資料庫作業手順

内容

[1. 資料庫の在り処 1](#_Toc94547621)

[2. 資料庫の基本的構造 1](#_Toc94547622)

[3. HTMLの基本的構造 2](#_Toc94547623)

[4. 具体的作業手順 4](#_Toc94547624)

[5. 具体的分担 5](#_Toc94547625)

# 資料庫の在り処

資料庫の保存場所：

削除

パスワード：

削除

クラウド上ではHTMLファイルが正常に作用しないので、作業用に自己のパソコンにダウンロードする。また、ネット上同時に共同作業ができるように、「Liya-shakumon.html」（後述）は編集用に一時的にdocx形式に変換しているが、訳注稿の移転作業が終了すると、そのままHTML形式に戻す。

# 資料庫の基本的構造

* 資料庫の中心ファイルは釈文ファイルである（「Liya-shakumon.html」）（現在は一時的に「Liya-shakumon.docx」に変換）
* 釈文ファイルから、図版ファイル・読み下し文ファイル・釈読ファイル・語釈ファイルという三種の関連ファイルへとリンクが貼られており、釈文ファイルから各種情報にアクセスできるようになっている。関連ファイルは次の三つのフォルダーに格納されている。
	+ Liya-zuhan
	+ Liya-yomikudashi
	+ Liya-shakudoku
	+ Goshaku
* 図版ファイル・読み下し文ファイル・釈読ファイルは、簡番号をファイル名としており、簡番号を通じて釈文ファイルとリンクされている。
* 語釈は、語を単位に作られており、ファイル名は語の日本語音を参照して作られているが、格納フォルダーは、人名・地名・官職名・身分呼称・労働内容という内容に応じて、次のようなサブフォルダに細分化れている。
	+ jinmei
	+ chimei
	+ kanshoku
	+ mibunkosho
	+ rodounaiyo
* 上記のファイルとフォルダーは、資料庫の以下のような基本構造を反映する。



今後は他の資料群も追加することで次のような構造を目指す。



* 資料庫には上記以外、次の四つのフォルダーが設けられている。
	+ SpryAssets（釈文ファイルの見出しを制御するファイルを格納、HTMLの基本構造を参照。今後はできればもう少し使いやすいナビゲーションを構築したいと考えるが、ここを変更することになる）
	+ Styles（釈文ファイルおよび関連ファイルの書式などを制御するファイルを格納、HTMLの基本構造を参照）
	+ Gaiji（画像データとしてユニコード以外の字形を蓄積。アクセス方法などは「里耶訳注稿外字一覧表.docx」を参照）
	+ Liya-gazo（外字以外で、釈文ファイルや関連ファイルに組み込む画像を蓄積。アクセス方法はGaijiを参照）
	+ ツール（Perlスクリプト等のツールを格納）→削除
	+ 使用済み作業ファイル（構築に使ったPerlスクリプト等の古い使用済みバージョン）→削除
* 資料庫には、上記以外次のファイルとショートカットが格納されている
	+ 訳注稿11（20220131）.docx（訳注稿の1月末時点の最新ファイル。釈文ファイル作成開始以降の変更には変更履歴が附せられている
	+ 作業用ファイル訳注稿11.docx（読み下し文ファイルや釈読ファイルへの関連情報の格納に用いる。情報格納後は、履歴を残さず、作業用ファイルから情報を削除する）
	+ 資料庫構築メモ.txt（資料構築に際して作った覚書）→削除
	+ 資料庫作業手順.docx（本ファイル）
	+ rekijitubangoichiran.html/rekijitujunichiran.html（暦日の一覧表、番号順と暦日順の二種。釈文ファイルにはリンク貼り済み。本来はGoshakuフォルダーにサブフォルダーを設けて格納すべし。今後変更する予定）
	+ Command Promptショートカット（Perlスクリプトを実行するため）（OneDrive上は作動しないため削除してあるが、今後の作業のため、パソコン上は設けた方が便利）

