

LuaTeX-ja パッケージ

LuaTeX-ja プロジェクトチーム

2011 年 11 月 2 日

# 目次

第 I 部	ユーザズマニュアル	3
1	はじめに	3
1.1	背景	3
1.2	pT <sub>E</sub> X から主な変更点	3
1.3	用語と記法	4
1.4	プロジェクトについて	4
2	使い方	5
2.1	インストール	5
2.2	注意点	5
2.3	plain T <sub>E</sub> X で使う	5
2.4	L <sup>A</sup> T <sub>E</sub> X で使う	6
2.5	フォントの変更	7
3	パラメータの変更	8
3.1	J <sub>A</sub> char の範囲の設定	8
3.2	kanjiskip と xkanjiskip	10
3.3	xkanjiskip の設定の挿入	11
3.4	ベースラインの移動	11
3.5	トンボ	12
第 II 部	リファレンス	12
4	フォントメトリックと日本語フォント	12
4.1	\jfont プリミティブ	12
4.2	JFM ファイルの構造	13
4.3	数式フォントファミリ	15
4.4	コールバック	16
5	パラメータ	17
5.1	\ltjsetParameter プリミティブ	17
5.2	パラメーター一覧	17
6	その他のプリミティブ	19
6.1	互換プリミティブ	19
6.2	\inhibitglue プリミティブ	19

7	LaTeX <sub>2<math>\epsilon</math></sub> 用のコントロールシーケンス	19
7.1	NFSS2 へのパッチ	19
7.2	トンボ	21
8	拡張	21
8.1	luatexja-fontspec.sty	21
8.2	luatexja-otf.sty	21
<b>第 III 部 実装</b>		<b>21</b>
9	パラメータの保持	21
9.1	用いられる寸法レジスタ, 属性レジスタ, whatsit ノード	21
9.2	LuaTeX-ja のスタックシステム	22
10	和文文字直後の改行	24
10.1	参考: pTeX の挙動	24
10.2	LuaTeX-ja の挙動	25
11	JFM グルーの挿入, kanjiskip と xkanjiskip	25
11.1	概要	25
11.2	「クラスタ」の定義	26

本ドキュメントはまだまだ未完成です。

## 第 I 部

# ユーザーズマニュアル

## 1 はじめに

LuaTeX-j<sub>a</sub> パッケージは、次世代標準 T<sub>E</sub>X である LuaTeX の上で、pT<sub>E</sub>X と同等/それ以上の品質の日本語組版を実現させようとするマクロパッケージである。

### 1.1 背景

従来、「T<sub>E</sub>X を用いて日本語組版を行う」といったとき、エンジンとしては ASCII pT<sub>E</sub>X やその拡張物が用いられることが一般的であった。pT<sub>E</sub>X は T<sub>E</sub>X のエンジン拡張であり、(少々仕様上不便な点はあるものの) 商業印刷の分野にも用いられるほどの高品質な日本語組版を可能としている。だが、それは弱点にもなってしまった：pT<sub>E</sub>X という(組版的に)満足なものがあつたため、海外で行われている数々の T<sub>E</sub>X の拡張 例えば ε-T<sub>E</sub>X や pdfT<sub>E</sub>X や、TrueType, OpenType, Unicode といった計算機で日本語を扱う際の状況の変化に追従することを怠ってしまったのだ。

ここ数年、若干状況は改善されてきた。現在手に入る大半の pT<sub>E</sub>X バイナリでは外部 UTF-8 入力を利用可能となり、さらに Unicode 化を推進し、pT<sub>E</sub>X の内部処理まで Unicode 化した upT<sub>E</sub>X も開発されている。また、pT<sub>E</sub>X に ε-T<sub>E</sub>X 拡張をマージした ε-pT<sub>E</sub>X も登場し、T<sub>E</sub>X Live 2011 では pL<sup>A</sup>T<sub>E</sub>X が ε-pT<sub>E</sub>X の上で動作するようになった。だが、pdfT<sub>E</sub>X 拡張 (PDF 直接出力や micro-typsetting) を pT<sub>E</sub>X に対応させようという動きはなく、海外との gap は未だにあるのが現状である。

しかし、LuaTeX の登場で、状況は大きく変わるようになった。Lua コードで ‘callback’ を書くことにより、LuaTeX の内部処理に割り込みをかけることが可能となった。これは、エンジン拡張という真似をしなくても、Lua コードとそれに関する T<sub>E</sub>X マクロを書けば、エンジン拡張とほぼ同程度のことができるようになったということの意味する。LuaTeX-j<sub>a</sub> は、このアプローチによって Lua コード・T<sub>E</sub>X マクロによって日本語組版を LuaTeX の上で実現させようという目的で開発が始まったパッケージである。

