

# PJA 日本語 EPUB 作成補助ツール

Version 1.3.0

2020 年 8 月 7 日  
有限会社サイパック

## ■ はじめに

PJA は、アクセシブルな日本語の EPUB を作成するための補助ツールです。Dolphin Publisher の EPUB 作成機能では、原本ページやルビ等について編集できません。このような Publisher が実現できない機能を補助的に実現するのが PJA の主たる用途です。また、他の製作ソフトウェアで EPUB・DAISY 2.02・HTML の編集時に、ルビやわかち書き情報を加える汎用のツールとしても利用できます。

PJA による変換は前変換と後変換の 2 つに分かれます。

### ◆ 前変換

前変換では、本文中の日本語テキストを Mecab という形態素解析エンジンで解析し、ルビや分かち書きの情報を一括して PJA の独自記法でテキストとして書き出します。

デイジー教科書を編集します。

↓ 前変換

デイジー | 教科書<きょうかしよ>を | 編集<へんしゅう>します。

### ◆ 後変換

後変換では、PJA の独自記法で記述された箇所を解析し、HTML のタグに変換します。

デイジー | 教科書<きょうかしよ>を | 編集<へんしゅう>します。

↓ 後変換

デイジー<ruby><rb>教科書</rb><rt>きょうかしよ</rt></ruby>を<wbr/><ruby><rb>編集</rb><rt>へんしゅう</rt></ruby>します。

Dolphin Publisher の固有の変換処理として以下の処理を行います。

- ・ 空の meta タグの削除
- ・ ページ番号箇所への epub:type="page-break" 属性の付与
- ・ ナビゲーションファイルにページ一覧の情報を追加

また、opf ファイルに、ページのめくり順の情報を加えます。

## ◆ 製作の流れ

PJA は以下のような製作の流れでの使用を想定しています。

PJA 前変換 ↓	EPUB の編集がある程度進んだ段階で、PJA の前変換を行い、一括してルビと分かち書きの情報を付与します。
ルビ・分かち書きの修正 ↓	機械的に付与された PJA 独自記法によるルビや分かち書きは間違いがあるため、EPUB の編集ソフトやテキストエディタ等で適宜修正します。
PJA 後変換 ↓ 完成	PJA 独自記法のルビと分かち書きの箇所を HTML のタグに変換します。また、Dolphin Publisher 固有の修正を行うこともできます。

## ■ PJA の動作する OS

### ・ Windows 版 PJA

Windows 7 及び Windows 10 で動作する 32bit アプリケーションです。GUI 版とコマンドライン版の二種類の実行ファイルがあります。Windows 版はどなたでも無償でご利用いただけます。

