追加、修正をほどこす形で、包摂規準を定めています。

青空文庫の漢字の入力は、JIS X 0208 に明示された 185 番までの包摂規準に加え、JIS X 0213 で追加された 186 ~ 199 番を受け入れた上で、底本に忠実に行います。(JIS X 0208 と 0213 規格票の包摂関連項目)

JIS X 0208 の第 1 水準第 2 水準にない漢字は、外字として注記します。
(WIN 版入力時に使える文字、 MAC 版入力時に使える文字)

(2) 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。

Windows には、MSIME というかな漢字変換ソフトが標準装備されています。MSIME では、設定によって、和文モードでもスペースが半角になる場合があります。入力にとりかかる前に、設定を確認しておきましょう。

1) MSIME のツールバーから「プロパティ」を選びます。



2) 「基本設定」タブ (MSIME98 以降は「全般」タブ) で、スペースの設定を「全角スペース」にしておいてください。



ツールバーが出てこない設定の場合は、「スタート」メニュー 「設定」 「コントロールパネル」 「キーボード」で「言語」タブを選ぶと、インストールされているかな漢字変換ソフトが表示されます。この中から選んでダブルクリックすると、「プロパティ」の画面になります。

【行頭の括弧の字下げ】

一部の書籍では、新しい段落の始まりに括弧 (「」 , 「(」 など) がくる際、1 字または半字下げである場合があります。(右の例では、1 字下げであります。) こうした場合にも、下のように、スペースは入れずに入力してください。

「平兵衛さんも世話好きさね。それにしても、あ
ようになったか。まあ、お父《とっ》さんともあ
もともと養父金兵衛は木曾谷での分限者《ぶげ、
田屋惣右衛門《ますだやそうえもん》父子の衣鉢

【字下げ】

テキスト・ファイルでは、字下げは形どおりに再現できません。底本に字下げした組みがあるときは、入力者注で表します。

字下げがはじまる 1 行前に、下の例のように [# ここから 字下げ] と入れ、行をあらためて、底本の文章を入力してください。「 字下げ」

220
「平兵衛さんも世話好きさね。
か。まあ、お父さんともあとで相
もともと養父金兵衛は木曾谷で
で、家では造り酒屋のほかに質屋

の「」の所は、全角のアラビア数字で入力してください。

すでに何字分下げるかは注記してあるのですから、あらためて字下げされた字数分、全角スペースを入れる必要はありません。

字下げが終わったら、下のように改行して [# ここで字下げ終わり] と入れてください。

字下げ箇所の前後で行が開けてあれば、下のように、入力テキスト中にも、空白の行を入れてください。

「ああ、風情《ふぜい》なものじゃの。」
能の狂言の小舞《こまい》の謡《うたい》に、

[# ここから 3 字下げ]

いたいけしたるものあり。張子《はりこ》の顔や、鐘稚児《ねりちご》。しゅくしゅく結びに、さき結び、やましな結びに風車《かざぐるま》。瓢箪に宿る山雀、胡桃《くるみ》にふける友鳥《ともどり》……

[# ここで字下げ終わり]

「いまはじめて相分《あいわか》った。——些少《ちと》じゃが餌《え》の料《りょう》を取らせよう。」

【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えてあることがよくあります。上の字下げの場合と同様に、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。

（３）外国語

【アルファベット】

JIS X 0208 にある文字を使って進めていく青空文庫の入力では、ラテン文字、ギリシア文字、（ロシア語などの表記に使われる）キリール文字の三種類のアルファベットが使えます。

このうちラテン文字については、一般のパソコンでは、半角のものと全角のものを使い分けられます。

青空文庫の入力でも、半角、全角双方のラテン文字を使います。

アルファベットの語句は、原則として、英文モードで半角のラテン文字を使い、

僕は Victor Hugo の Notre Dame を読んだとき、
のように入力してください。

半角ラテン文字の語間は、上のように、半角開けてください。

ギリシア文字、キリール文字は、和文モードの全角でしか入力できません。全角の文字を用いて、

そこには
とあった。
のように入力してください。

ギリシア文字、キリール文字の語間は、上のように、全角開けてください。

かな、漢字、記号（句読点と括弧をのぞく）とアルファベットの境は、

外道哲学と Sokrates 前の希臘哲学
説話には一々 moral が附く。

そこには
とあった。
のように、半角開けてください。

アルファベット文字列に組み込まれた、ピリオドやコンマ、中点、コロ、セミコロ、疑問符、感嘆符などの各種記号は、アルファベットとして扱います。

半角のものがあれば、半角で入力し、かな、漢字、記号（句読点と括弧をのぞく）と接するときは、

Il ne vivra pas ! と云った。

青空文庫の玄関口は <http://aozora.gr.jp/> です。

のように、半角開けてください。

句読点とアルファベットの境には、

西洋にいた時から、Archive とか

のように、アキを入れないでください。

日本語が入る括弧とアルファベットの境には、

そこには (苦は苦になら

ぬ) とあった。

のように、アキを入れないでください。

括弧の中にアルファベットが入る場合、アルファベットの前後には、

“ Kosinski soll leben ! ”