### 1.2 pT<sub>E</sub>X からの主な変更点

LuaTeX-j<sub>a</sub> は、pT<sub>E</sub>X に多大な影響を受けている。初期の開発目標は、pT<sub>E</sub>X の機能を Lua コードにより実装することであった。しかし、開発が進むにつれ、pT<sub>E</sub>X の完全な移植は不可能であり、また pT<sub>E</sub>X における実装がいささか不可解になっているような状況も発見された。そのため、LuaTeX-j<sub>a</sub> は、もはや pT<sub>E</sub>X の完全な移植は目標とはしない。pT<sub>E</sub>X における不自然な仕様・挙動があれば、そこは積極的に改める。

以下は pT<sub>E</sub>X からの主な変更点である。

- 和文フォントは「実際の」フォント、和文フォントメトリック (JFM と呼ぶ)、そして ‘variation’ と呼ばれる文字列の組である。
- 日本語の文書中では改行はほとんどどこでも許されるので、pT<sub>E</sub>X では和文文字直後の改行は無視される (スペースが入らない) ようになっていた。しかし、LuaTeX-j<sub>a</sub> では LuaTeX の仕様のためにこの機能は完全には実装されていない。
- 2 つの和文文字の間、和文文字と欧文文字の間に入るグルー / カーン (JAglue と呼ぶ) の挿入処理が 0 から書き直されている。
  - LuaTeX の内部での文字の扱いが「ノードベース」になっているように (例えば、of{}fice で合字は抑制されない)、JAglue の挿入処理も「ノードベース」である。

- さらに、2つの文字の間にある行末では効果を持たないノード（例えば`\special` ノード）や、イタリック補正に伴い挿入されるカーンは挿入処理中では無視される。
- 注意：上の2つの変更により、従来 `JAgue` の挿入処理を分断するのに使われていたいくつかの方法は用いることができない。具体的には、次の方法はもはや無効である：
  - ちょ{}っと    ちょ\っと
 もし同じことをやりたければ、空の `hbox` を間に挟めばよい：
  - ちょ\hbox{}っと
- 処理中では、2つの和文フォントは、「実際の」フォントのみが異なる場合に同一視される。
- 現時点では、縦書きは `LuaTeX-ja` ではサポートされていない。

詳細については第 III 部を見よ。

### 1.3 用語と記法

本ドキュメントでは、以下の用語と記法を用いる：

- 文字は2種類に分けられる：
  - `JAchar`: ひらがな、カタカナ、漢字、和文用の約物といった和文文字のことを指す。
  - `ALchar`: アルファベットを始めとする、その他全ての文字を指す。
 そして、`ALchar` の出力に用いられるフォントを「欧文フォント」と呼び、`JAchar` の出力に用いられるフォントを「和文フォント」と呼ぶ。
- サンセリフ体で書かれた語（例：`prebreakpenalty`）は日本語組版用のパラメータを表し、これらは `\ltjsetparameter` コマンドのキーとして用いられる。
- 下線付きのタイプライタ体で書かれた語（例：`fontspec`）は `LATEX` のパッケージやクラスを表す。
- 「プリミティブ」という語を、`LuaTeX` のプリミティブだけではなく `LuaTeX-ja` のコアモジュールで定義されたコントロールシーケンスに対しても用いる。
- 本ドキュメントでは、自然数は0から始まる。

### 1.4 プロジェクトについて

プロジェクト Wiki    プロジェクト Wiki は構築中である。

- <http://sourceforge.jp/projects/luatex-ja/wiki/FrontPage> (日本語)
- <http://sourceforge.jp/projects/luatex-ja/wiki/FrontPage%28en%29> (英語)

本プロジェクトは `SourceForge.JP` のサービスを用いて運営されている。

開発メンバー

- |         |          |         |
|---------|----------|---------|
| • 北川 弘典 | • 前田 一貴  | • 八登 崇之 |
| • 黒木 裕介 | • 阿部 紀行  | • 山本 宗宏 |
| • 本田 知亮 | • 齋藤 修三郎 |         |

## 2 使い方

### 2.1 インストール

LuaTeX-ja パッケージのインストールには、次のものが必要である。

- LuaTeX (バージョン 0.65.0-beta 以降) とその支援パッケージ、TeX Live 2011 や W32TeX の最新版ならば問題ない。
- LuaTeX-ja のソースアーカイブ (もちろん :)