辞書の文字コードが UTF-8 以外で形態素解析エンジン MeCab を単体でシステムにインストールしていると、結果が文字化けしますのでご注意ください。

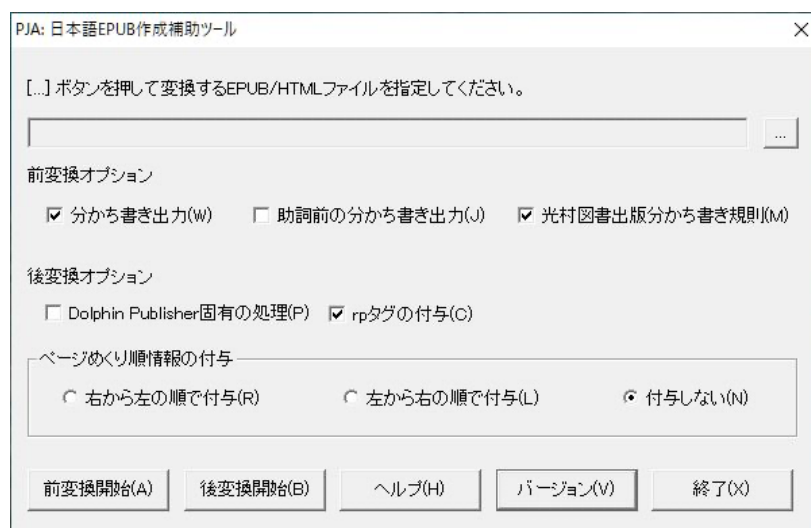
### ・ Mac / Linux 版 PJA

コマンドライン版は Mac や Linux でも動作し、サーバへの組み込み用途等にも利用できます。現在、一般への配付は行っておりません。

## ■ GUI 版 PJA

PJA をインストールすると、スタートメニューに GUI で動作する PJA が登録されます。

PJA を起動すると、次のダイアログが表示されます。



[...]ボタンを押すと、ファイル指定のダイアログが表示されます。ここで変換元とする EPUB あるいは HTML ファイルを指定します。

[前変換開始]もしくは[後変換開始]のボタンを押すと、変換元のファイルに対して、前述のような変換を行い、デスクトップ上に変換結果のファイルが生成されます。

変換間違い防止のため、前変換時には変換元のファイル名に **pre-**が自動付与されたファイルがデスクトップ上に生成されます。後変換時には変換元のファイル名に **post-**が自動付与されたファイルがデスクトップ上に生成されます。

### ◆ 前変換オプション

分ち書き情報を出力しないオプションと、助詞の前の分ち書き情報を出力しないオプションがあります。分ち書き情報を出力しない指定とした場合には、助詞の前の分ち書き情報も出力されません。

光村図書出版分ち書き規則にチェックを入れて前変換を行うと、光村図書出版の国語教科書の分ち書きにできるだけ沿った分ち書きを行います。

### ◆ 後変換オプション

#### ・ Dolphin Publisher の固有の処理

[Dolphin Publisher の固有の処理]にチェックを入れると、前述の Dolphin Publisher に固有の変換処理を行います。Dolphin Publisher 以外の EPUB 編集ツール等を使用する場合には、この指定をオフにして変換してください。

#### ・ ページめくり順情報の付与

PJA では EPUB の opf ファイルに対しページのめくり順の情報を付与できます。縦書きの EPUB を製作する際には、ページのめくり順を右から左の順に指定しないと、見開き表示で読書する場合に非常に読みにくくなるので、注意が必要です。

#### ・ rp タグの付与

ルビの前後に rp タグを使った括弧の情報を付与します。

### ■ コマンドライン版 PJA

インストール先（デフォルトでは C:\Program Files (x86)\CYPAC\PJA）に pja.exe の実行ファイルがあります。こちらは、DOS BOX の内部でコマンドラインのプログラムとして動作します。-h のオプションを付けて動作させると、次のように簡単なヘルプが表示されます。

pja [options] source\_path destination\_path

OPTION:

-a, --pre EPUB pre-process  
 -b, --post EPUB post-process (default)  
 -ha, --pre-html HTML pre-process  
 -hb, --post-html HTML post-process

PRE-PROCESS OPTIONS:

-w, --pre-wakachi-off don't output wakachi information.  
 -j, --pre-jyoshi-wakachi output jyoshi wakachi information.  
 -m, --pre-mitumura-wakachi use mitumura wakachi rule.

POST-PROCESS OPTIONS:

-p, --post-publisher-uniq-process-off don't do Dolphin Publisher uniq process.  
 -r, --rtol, --page-progression-rtol output right to left page progression.  
 -l, --ltor, --page-progression-ltor output left to right page progression.  
 -n, --none, --page-progression-none don't output page progression. (default)  
 -br, --without-rp don't output braces pair with rp tags.

コマンドライン版の変換動作は、GUI 版と同一です。

## ■ mitumurawakati.exe

光村図書出版の分かち書き規則によって、テキストファイルを分かち書きするコマンドラインで動作する mitumurawakati.exe という実行ファイルがインストール先に格納されています。入力ファイルと出力ファイルをコマンドラインで指定して、テキストの内容を分かち書きします。

-h のオプションを付けて動作させると、次のような簡単なヘルプが表示されます。

mitumurawakati.exe [options] source\_path destination\_path

OPTION:

-r, --ruby output ruby information  
 -m, --mitumura don't use mitumura wakati rule  
 -j, --jyoshi output jyoshi wakati

このコマンドで扱えるテキストファイルの文字コードは UTF-8 限定です。

## ■ 後変換時のページ番号箇所の修正

以下の例のように span タグに class="page-normal" の属性が付いている要素の配下のテキストをページ番号と見なします。(Publisher では DAISY 2.02 制作時のページ番号付けと同じ操作で、この class="page-normal" が付与されます。)

```
<p>
  <span id="dol_1_1_xdyg_0001" class="page-normal">10</span>
</p>
```

本ツールは、この箇所に epub:type="pagebreak" の属性を追記します。

```
<p>
  <span id="dol_1_1_xdyg_0001" class="page-normal" epub:type="pagebreak">10</span>
</p>
```