バヴァリア (Bavaria, Bayern) は、

苦は苦にならぬ ()

のように、アキを入れないでください。

アルファベットの入る括弧には、全角のものを使ってください。括弧の前後には、アキを入れないでください。

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は、和文モードで、全角のアルファベットを用いて入力してください。

縦組みの底本中で、縦に組まれているアルファベットは、和文モードで、全角のものを用いて入力してください。

全角で入力した、アルファベットの略号、縦アルファベットが、かな、漢字、記号と接するところには、

A T & T、N T T はともに、

a、b、c と黒板に

のように、アキを入れないでください。

【アクセント符号付きのラテン文字】

アクセサン (´) (`) や、ウムラウト (¨) などが付いたラテン文字には、異機種間の互換性がありません。アクセント符号なしのラテン文字を入力し、

フランスには raffine [# e はアクセサンテギュ (´) 付き] という語がある。

のように、どの字にどのアクセント符号が付いているのかを、入力者注で示してください。(WIN 版異機種間の互換性、MAC 版異機種間の互換性)

アクセント符号に関する注記は、以下のように、単語の後に付けてください。

刻み附けた *écriture* [# 1 文字目の e はアクセサンテギュ (´) 付き] *runique* であろう。

ラテン文字の語と注記の括弧 ([) のあいだには、上の例のように、アキを入れないでください。ラテン文字のうしろに入るアキは、注記の括弧 (]) のあとに置いてください。

アクセント符号の名称は、以下にならってください。

´ アクサンテギュ

` アクサングラフ

^ アクサンシルコンフレックス

¨ ウムラウト

(4) 数字と記号

【数字】

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

半角数字中の「,」は、半角で入力してください。

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】

「(」)」,「「」」などの括弧記号、「,」,「.」,「・」,「?」,「!」などの記述記号は、原則として、和文モードで入力してください。ただし、アルファベットとアラビア数字の、半角文字列中で用いられる場合には、英文モードで半角のものを入力してください。

かな漢字変換ソフト（MSIME、ATOK など）の設定によっては、句読点などが自動的に半角や「,」「.」に変換されてしまうことがあります。入力にとりかかる前に、設定を確認しておきましょう。

MSIME、ATOK 共に、句読点の種類（「。」「か」「.」「,」か）は、底本の句読点に合わせてください。ツールバーから、MSIME は「プロパティ」を、ATOK は「メニュー」「プロパティ」を選べると、設定画面になります。

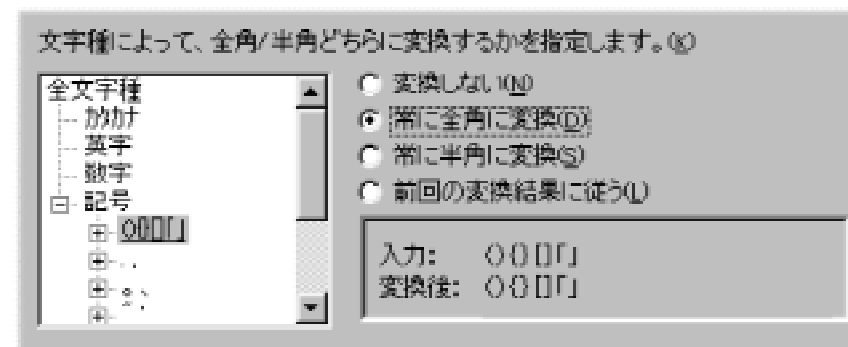
MSIME の設定画面



ATOK の設定画面



また、MSIME の初期設定では、記号の半角 / 全角変換が「前回の変換に従う」となっていることがあります。「全角・半角」タブ（MSIME98以降は「オートコレクト」タブ）で、句読点、括弧の設定を「常に全角」にしておいてください。



【特殊記号】

Windows の MSIME や ATOK、Macintosh のことえりや ATOK の文字パレット（WIN 版文字パレットを使う方法、MAC 版文字パレットを使う方法）にある「ローマ数字」などの特殊文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。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。（WIN 版異機種間の互換性、MAC 版異機種間の互換性）

【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。

Windows の MSIME や ATOK、Macintosh のことえりや ATOK の文字パレットの「単位」の項目には全角 1 文字の単位記号が用意されています。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。（WIN 版異機種間の互換性、MAC 版異機種間の互換性）