インストール方法は以下ようになる：

1. ソースアーカイブをダウンロードする。

現時点では、LuaTeX-ja の公式リリースはないので、レポジトリからアーカイブを取得しなければならない。次のようにすることで、Git レポジトリを取得することができる：

```
$ git clone git://git.sourceforge.jp/gitroot/luatex-ja/luatexja.git
```

もしくは、master ブランチの HEAD アーカイブを以下からダウンロードすることができる：

```
http://git.sourceforge.jp/view?p=luatex-ja/luatexja.git;a=snapshot;h=HEAD;sf=tgz.
```

開発中の最新の成果は master ブランチには含まれていないかもしれないことに注意。

2. アーカイブを展開する。src/ をはじめとしたいいくつかのディレクトリができる。
3. src/ の中身全てを自分の TEXMF ツリーにコピーする。
4. もし mktexlsr を実行する必要があるらそうする。

### 2.2 注意点

- 原稿のソースファイルの文字コードは UTF-8 でなければならない。EUC-JP や Shift-JIS は使用できない。
- いくつかのパッケージと衝突する。  
例えば、JAchar の範囲の設定がデフォルトのままだと、現行のバージョンでは `unicode-math` パッケージと共存できない。以下の行をプリアンプルに追加することで数学記号が正しく出るようになるが、副作用としていくつかの和文文字が ALchar として扱われるようになってしまう：

```
\ltjsetparameter{jacharrange={-3, -8}}
```

### 2.3 plain TeX で使う

LuaTeX-ja を plain TeX で使うためには、単に次の行をソースファイルの冒頭に追加すればよい：

```
\input luatexja.sty
```

これで (ptex.tex のように) 日本語組版のための最低限の設定がなされる：

- 以下の 6 つの和文フォントが定義される：

字体	フォント名	‘10 pt’	‘7 pt’	‘5 pt’
明朝体	Ryumin-Light	<code>\tenmin</code>	<code>\sevenmin</code>	<code>\fivemin</code>
ゴシック体	GothicBBB-Medium	<code>\tengt</code>	<code>\seventgt</code>	<code>\fivegt</code>

- ‘Q(級)’ は日本の写植で用いられる単位で、 $1\text{Q} = 0.25\text{mm}$  である。この長さは `\jq` に保持されている。
- ‘Ryumin-Light’ と ‘GothicBBB-Medium’ は PDF ファイルに埋め込まずに名前参照のみで用いることが広く受け入れられており、この場合 PDF リーダーが適切な外部フォントで代用する（例えば、Adobe Reader では Ryumin-Light は小塚明朝で代替される）。そこで、これらをデフォルトのフォントとして採用する。
- 欧文フォントの文字は和文フォントの文字よりも、同じ文字サイズでも一般に小さい。そこで、これらの和文フォントの実際のサイズは指定された値よりも小さくなるように設定されており、具体的には指定の 0.962216 倍にスケールされる。
- `JAchar` と `ALchar` の間に入るグルー (`xkanjiskip`) の量は次のように設定されている：

$$(0.25 \cdot 0.962216 \cdot 10\text{pt})_{-1\text{pt}}^{+1\text{pt}} = 2.40554\text{pt}_{-1\text{pt}}^{+1\text{pt}}$$

## 2.4 L<sup>A</sup>T<sub>E</sub>X で使う

L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> を用いる場合も基本的には同じである。日本語組版のための最低限の環境を設定するためには、`luatexja.sty` を読み込むだけでよい：

```
\usepackage{luatexja}
```

これで pL<sup>A</sup>T<sub>E</sub>X の `plfonts.dtx` と `pldefs.ltx` に相当する最低限の設定がなされる：

- JY3 は和文フォント用のフォントエンコーディングである（横書き用）。将来的に、LuaT<sub>E</sub>X-j<sub>a</sub> で縦書きがサポートされる際には、JT3 を縦書き用として用いる予定である。
- 2 つのフォントファミリ `mc` と `gt` が定義されている：

字体	ファミリ	<code>\mdseries</code>	<code>\bfseries</code>	スケール
明朝体	<code>mc</code>	Ryumin-Light	GothicBBB-Medium	0.962216
ゴシック体	<code>gt</code>	GothicBBB-Medium	GothicBBB-Medium	0.962216

どちらのファミリにおいても、その bold シリーズはゴシック体の medium シリーズであることに注意。これは初期の DTP において和文フォントが 2 つ（それがちょうど Ryumin-Light, GothicBBB-Medium だった）しか利用できなかった時の名残であり、pL<sup>A</sup>T<sub>E</sub>X での標準設定とも同じである。