# HTMLの基本的構造

* HTMLファイル自体は、「<abc>」「</abc>」という書式のマーカーを付けたテキストとなっている。ブラウザーで開くと、テキストのみ、マーカーに従って書式を付けて表示され、EmEditorでは、マーカーもテキスト情報として表示されるが、テキストが黒色なのに対し、マーカーは諸種の色で区別して表示される。「<abc>」はマークしたいテキストの始まり、「</abc>」はその終わりを示す。
* マーカーには、HTMLという機械言語によって機能・書式が定められているものもあれば、陶安が適宜作ったものもある。マーカーの定義は、stylesフォルダに格納されているcss形式のファイル（「Ejina cs5.css」）で行われている。自作マーカーのみならず、HTML所与マーカーに対しても、cssファイルで書式を定義・変更することができる。
* cssファイルに関する補足説明（とりあえずは読み飛ばしても差し支えない）

cssファイルでは直接にマーカーに対して書式を指定することもできれば、マーカーに限定的な変更を加えるクラス（class）や識別子（identifier）を定義することもできるが、表を除けば、できるだけクラスや識別子を使わない。（cssファイルの基本形式は、セレクター（selector）に対して、書式などの指定を行う形になっているが、HTMLファイルのマーカーに対して定義を行う場合には、マーカー（「<abc>」）の文字部分（「abc」）がセレクターになり、クラスや識別子として使われるセレクターは、HTMLファイルでは「<abc class=”xyz”>」もしくは「「<abc id=”xyz”>」と書き換える」という形でマーカーの中に書き込まれる）

* 釈文ファイルは大きく四つの部分、関連ファイルは二つの部分に分かれる。
	+ ヘッダー（基礎設定を含む部分。「<head＞」「</head>」というマーカーで囲まれている。釈文ファイルは行１から行９まで）
	+ 見出しの指定（釈文ファイルのみ、「<header＞」「</header>」で囲まれている。行10から行479まで）
	+ 本体（テキストの本体部分、「<body>」「</body>」で囲まれている。釈文ファイルは行481から14144まで）
	+ スクリプト指定（釈文ファイルのみ、見出しに使うjavascriptを指定。本体「</body>」の直前に置かれている
* 作業の対象となるのは本体のみ。（見出し等は将来より機能性の高いものに作り替えていく予定）
* 本体に使うマーカーはできるだけ種類を限定して、釈文ファイルでは次のものを用いる。
	+ <p></p>：paragraph（段落）をマークする。</p>で改行が行われる。
	+ <h2></h2>：表題をマークする。<h2>の中に「id="abc"」はアイデンティファイアである。見出し部分にも同じidが設定されており、見出しから該当箇所にジャンプするためにある。また、HTMLでは予め<h1>～<h6>の標題マーカが用意されており、より機能性の高い見出しをつくるには将来活用することになろう）
	+ <a></a>：anchor elementというが、リンクを貼るための足場となる。リンクを貼りたい文字列を<a></a>で囲んで用いる。リンクは「href="ファイル名"」という形式を取り、ファイル名の後ろに「#abc」を置くと、該当ファイルの冒頭ではなく、「id=abc」の箇所に直行する。「target="\_blank"」は開き方の指定、新しいタブ若しくはウィンドウで開くという意味。具体例は次の通り。
		- 簡番号にリンクを貼る場合：