【ハイフンとダッシュ】

半角のアルファベット、数字文字列中で使われているハイフンには、半角の「-」を使ってください。

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には、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、中点 (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。

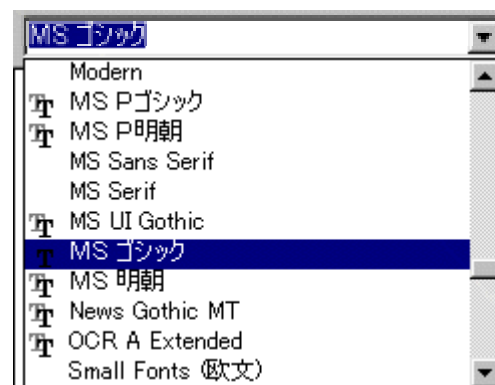
【全角と半角】

プロポーショナルフォント (文字の形によって、表示される文字の幅が変化するフォント) をお使いの場合、入力中の画面では、数字などの全角と半角との区別がつきにくいことがあります。特に、Windows では、Macintosh に比べて、この傾向が強いようです。

「画面では半角に見えているようなのに、全角になっていた」、あるいはその逆のケースが多発する場合は、お使いのワープロソフトやテキストエディタの「フォント設定」を、「等幅フォント」に変えてみると良いかもしれません。「等幅フォント」では、全角と半角の文字幅は、常に 2 対 1 で表示されます。

Windows では、「MS Pゴシック」「MS P明朝」など、フォント名に「P」がつくものがプロポーショナルフォント、「MS ゴシック」「MS 明朝」など、「P」がつかないものが等幅フォントです。既定値の設

定では、「MS Pゴシック」となっているソフトがほとんどです。



(5) 特殊な表記

【ルビ】

ルビ処理の基本

ルビは、ルビの付く文字列のあとに、「《》」(区点コード 0152・0153、JIS コード 2154・2155、シフト JIS コード 8173・8174) でくくって入力します。(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。)

印刷物では通常、「がっこう」、「きょうしつ」のように、ルビの拗促音が小書きされません。けれど青空文庫のファイルでは、「学校《がっこう》」、「教室《きょうしつ》」などと、ルビの拗促音を小書きしてください。ただし、拗促音を小書きしない旧かな作品では、ルビにも小書きを用いないでください。

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。

分けるべきか、まとめるべきか迷ったときは、まとめてください。

ルビの付く文字列が始まる前には、「|」(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。)を入れます。

書き込まれた「|」は、「《》」におさめられたルビを、プログラムを使って底本にあるような形に再現しようとする際に、処理の手がかりとして利用できます。

底本に次のようにあれば、

武州^{おうめ}青梅の宿

武州 | 青梅《おうめ》の宿
と入力してください。

「|」を省く場合

ルビのかかる文字列の区切りにならず「|」が入っていると、自動処理には便利でも、テキストを読む人にはわずらわしいでしょう。

そこで、「|」なしでもルビのかかる範囲を特定できる場合には、「|」を省きます。

ルビのかかる部分が漢字だけで構成され、かな、記号、空白など、「漢字以外の文字」に続いている場合には、「漢字」と「漢字以外」という文字種の違いを手がかりに、ルビのかかる範囲を特定できます。

耳^{ほて}まで火照って

は、ルビのかかる「火照」がかなに続いているから、「|」を省いて、

耳まで火照《ほて》って
と入力します。

ルビは多くの場合、ひらがなや句読点などに続く漢字文字列に付いています。そのためファイル中では、ほとんど、

夢と現《うつつ》との闘《しきい》はない
帆をたて直し、艦《ろ》を押して

などと、「|」なしで入力することになります。

繰り返し記号「々」は、漢字として扱います。

すると^{やや}少々度を失った

とあれば、「|」を省いて、

すると少々《やや》度を失った
と入力します。

将来、入力に用いる文字を JIS X 0213 の範囲に拡張した際は、「々」(二の字点、面区点番号 1-2-22)も、「々」と同様に漢字として扱います。現時点で二の字点は、外字注記することになりますが、その際は、

确实さで、益 [#二の字点、面区点番号 1-2-22、33-3]《ますます》はっきりと
のように書いてください。

アルファベットの単語は、【アルファベット】の項にあるように、原則として前後に半角のアキを入れて入力します。

アルファベットの単語にルビが付いている場合は、空白(まれに句読点、括弧)との文字種の違いによって、ルビのかかる範囲を特定できます。

兄^{ファナチケル}のような Fanatiker とは

とあれば、「Fanatiker」の前後を半角開けますから、「|」を省いて、

兄^{ファナチケル}のような Fanatiker 《ファナチケル》 とは
と入力します。

この際、単語のあとにくる半角アキは、「《」の前ではなく、「》」のあとに置きます。

アルファベットの句や文にルビが付く場合は、単語ごとにルビを付けます。

“^{コジンスキイ} Kosinski ^{ゾル} soll ^{レエベン} leben !”