- 数式モード中の和文文字は `mc` ファミリで出力される。

しかしながら、上記の設定は日本語の文書にとって十分とは言えない。日本語文書を組版するためには、`article.cls`, `book.cls` といった欧文用のクラスファイルではなく、和文用のクラスファイルを用いた方がよい。現時点では、`jlclasses` (pL<sup>A</sup>T<sub>E</sub>X の標準クラス) と `jsclasses` (奥村晴彦氏によるクラスファイル) に対応するものとして、`ltjclasses`, `ltjsclasses` がそれぞれ用意されている。

`\CID`, `\UTF` と OTF パッケージのマクロ pL<sup>A</sup>T<sub>E</sub>X では、JIS X 0208 にない Adobe-Japan1-6 の文字を出力するために、齋藤修三郎氏による `otf` パッケージが用いられていた。このパッケージは広く用いられているため、LuaT<sub>E</sub>X-j<sub>a</sub> においても `otf` パッケージの機能の一部をサポートしている。これらの機能を用いるためには `luatexja-otf` パッケージを読み込めばよい。

1 森  
 2 \UTF{9DD7}外と内田百\UTF{9592}とが\UTF{9AD9}島  
 屋に行く。 森鷗外と内田百閒とが高島屋に行く。  
 3 葛飾区の吉野家， 葛飾区の吉野家  
 4 \CID{7652}飾区の\CID{13706}野家，  
 5 葛飾区の吉野家

## 2.5 フォントの変更

意見：数式モード中の和文文字 pTeX では数式中での和文文字の使用がサポートされていたので，以下のようなソースが存在した：

```

1 $f_{高温}$~($f_{\text{high temperature}}$).      f高温 (fhigh temperature).
2 \[ y=(x-1)^2+2\quad{よって}\quad y>0 \]          y = (x - 1)2 + 2   よって   y > 0
3 $5\in{}素:=\{\,p\in\mathbb N:\text{\$p\$ is a
   prime}\,\,\}$.      5 ∈ 素 := {p ∈ N : p is a prime}.

```

LuaTeX-jp プロジェクトでは，数式モード中での和文文字はそれらが識別子として用いられるときのみ許されると考えている．この観点から，

- 上記数式のうち 1, 2 行目は正しくない．なぜならば‘高温’が意味のあるラベルとして，‘よって’が接続詞として用いられているからである．
- しかしながら，3 行目は‘素’が識別子として用いられているので正しい．

したがって，LuaTeX-jp プロジェクトの意見としては，上記の入力は次のように直されるべきである：

```

1 $f_{\text{高温}}$~%
2 ($f_{\text{high temperature}}$).      f高温 (fhigh temperature).
3 \[ y=(x-1)^2+2\quad
4 \mathrel{\text{よって}}\quad y>0 \]          y = (x - 1)2 + 2   よって   y > 0
5 $5\in{}素:=\{\,p\in\mathbb N:\text{\$p\$ is a
   prime}\,\,\}$.      5 ∈ 素 := {p ∈ N : p is a prime}.

```

また LuaTeX-jp プロジェクトでは，和文文字が識別子として用いられることはほとんどないと考えており，したがってこの章では数式モード中の和文フォントを変更する方法については記述しない．この方法については第 II 部を参照のこと．

plain TeX plain TeX で和文フォントを変更するためには，\jfont プリミティブを用いなければならない．第 II 部を参照せよ．

NFSS2 L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> については，LuaTeX-jp では単にフォント選択システムを pL<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> (plfonts.dtx) から採用している．

- 2 つのコントロールシーケンス \mcdefault と \gtdefault がそれぞれ明朝体とゴシック体のデフォルトのフォントファミリを指定するために用いられる．初期値：\mcdefault は mc，\gtdefault は gt．
- \fontfamily, \fontseries, \fontshape, そして \selectfont が和文フォントの属性を変更するために使用できる．



	エンコーディング	ファミリー	シリーズ	シェーブ	選択
欧文フォント	<code>\romanencoding</code>	<code>\romanfamily</code>	<code>\romanseries</code>	<code>\romanshape</code>	<code>\useroman</code>
和文フォント	<code>\kanjiencoding</code>	<code>\kanjifamily</code>	<code>\kanjiserie</code>	<code>\kanjishape</code>	<code>\usekanji</code>
両方	—	—	<code>\fontseries</code>	<code>\fontshape</code>	—
自動選択	<code>\fontencoding</code>	<code>\fontfamily</code>	—	—	<code>\usefont</code>