<a href="Liya-zuhan/8-1477b.jpg" target="\_blank">8-1477b</a>

（</a>の前の「8-1477b」は釈文ファイルの文字列）

* + - 語釈にリンクを貼る場合：

<a href="Goshaku/kanshoku/i11.html" target="\_blank">尉</a>

* + <beppitu1></beppitu1>～<beppitu4></beppitu4>：別筆のイタリックや下線を指定する。自作マーカーで、書式の定義はcssファイルにある。
	+ <box></box>：囲い文字をつくる自作マーカー、書式はcssファイルにある。
* 関連ファイルには、注・引用文のインデントやテーブルを設定するために、また次のマーカーを使う。
	+ <div></div>,<blockquote></blockquote>：注や引用文のインデントを下げるためにつかう。<div></div>には本来そういう意味がないが、cssファイルでそのように設定した。（なお、<blockquote>はインデントのみなのに対し、<div>は一字下げも伴う）
	+ <table></table>, <tr></tr>,<td></td>：表形式を表現するためのマーカー。<td></td>は最初の単位のセールを表す。セールはさらに行単位で<tr></tr>で囲まれる。最後に表全体は<table></table>で囲まれる。複数のセールを横（行）や縦（列）の中で結合するためには、「rowspan」もしくは「colspan」という限定詞（attribute）をセールマーカの「<td>」に付ける（「<th colspan="5">」等）。「<td></td>」を「<th></th>」に変える（セール単位）か、「<tr></tr>」を「<thead></thead>」で囲む（行単位）と、セールもしくは行が、表の標題として認識される。（「<thead></thead>」に対してさらに表本体を表す「<tbody></tbody>」というマーカーを使うこともある。（これはやや複雑であるが、基本的にネット上の変換サービスがやってくれるから、あまり気にする必要はない）
	+ <ruby></ruby>,<rp></rp>,<rt></rt>：ルビを付けるためのマーカー。書式は「<ruby>字<rp>(</rp>,<rt>ルビ</rt><rp>)</rp></ruby>」となる。（現在はネット上の変換サービスの仕様のため未使用。今後Perlスクリプトを使って一括して付ける予定）
	+ <i></i>,<u></u>：イタリックや下線を表示するマーカー。釈文ファイルでは <beppitu1></beppitu1>～<beppitu4></beppitu4>を使うが、読み下し文の表をネット上の変換サービスで自動生成する関係上、取り合えず<i></i>,<u></u>などをそのまま使うしかない。自動生成の表にはその他にも若干欠点があるが、手動でテキストを編み込むよりは簡単。将来一括して修正す予定。
	+ <a></a>：表の注にも前述のanchor elementを使う。表の本体には、<a href="#\_edn注番号" name="\_ednref注番号" title="">注番号</a>という形でリンクを貼り、注は<a href="#\_ednref注番号" name="\_edn注番号" title="">注番号</a>で注番号を表示する。これも実際はネット上の変換サービスがやってくれる。

# 具体的作業手順

## 作業開始・終了時のファイル更新

* OneDrive上でHTMLが作動しないため、各自パソコン上作業をするが、多人数編集による混乱を避けるために、次のような扱いを定める。
	+ 同一の作業時間帯には常に一人のみ編集を行う（現在は月・木・金の午前中は青木さん、水曜日の午前中は飯田さん）
	+ 作業開始時には、手持ちの更新記録より更新時間が新しいファイルをOneDriveからパソコンにダウンロードする。
	+ 作業終了時には、パソコン上の更新ファイルをOneDriveにアップロードし、フォルダーごとに更新時間をテキストファイルに記録する
* 上記のように手持ちパソコン上作業できるように、最初の作業に先立って、OneDriveの資料庫フォルダーを丸ごとダウンロードして置く必要がある。（OneDriveのファイル検査によって警告の付けられたファイルは一括ダウンロードでダウンロードされないから、ファイルごとに警告を無視する選択をしてダウンロードする必要がある。一括ダウンロードから漏れたファイルの一覧表は「\_\_\_All\_Errors.txt」というファイルに自動的に格納されるので、そこから確認して個別にダウンロードを行う）