とあれば、

“ Kosinski 《コジンスキイ》 soll 《ゾル》 leben 《レエベン》 !”
と入力します。

アルファベットには、英文モードで入力された半角のものと、和文モードで入力された全角のものがありますが、ともに上のように処理してください。

「 | 」を省かない場合

「 | 」なしでは、ルビのかかる範囲の特定が難しい次のような場合には、「 | 」を用います。

冒頭の例のように、漢字文字列の途中からルビが付く場合には、文字種の違いを手がかりにできません。

そこで、「 | 」を用いてルビのかかる範囲を特定し、

武州 | 青梅《おうめ》の宿
と入力します。

ルビの付く文字列に、漢字とかななど、異なった文字種が混在している場合にも、文字種の違いを手がかりにできません。

フエヤリー・ストーリーズ
そんなお伽話は、

とあれば、「 | 」を用いて、

そんな | お伽話《フエヤリー・ストーリーズ》は、
と入力します。

スコットランドヤード
霧のロンドン警視庁…

とあれば、

霧の | ロンドン警視庁《スコットランドヤード》…
と入力します。

かま え もん
いいか釜右エ門。

の「エ」は、カタカナです。「 | 」を用いて、
いいか | 釜右エ門《かまえもん》。

と入力します。

複数のアルファベットの単語に、一つのまとまったルビが付く場合には、「 | 」を用いてルビのかかる範囲を特定してください。

とらば
彼は Au revoir と、

とあれば、「彼は」のあとに「 | 」を入れて、

彼は | Au revoir 《さらば》 と、
と入力します。

アルファベット入力時に入れる半角アキは、上の例では「 | 」と「 Au」のあいだと、「 》」 と「 と」のあいだに置きます。

ルビ処理の応用

ルビ処理のルールを確立し、これを遵守していけば、プログラムによるルビ付けの自動処理や、洗練された音声読み上げなどへの道が開けます。

「テキストビューワー」<http://www.izu.co.jp/~at-sushi/aozora/viewer.html> のページにあるように、すでに青空文庫形式のルビ記号に対応したソフトが、いくつも開発されています。

【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、入力者注で処理してください。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。(入力者注の形式)

傍点には「\」だけでなく、「\」_丸「」_丸「」_丸「」_丸「」_丸「」_丸「●」_丸などが使われることもあります。

「\」が使われている場合は、「傍点」の代わりに「白ゴマ傍点」と入力してください。「」_丸は「丸傍点」_丸「」_丸は「白丸傍点」_丸「」_丸は「黒三

角傍点、**「」**は「白三角傍点」、**「」**は「二重丸傍点」、**「●」**は「蛇の目傍点」と書いてください。

【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、青空文庫で使える文字には含まれていません。代わりに「/ \」（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「/ \」を入力します。

印刷物では、「々」「>」などが行頭にくる場合には、「人々 人々」のように、踊り字を一字目の漢字に置き換えるのが普通です。この処理が行われたと判断できるときは、二つ目の漢字を踊り字に戻し、「人々」といった形で入力してください。

達雄はいらくした。

【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てください。

【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、入力者注で処理してください。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】

国家による検閲が盛んに行われた時代に発行された書籍では、問題があるとされた箇所を「×」などで置き換えた、伏せ字が使われている場合があります。伏せ字は、底本で使われている記号を、使われている個数分入力してください。

伏せ字があった場合は、ファイルの冒頭に置く【テキスト中に現れる記号について】の最後に、以下のように項目を追加してください。

×：伏せ字

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です。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と思われる場合は、

宮沢賢治 [# 「賢治」は底本では「憲治」と誤植]

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。
誤植と断定することはできないけれど、表記が疑わしい場合、もしくは、明らかに誤りではあるけれど、正しい内容が特定できない場合には、

彼は 12381 [# 「12381」はママ] 年に生まれ、
のように、「ママ」を用いて注記してください。

【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。

3 入力-2 (Win)