このようなページ番号の箇所の id と、ページ番号を元に、EPUB のナビゲーションドキュメント (nav.xhtml) を書き換えてページ番号の一覧情報を追加します。

## ◆ 後変換時の空の meta タグの削除

Dolphin Publisher では EPUB の HTML の <head> の直下に全く属性を持たない空の meta タグ <meta/> が出力される場合があります。PJA ではこのような空の meta タグを後変換の際に削除します。

## ■ PJA の独自記法

### ◆ ルビ表記

デイジー教科書用 EPUB3 には、原本通りのルビと総ルビの 2 通りのルビの表現方法があります。(原本通りのルビの場合には ruby タグに class="original" の属性を付与するコンベンションです。)

以下の例のように |, <, >, 《, 》 の記号を使って、ルビの親文字とルビを表現します。スタートは | とします。

- ・グループルビ (原本ルビ)

```
<span id="dol_1_1_xdyg_0002">デイジー | 教科書<きょうかしょ>の</span>
↓
```

<span id="dol\_1\_1\_xdyg\_0002">デイジー<ruby class="original"><rb>教科書</rb><rt>きょうかしよ</rt></ruby>の</span>

- ・ グループルビ（総ルビ）

<span id="dol\_1\_1\_xdyg\_0002">デイジー | 教科書《きょうかしよ》の</span>

↓

<span id="dol\_1\_1\_xdyg\_0002">デイジー<ruby><rb>教科書</rb><rt>きょうかしよ</rt></ruby>の</span>

- ・ モノルビ（原本ルビ）

<span id="dol\_1\_1\_xdyg\_0002">デイジー | 教<きょう | 科<か | 書<しよ>の</span>

↓

<span id="dol\_1\_1\_xdyg\_0002">デイジー<ruby class="original"><rb>教</rb><rt>きょう</rt><rb>科</rb><rt>か</rt><rb>書</rb><rt>しよ</rt></ruby>の</span>

- ・ モノルビ（総ルビ）

<span id="dol\_1\_1\_xdyg\_0002">デイジー | 教《きょう | 科《か | 書《しよ》の</span>

↓

<span id="dol\_1\_1\_xdyg\_0002">デイジー<ruby><rb>教</rb><rt>きょう</rt><rb>科</rb><rt>か</rt><rb>書</rb><rt>しよ</rt></ruby>の</span>

- ・ モノルビ（原本ルビと総ルビ混在）

以下の例では「教科」に対して原本ルビが振られ、「書」に対しては総ルビのルビが振られます。

<span id="dol\_1\_1\_xdyg\_0002">デイジー | 教<きょう | 科<か | 書《しよ》の</span>

↓

<span id="dol\_1\_1\_xdyg\_0002">デイジー<ruby class="original"><rb>教</rb><rt>きょう</rt><rb>科</rb><rt>か</rt></ruby><ruby><rb>書</rb><rt>しよ</rt></ruby>の</span>

以下の例では「教科」に対して総ルビが振られ、「書」に対しては原本ルビのルビが振られます。

<span id="dol\_1\_1\_xdyg\_0002">デイジー | 教《きょう | 科《か | 書<しよ>の</span>

↓

<span id="dol\_1\_1\_xdyg\_0002">デイジー<ruby><rb>教</rb><rt>きょう</rt><rb>科</rb><rt>か</rt></ruby><ruby class="original"><rb>書</rb><rt>しよ</rt></ruby>の</span>

### ◆ 分かち書きの変換

分かち書きの位置を示すためには EPUB の標準として採用されることになった<wbr/>のタグを使用します。

テキスト中に半角文字の~と\_用いて分かち書き位置を示すと、それぞれ<wbr class="sub"/>および<wbr/>に変換します。

<span>今日~は\_良い\_天気です。</span>

↓

<span>今日<wbr class="sub">は<wbr/>良い<wbr/>天気です。</span>

### ◆ エスケープ処理

#### ルビの開始を示す文字のエスケープ

本文中に | の記載を入れたい場合には頭の特異文字を 2 つ記載すると以下のように、1 文字の特異文字として変換できます。

<span id="dol\_1\_1\_xdyg\_0002">デージー | | 教科書《きょうかしよ》の</span>

↓

<span id="dol\_1\_1\_xdyg\_0002">デージー | 教科書《きょうかしよ》の</span>

| が前置されないで、文中に《》や〈〉がテキストとして現れる場合には、《》や〈〉はそのまま変換されずに出力されます。

#### 分かち書きの文字のエスケープ

本文中に分かち書きを示す~あるいは\_を用いたい場合には、この文字を 2 回繰り返します。本ツールでは、2 回の繰り返しを 1 文字に変換します。

~~ → ~

\_\_ → \_

## ■ 著作権

Copyright© 2019-2020 CYPAC Co, Inc.