## 作業内容

* 今回の主要な作用は、「作業用訳注稿11.docx」から読み下し文の表および全ての注を然るべき関連情報ファイルに格納することを内容とする。
* 読み下し文の表はhttps://wordhtml.com/のクラウドサービスを使う。変換の流れは次の通りである。
	+ ページにアクセスして、「WordEditor」の空白部分に表を貼り付ける
	+ 「align center」（中央揃え）や「increase indent」を使って書式を揃える（仕様上「<blockquote></blockquote>」以外の仕方でインデントがつけられるが、将来一括して変更する予定）
	+ 「HTML」タグをクリックして窓を切り替える
	+ 「clean」の左側の設定ボタンは、2列目の「deletes empty tags」と「deletes tags with one space」以外は全てチェックを外した上、「clean」をクリック
	+ 窓左上の三つ目のボタン「set code intention」をクリックしてhtmlファイルの行頭インデントを設定する
	+ HTML窓の内容を全て選択（CTLR+A）してコピー（CTLR+C）し、EmEditorで開いた資料庫の該当読み下しファイルの末尾に貼り付ける（CTLR+V）。
	+ 上書きの上ブラウザーで正常な表示を確認する。
	+ 作業用訳注稿ファイルから読み下し表を削除する
	+ 注意を要するのは、表のインデントである。訳注稿ではインデントを空のセールで表現している場合には変換する前にワードの中でまず空セールを後続のセールと結合する（Alt+a+m）。（通常のインデント機能を使ってインデントを表現する場合もあるが、その場合にはそのままクラウドサービスの窓に貼り付ければよい）
	+ 更に注意を要するのは、読み下し文の注に語釈などが混じっていることである。前述の変換の前にそれらを語釈ファイルに移動させる必要がある。
* 釈読問題や語釈は、関連ファイルに雛型が組み込まれているので、その雛型に従って、訳注稿の注に含まれる関連情報を挿入していく。（語釈の場合は、ファイル未作成の状況もある。語釈のファイルは索引稿に従って作成されている関係で、それ以外のファイルは未作成。そうした場合には、取り合えず一つのワードファイルに情報を集めて、今後一括して関連ファイルをつくる。）
* 関連ファイルに格納された情報は、作業用訳注稿11.docxから削除する。
* ・釈読情報・読み下し等のない場合はそれでもリンクなどを残す（今後共同講読などによって必要になることがあるからである。空のファイルが数多く残ることになるが、開けてしまえばすぐ分かることなので問題がない）
* ・釈文ファイルには改行を「＃/#」で表している箇所もあるが、編集のついでに改行（<p></p>」に改めるべし
* 訳注稿ファイルについているコメントは無視してよい。今後別の作業で拾い集める。
* 表の別の変換サービスには[https://word2cleanhtml.com/](https://word2cleanhtml.com/cleanit)があるが、不明な原因により現在は正常に作動しない。変換の流れは次の通りである。
	+ ページにアクセスして、「Paste your document here」に表を貼り付ける
	+ 変換方法の指定は次の通りとする。

☑Remove empty paragraphs

□Convert <b> to <strong>, <i> to <em>

□Replace non-ascii with HTML entities

□Replace smart quotes with ascii equivalents

☑Indent with tabs, not spaces

☑Replace non-breaking spaces with ordinary spaces

* + 「convert to clean html」をクリックして結果を該当読み下し文ファイルの末尾に貼り付ける

# 具体的分担（2月～3月）

* 上記の作業は、火曜日と木曜日に、二人で手分けして行う。一人は5-01から昇順に、もう一人は8-2552から降順に逐次手元のは作業用訳注稿11.docxから関連情報を該当ファイルに移していく。里耶（壹）以外の作業は後日協議をする。作業中連絡しあえる（Zoomかメール）状態が望ましい。
* 月曜日には飯田さんのみの作業となるが、五一広場の資料庫作成を試験的に行う。作業工程についてはまた別途協議をする。

# 具体的分担（4月～）

* 飯田さんは引き続き昇順の移転作業と五一広場資料庫の試作を行う。
* 青木さんは、鷲尾さんの作業を継承する。鷲尾さんは降順に作業していた。鷲尾さんの最後の作業日は3月31日で、「作業用訳注稿11.docx」と作業状況を引き継ぐ必要がある。
* 青木さんが作業になれたら、飯田さんと同様に、別の資料庫の試作を行う。肩水漢簡か居延漢簡から開始するのがよかろう。
* 飯田さんと青木さんの作業日は月曜日のみ重なる。月曜日はZoomで連絡を取り合って里耶の移転作業を進めるのが望ましい。