漢字を探す

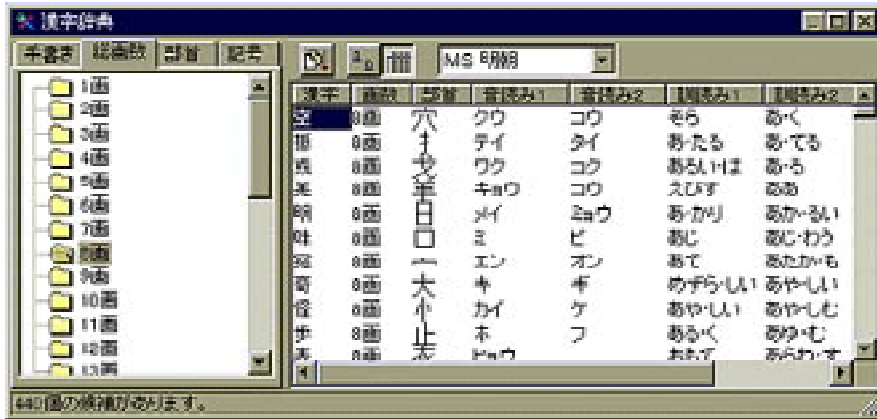
テキストを入力していて、必ず悩まされる問題があります。それは目的の漢字が、かな漢字変換ソフト（フロント・エンド・プロセッサ、略して FEP ともいいます。Windows では標準装備の「MSIME」、ジャストシステム社の「ATOK」が代表的です）の候補に出てこないことです。

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しても、「ATOK14」までで表示されるのは「入洛」、「聚楽」などだけです。これは「聚」と「落」という2つの漢字がともにパソコンで使えるにもかかわらず、「聚落という熟語」が「ATOK14 までの辞書には登録されていない」ことを示します。

かな漢字変換ソフトから呼び出せない漢字の探し方は、大きく分けて三つあります。仮名漢字変換ソフトの文字パレットと呼ばれるツールを使う方法。参考図書などで調べた文字のコードで入力する方法。そして、青空文庫が独自に用意した「新 JIS 漢字総合索引」を使う方法です。

【文字パレットを使う方法】

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。MSIME2000 の場合、タブがプルダウンメニューに変わっていますが、基本的な使い方は同じです。

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。ATOK10、ATOK13 など、他のバージョンの ATOK にも、文字パレットは用意されています。

文字パレットの詳しい使い方は、かな漢字変換ソフトのマニュアルを参照してください。

【コードで入力する方法】

『JIS 漢字字典』（日本規格協会）や、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』といったものを使うと、探したい文字のコードを調べられます。『新版漢語林』（大修館書店）など、一部の漢和辞典も、JIS 漢字コードのコードを併記するようになっています。

こうした参考図書ではしばしば、一つの漢字に対し、「区点コード」「JIS コード」「シフト JIS コード」など、いくつかのコードが示してあります。これらは、コード表中で文字が占める位置を、どのようなルールに従って表現するかの違いによるものです。JIS 漢字コードという一つのものを表現する、三つの方法と理解してください。

MSIME の場合

- 1) 日本語入力をオンにして、下左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。（下中右図）



ATOK の場合

1) 日本語入力をオンにすると、下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、下図のようにシフト 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。



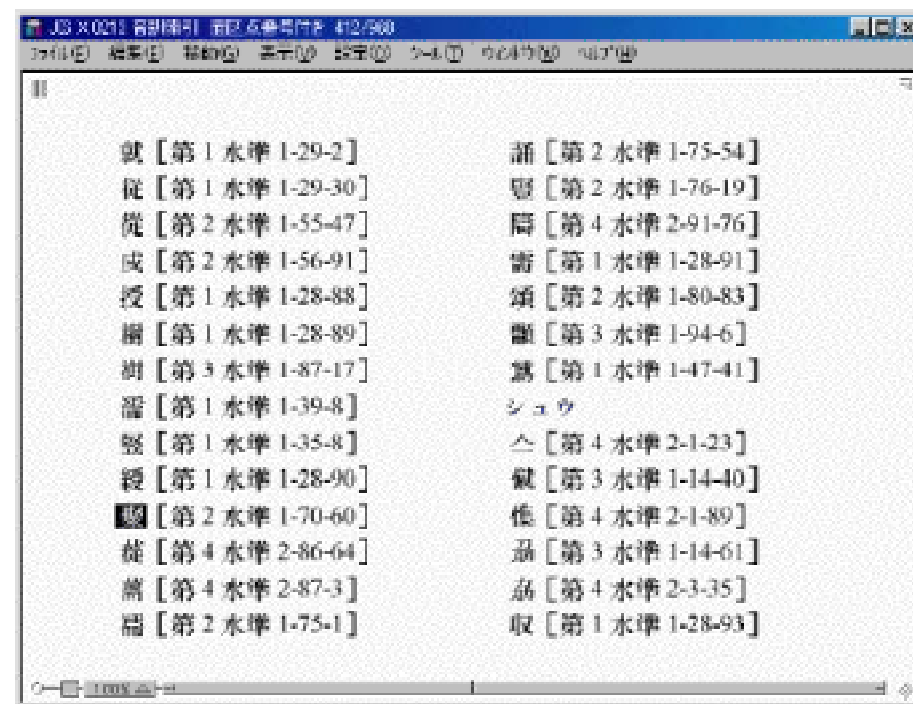
3) 目的の文字のシフト JIS コード (区点コード、JIS コード) を入力します。これで、目的の文字が現れます。現れた文字をクリックすれば、カーソルのある位置に入力されます。