PJA は有限会社サイパックが著作権を有するソフトウェアです。Windows 版のインストールイメージは無償でご利用いただけます。ソースコードは公開しておりません。

PJA の内部では、以下のオープンソースのソフトウェアや辞書データを使用しています。素晴らしいオープンソースのソフトウェアに感謝いたします。

### ◆ Mecab

MeCab: Yet Another Part-of-Speech and Morphological Analyzer

MeCab is copyrighted free software by Taku Kudo <taku@chasen.org> and Nippon Telegraph and Telephone Corporation, and is released under any of the GPL (see the file GPL), the LGPL (see the file LGPL), or the BSD License (see the file BSD).

Copyright(C) 2001-2012 Taku Kudo

Copyright(C) 2004-2008 Nippon Telegraph and Telephone Corporation

<https://taku910.github.io/mecab/>

### ◆ IPA 辞書

Copyright 2000, 2001, 2002, 2003 Nara Institute of Science and Technology. All Rights Reserved.

Use, reproduction, and distribution of this software is permitted. Any copy of this software, whether in its original form or modified, must include both the above copyright notice and the following paragraphs.

Nara Institute of Science and Technology (NAIST), the copyright holders, disclaims all warranties with regard to this software, including all implied warranties of merchantability and fitness, in no event shall NAIST be liable for any special, indirect or consequential damages or any damages whatsoever resulting from loss of use, data or profits, whether in an action of contract, negligence or other tortuous action, arising out of or in connection with the use or performance of this software.



A large portion of the dictionary entries originate from ICOT Free Software. The following conditions for ICOT Free Software applies to the current dictionary as well.

Each User may also freely distribute the Program, whether in its original form or modified, to any third party or parties, PROVIDED that the provisions of Section 3 ("NO WARRANTY") will ALWAYS appear on, or be attached to, the Program, which is distributed substantially in the same form as set out herein and that such intended distribution, if actually made, will neither violate or otherwise contravene any of the laws and regulations of the countries having jurisdiction over the User or the intended distribution itself.

## NO WARRANTY

The program was produced on an experimental basis in the course of the research and development conducted during the project and is provided to users as so produced on an experimental basis. Accordingly, the program is provided without any warranty whatsoever, whether express, implied, statutory or otherwise. The term "warranty" used herein includes, but is not limited to, any warranty of the quality, performance, merchantability and fitness for a particular purpose of the program and the nonexistence of any infringement or violation of any right of any third party.

Each user of the program will agree and understand, and be deemed to have agreed and understood, that there is no warranty whatsoever for the program and, accordingly, the entire risk arising from or otherwise connected with the program is assumed by the user.

Therefore, neither ICOT, the copyright holder, or any other organization that participated in or was otherwise related to the development of the program and their respective officials, directors, officers and other employees shall be held liable for any and all damages, including, without limitation, general, special, incidental and consequential damages, arising out of or otherwise in connection with the use or inability to use the program or any product, material or result produced or otherwise obtained by using the program,

regardless of whether they have been advised of, or otherwise had knowledge of, the possibility of such damages at any time during the project or thereafter. Each user will be deemed to have agreed to the foregoing by his or her commencement of use of the program. The term "use" as used herein includes, but is not limited to, the use, modification, copying and distribution of the program and the production of secondary products from the program.

In the case where the program, whether in its original form or modified, was distributed or delivered to or received by a user from any person, organization or entity other than ICOT, unless it makes or grants independently of ICOT any specific warranty to the user in writing, such person, organization or entity, will also be exempted from and not be held liable to the user for any such damages as noted above as far as the program is concerned.

#### ◆ ZLIB

#### ZLIB DATA COMPRESSION LIBRARY

zlib 1.2.11 is a general purpose data compression library. All the code is thread safe. The data format used by the zlib library is described by RFCs (Request for Comments) 1950 to 1952 in the files <http://tools.ietf.org/html/rfc1950> (zlib format), rfc1951 (deflate format) and rfc1952 (gzip format).