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。

【新 JIS 漢字総合索引の利用】



新 JIS 漢字総合索引は、「部首・画数索引」「音訓索引」と、青空文庫の入力でよく使う記号を集めた「記号一覧」からなる、電子索引です。部首・画数索引を使えば、部首と画数から、音訓索引を使えば漢字の読みから、目指す漢字を探せます。(右は、「じゅ」という読みから、音訓索引を使って「聚」を探した例。)

見つかった漢字は、コピーしてエディターやワープロにペーストでき

ます。

この索引を使うためには、いくつかのソフトとフォント、ファイルが必要です。求められるものを、「青空文庫」「明日の硯箱」
<http://sumomo.sakura.ne.jp/%7Eaozora/jisx0213/0213tools/>にリストアップしました。導入時に少し手間がかかりますが、使うことのできる文字を確実に探せます。必要なフォントやファイルは全て無料で入手できますので、ゆっくりステップを踏んで、是非、索引の導入にチャレンジしてください。

新 JIS 漢字総合索引には、第 1 ～ 第 4 水準までの漢字を収録してありますが、青空文庫の入力に使えるのは、第 1 第 2 水準の文字に限られます。第 3 第 4 水準の漢字は、使わないでください。

新 JIS 漢字総合索引の成り立ちと使い方は、【「新 JIS 漢字総合索引」を使ってみよう】で詳しく説明しています。この解説を含む、「新 JIS 漢字時代の扉を開こう！」
<http://www.aozora.gr.jp/newJIS-Kanji/newJIS1.html>と名付けた文書では、JIS 漢字コードに関して、青空文庫の活動を続けていく上で必要な情報をまとめて解説しています。

【入力時に使える文字】

青空文庫の入力は、JIS X 0201 で定義され、半角で表示されるラテン文字、数字、記号と、JIS X 0208 で定義され、全角で表示される漢字(第 1 水準と第 2 水準) かな、ラテン文字、ギリシア文字、キリール文字、数字、記号によって行います。

JIS X 0208 (JIS X 0201 の文字は、ここにすべて含まれます。) がない文字を、外字と呼びます。

Windows や Macintosh では、JIS X 0208 のコード表の空き領域に、独自に文字を割り振って、外字の一部を使えるようにしています。けれど、メーカー独自の拡張によって使えるようになった外字には、他の機種で使えるという保証がありません。

独自拡張によって使えるようになった外字は、青空文庫では使いません。

半角で表示される、JIS X 0201 のカタカナも使いません。

【外字の処理】

外字は、入力者注で処理してください。例えば、

喉を掻き撓って

の「撓」は、青空文庫で使える第 1 第 2 水準の中にはありません。こうした場合には、

喉を掻き [# 「てへん + 劣」読みは「むし」 30-16] って
のように、代わりに「」(区点コード 0208、JIS コード 2228、シフト JIS コード 81A6) を入れ、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。

外字の形を説明する際、左右に並んでいる部分は、「+」(区点コード 0160、JIS コード 215C、シフト JIS コード 817B) でつないでください。「たれ」や「によう」との組み合わせにも、「+」を用いてください。

「目 + 争」「登 + おおざと」「やまいだれ + 音」「しんにゅう + 台」など

上下の関係にあるものは、「/」(区点コード 0131、JIS コード 213F、シフト JIS コード 815E) でつないでください。

「くさかんむり / 弓」「白 / 十」「山 / 松」など

上下、左右の関係にあるものを組み合わせて示す際は、「()」(区点コード 0142・0143、JIS コード 214A・214B、シフト JIS コード 8169・816A) を補って、位置関係をはっきりさせてください。

森

「水 / (水 + 水)」

醫

「(医 + 殳) / 巫」

嶺

「山 + (老 / 日)」

等

「+」や「/」で表しにくい場合には、以下のように説明してもかまいません。

[# 「糾」の「糸」に代えて「虫」読みは「きゅう」 145-1]

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、

喉を掻き^{むし}撓って

となっている場合には、

喉を掻き [# 「てへん+劣」 30-16] 《むし》 っ
てのように入力します。

外字注記はかつて、以下のように書いていました。

喉を掻き [# 「 」は「てへん+劣」 読みは「むし」 30-16]
っ
過去のファイルにさかのぼって、下線部を略した現在の形にそろえるこ
とは当面行いませんが、より簡潔な現在の形を推奨します。