All functions of the compression library are documented in the file `zlib.h` (volunteer to write man pages welcome, contact [zlib@gzip.org](mailto:zlib@gzip.org)). A usage example of the library is given in the file `test/example.c` which also tests that the library is working correctly. Another example is given in the file `test/minigzip.c`. The compression library itself is composed of all source files in the root directory.

To compile all files and run the test program, follow the instructions given at the top of `Makefile.in`. In short `./configure; make test`, and if that goes well, `make install` should work for most flavors of Unix. For Windows, use one of the special makefiles in `win32/` or `contrib/vstudio/`. For VMS, use `make_vms.com`.

Questions about zlib should be sent to <zlib@gzip.org>, or to Gilles Vollant <info@winimage.com> for the Windows DLL version. The zlib home page is <http://zlib.net/>. Before reporting a problem, please check this site to verify that you have the latest version of zlib; otherwise get the latest version and check whether the problem still exists or not.

PLEASE read the zlib FAQ [http://zlib.net/zlib\\_faq.html](http://zlib.net/zlib_faq.html) before asking for help.

Mark Nelson <markn@ieee.org> wrote an article about zlib for the Jan. 1997 issue of Dr. Dobbs's Journal; a copy of the article is available at <http://marknelson.us/1997/01/01/zlib-engine/>.

The changes made in version 1.2.11 are documented in the file ChangeLog.

Unsupported third party contributions are provided in directory contrib/.

zlib is available in Java using the java.util.zip package, documented at <http://java.sun.com/developer/technicalArticles/Programming/compression/>.

A Perl interface to zlib written by Paul Marquess <pmqs@cpan.org> is available at CPAN (Comprehensive Perl Archive Network) sites, including <http://search.cpan.org/~pmqs/IO-Compress-Zlib/>.

A Python interface to zlib written by A.M. Kuchling <amk@amk.ca> is available in Python 1.5 and later versions, see <http://docs.python.org/library/zlib.html>.

zlib is built into tcl: <http://wiki.tcl.tk/4610>.

An experimental package to read and write files in .zip format, written on top of zlib by Gilles Vollant <info@winimage.com>, is available in the contrib/minizip directory of zlib.

Notes for some targets:

- For Windows DLL versions, please see win32/DLL\_FAQ.txt
- For 64-bit Irix, deflate.c must be compiled without any optimization. With

- O, one libpng test fails. The test works in 32 bit mode (with the -n32 compiler flag). The compiler bug has been reported to SGI.
- zlib doesn't work with gcc 2.6.3 on a DEC 3000/300LX under OSF/1 2.1 it works when compiled with cc.
- On Digital Unix 4.0D (formerly OSF/1) on AlphaServer, the cc option -std1 is necessary to get gzprintf working correctly. This is done by configure.
- zlib doesn't work on HP-UX 9.05 with some versions of /bin/cc. It works with other compilers. Use "make test" to check your compiler.
- gzopen is not supported on RISCOS or BEOS.
- For PalmOs, see <http://palmzlib.sourceforge.net/>

#### Acknowledgments:

The deflate format used by zlib was defined by Phil Katz. The deflate and zlib specifications were written by L. Peter Deutsch. Thanks to all the people who reported problems and suggested various improvements in zlib; they are too numerous to cite here.

#### Copyright notice:

(C) 1995-2017 Jean-loup Gailly and Mark Adler

This software is provided 'as-is', without any express or implied warranty. In no event will the authors be held liable for any damages arising from the use of this software.

Permission is granted to anyone to use this software for any purpose, including commercial applications, and to alter it and redistribute it freely, subject to the following restrictions:

1. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be

appreciated but is not required.

2. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.
3. This notice may not be removed or altered from any source distribution.

Jean-loup Gailly

jloup@gzip.org

Mark Adler

madler@alumni.caltech.edu

If you use the zlib library in a product, we would appreciate \*not\* receiving lengthy legal documents to sign. The sources are provided for free but without warranty of any kind. The library has been entirely written by Jean-loup Gailly and Mark Adler; it does not include third-party code.

If you redistribute modified sources, we would appreciate that you include in the file ChangeLog history information documenting your changes. Please read the FAQ for more information on the distribution of modified source versions.