【異機種間の互換性】

JIS X 0208 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 で入力した修飾数字
（MSIME の特殊文字、ATOK の丸数字）の「 付きの 1 （区点コード 1301、
JIS コード 2D21、シフト JIS コード 8740）は、Macintosh では「（日）」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 X 0208 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
機種間の互換性はありません。

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はありません。

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
性はありません。

【微妙な字体差と包摂規準】

底本で使われている漢字と入力し
たもので、細部の形が微妙に異な
っていることがあります。「果たして
この字を使ってよいのか？」と、迷
う場合もあるでしょう。その際、「よ
いか悪いか」の判断は、用いている
漢字コードが「字体の揺れ幅」につ
いて定めた、「包摂規準」（1 ～ 185
番）にのっとって下します。青空文
庫は JIS X 0208 を使っていますから、
その規準に従って、判断します。（加
えて、JIS X 0213 で新たに定義され
た 186 ～ 199 番にも従います。）

包摂規準とは、「この差ではコード
を区別しない」という約束事の集まり
です。ルールのすべては、JIS 漢字
コードの規格書（『7 ビット及び 8 ビ
ットの 2 バイト情報交換用符号化漢
字集合』日本規格協会）の、「漢字の
字体の包摂規準」（11 ～ 23 ページ）
に定められています。（「JIS X 0208
と 0213 規格票の包摂関連項目」
http://www.aozora.gr.jp/hosetsu_kijyun/index.html の「JIS X 0208」の項、参照。）

ここに定められた包摂規準に従えば、本来、右に示す 29 文字の A と B
のパターンは区別されなければいけません。ところが過去の規格改訂時
に犯した誤りとつじつまを合わせるために、例外として「同じコードで
入力する」とされています。これらに関しては、底本と表示文字に A と
B の違いがあったとしても、その差は無視してください。

A	B	A	B
啞	啞	箆	箆
焰	焰	搦	搦
鷗	鷗	填	填
嚙	嚙	顛	顛
俠	俠	禱	禱
軀	軀	浼	浼
齟	齟	囊	囊
翹	翹	澆	澆
屢	屢	醜	醜
繡	繡	頰	頰
蔣	蔣	麵	麵
醬	醬	萊	萊
蟬	蟬	蟬	蟬
搔	搔	攢	攢
驪	驪		

JIS X 0208 の包摂規準では、しんにゅうの点が一つか二つか、草冠の横画が切れているか否かで、コードは区別されません。画面やプリントアウトの表示と底本にこの差があったとしても、違いは無視してください。

細かな字体差を無視して入力しても良いのか、それとも外字として注記すべきか判断を迫られたときは、JIS X 0208 の「漢字の字体の包摂規準」を参照してください。見きわめが付かなければ、info@aozora.gr.jp に問い合わせてください。

【旧字入力の支援ツール】

旧字作品の入力ファイルには、使うべきでない新字や俗字が、しばしば紛れ込んでしまいます。

これらを見つけ出すために、「校閲君」と名付けたチェック・ツールを用意しました。詳しい使い方は、「旧字ファイルの新字・俗字を、校閲君で洗い出そう！」<http://www.aozora.gr.jp/tools/kouetsukun/>で説明しています。

旧字作品の入力が終わったら、青空文庫宛にファイルを送る前に、必ず校閲君でチェックしてください。

作業の途中で校閲君によるチェックをかければ、どんなものが新字、俗字になりがちか確認できます。その後の作業を正確に進める上では、有効でしょう。

かつて旧字作品の入力、校正を体験された方は、自分の関わったファイルを校閲君でチェックし直してください。

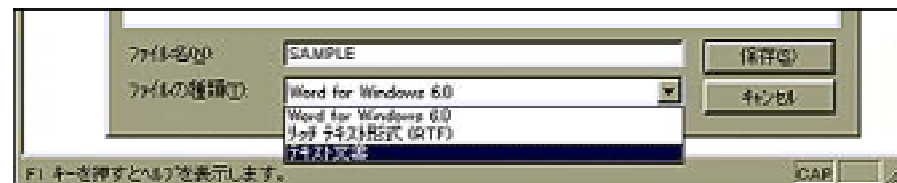
旧字、正字に置き換えるべきものが見つかったら、info@aozora.gr.jp に、ファイルの修正を求めてください。

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 X 0208 で定められている文字と、JIS X 0201 で定められてい

る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。

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つきます。

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「確実に誤りが含まれている」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ください。（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、info@aozora.gr.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、パソコンに比べ、処理速度の遅いメディアです。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。

入力ずみのテキスト・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。

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、当然ながら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。

4 校正

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とは違って「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、原則として、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、プリントアウトに記入してください。

作業が終わったら、プリントアウトと底本を共に、返送してください。宛先は、送付時にお知らせします。

青空文庫の作業に慣れた方には、入力済みのテキスト・ファイルと底本をお送りする形も、選んでいただけます。

テキスト・ファイルをお送りする際は、校正に加えて、ファイルの修正も行ってください。さらに、「どこをどう直したか」を示す、作業履歴ファイルも、別に作ってください。作業履歴の形式は特に定めません。以下のように、何をどう直したかがリストアップされていれば、それで結構です。

「嗚呼」と声が漏れた。 「嗚呼」と声が漏れた。

意味のない抵抗を 意味のない抵抗を

テキスト・ファイルによる校正が終了したら、圧縮をかけた修正済みテキスト・ファイルと作業履歴ファイルを、info@aozora.gr.jp 宛に送っ

てください。加えて、指定された宛先に、底本を返送してください。

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、あとで紹介する「主な校正記号」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。間違ってタイプされている箇所には消し線を引き、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あることがあります。

「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(初校といいます)と新たにプリントしたもの(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】

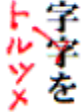




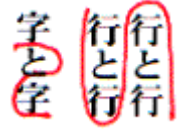
素読みの途中で、原著者の熟語の表記の勘違いなど、「パターンとして繰り返されやすいミス」が見つかることがあります。そのときには、そのまま先へ進むのではなく、いったん最初のページまでもどり、その漢字や用語について最初からもう一度チェックし直すようにしましょう。




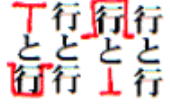

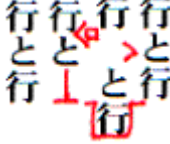
【完了後は1日寝かす】

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。

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。 削除してあとを 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>間違った修正を 取り消し、原文 のままにするこ 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 入するとき用 いる。</p>		<p>字と字、行と行 を入れ替える。</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字空きに するとき、>は 空きを詰めると きに。位置指定 記号を併用。</p>

5 ファイルの加工

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、「《》（ ）」に変更します。また、ルビの付く文字とつかない文字の境を示す区切り線「|」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に仕立てます。

【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＝外字部分には「 」を入れ、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。さらに、目次もしくは奥付ページの余白などに「外字グラフィックが使われているページ」を列記します。

【T-Time への対応】

（株）ボ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。ルビ付きの作品の HTML ファイルは、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。

【ファイルの作成】

テキスト・ファイルと HTML ファイルは、エディタやワープロでつくれます。HTML ファイル作成用には、さまざまなツールも用意されています。

エキスパンドブックづくりには、（株）ボイジャーが販売している開発ツール、エキスパンドブックが必要です。ブック・ファイルは、無償で利用できるエキスパンドブック・ブラウザで開けます。

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、Adobe 社の PDF（Portable Document Format）が、広く利用されています。

PDF ファイルは、Adobe 社が販売している Acrobat 日本語版によってつくることができ、Adobe Acrobat Reader 日本語版（無償）で開けます。

プリントアウトでは、PDF の特長がいけると判断しながらも、ファイル作成の負担を抑えたいと考えて、呼びかけ人はこれまで、この形式の

ファイルをつくってきませんでした。ただし、著作権の有効な作品で、登録希望者によってあらかじめ用意されている場合には、PDF も受け入れています。

TTZ、ドットブックの扱い

TTZ とは、T-Time の専用ファイル形式です。エキスパンドブックと同様、縦組みルビ付きで表示されたページを、めくりながら読める上に、ウインドウや文字の大きさも自由に変えられます。圧縮して保存されるので、ファイルサイズは小さくおさまります。

TTZ は、(株)ボージャーが販売している T-Time パブリッシャーズキットでつくることができ、無償で利用できる T-Time 機能限定版で開けます。(T-Time を使うためには、加えて、アップルコンピュータ(株)の QuickTime3 以上(無償)が必要です。)

青空文庫では、第3第4水準に対応した「明日の本棚」
<http://www.sumomo.sakura.ne.jp/~aozora/jisx0213/>で、TTZ を利用しています。

ドットブックは、ウェブブラウザで開く、特殊な形式の TTZ です。
(拡張子には TTZ 用の .ttz ではなく、.book が用いられます。)

ドットブックをつくるには、(株)ボージャーとのライセンス契約が必要です。ドットブックのファイルは、QuickTime3 以上、T-Time 機能限定版に加え、T-Time Plug (無償)があれば開けます。

青空文庫では、短いものの多い「随筆計画 2000」の作品を、ドットブックに仕立てています。

ご不明の点がありましたら、青空文庫 (info@aozora.gr.jp) 宛にご連絡ください。

青空文庫工作員 作業マニュアル Windows 版

1998 年 12 月作成

2001 年 11 月 1 日改訂

2001 年 11 月 26 日最終修正