ここで、`\fontencoding{<encoding>}`は、引数により和文側か欧文側かのどちらかが切り替わる。例えば、`\fontencoding{JY3}`は和文フォントのエンコーディングを JY3 に変更し、`\fontencoding{T1}` は欧文フォント側を T1 へと変更する。`\fontfamily` も引数により和文側、欧文側、あるいは両方のフォントファミリーが切り替わる。詳細は 7.1 節を参照すること。

- 和文フォントファミリーの定義には `\DeclareFontFamily` の代わりに `\DeclareKanjiFamily` を用いる。しかし、現在の実装では `\DeclareFontFamily` を用いても問題は生じない。

`fontspec` パッケージと同様の機能を和文フォントに対しても用いるためには、`luatexja-fontspec` パッケージをプリアンブルで読み込む必要がある。このパッケージは必要ならば自動で `luatexja` パッケージと `fontspec` パッケージを読み込む。

`luatexja-fontspec` パッケージでは、以下の 7 つのコマンドを `fontspec` パッケージの元のコマンドに対応するものとして定義している：

和文フォント	<code>\jfontspec</code>	<code>\setmainjfont</code>	<code>\setsansjfont</code>	<code>\newfontfamily</code>
欧文フォント	<code>\fontspec</code>	<code>\setmainfont</code>	<code>\setsansfont</code>	<code>\newfontfamily</code>
和文フォント	<code>\newjfontface</code>	<code>\defaultjfontfeatures</code>	<code>\addjfontfeatures</code>	
欧文フォント	<code>\newfontface</code>	<code>\defaultfontfeatures</code>	<code>\addfontfeatures</code>	

```

1 \fontspec[Numbers=OldStyle]{TeX Gyre Termes}
2 \jfontspec{IPAexMincho}
3 JIS-X-0213:2004→辻
4
5 \addfontfeatures{CJKShape=JIS1990}
6 JIS-X-0208:1990→辻

```

和文フォントについては全ての和文文字のグリフがほぼ等幅であるのが普通であるため、`\setmonojfont` コマンドは存在しないことに注意。また、これらの和文用の 7 つのコマンドでは Kerning feature はデフォルトでは off となっている。これはこの feature が JAgglue と衝突するためである (4.1 節を見よ)。

### 3 パラメータの変更

LuaTeX-ja には多くのパラメータが存在する。そして LuaTeX の振る舞いのために、その多くは TeX のレジスタにはなく、LuaTeX-ja 独自の 방법으로保持されている。そのため、これらのパラメータを設定・取得するためには `\ltjsetparameter` と `\ltjgetparameter` を用いなければならない。

#### 3.1 JAchar の範囲の設定

JAchar の範囲を設定するためには、まず文字範囲に 0 より大きく 217 より小さい自然数を割り当てる必要がある。これには `\ltjdefcharrange` プリミティブを用いる。例えば、次のように書くことで追加多言語面 (SMP) にある全ての文字と ‘漢’ の範囲番号が 100 に設定される。

```
\ltjdefcharrange{100}{"10000-"1FFFF,`漢}
```

この文字範囲への番号の割り当てはいつもグローバルであり、したがって文書の途中でこの操作をするべきではない。

もし指定されたある文字がある非零番号の範囲に属していたならば、これは新しい設定で書き換えられる。例えば、SMP は全て LuaTeX-j<sub>a</sub> のデフォルトでは 4 番の文字範囲に属しているが、上記の指定を行えば SMP は 100 番に属すようになり、4 番からは除かれる。

文字範囲に番号を割り当てた後は、jacharrange パラメータが JAchar として扱われる文字の範囲を設定するために用いられる。例えば、以下は LuaTeX-j<sub>a</sub> の初期設定である：

```
\ltjsetparameter{jacharrange={-1, +2, +3, -4, -5, +6, +7, +8}}
```

jacharrange パラメータには整数のリストを与える。リスト中の負の整数  $-n$  は「文字範囲  $n$  に属する文字は ALchar として扱われる」ことを意味し、正の整数  $+n$  は JAchar として扱うことを意味する。

初期設定 LuaTeX-j<sub>a</sub> では 8 つの文字範囲を設定している。これらは以下のデータに基づいて決定している。

- Unicode 6.0 のブロック。
- Adobe-Japan1-UCS2 による Adobe-Japan1-6 の CID と Unicode の間のマッピング。
- 八登崇之氏による upTeX 用の PXbase バンドル。

以下ではこれら 8 つの文字範囲について記述する。番号のあとのアルファベット ‘J’ と ‘A’ はデフォルトで JAchar として扱われるかどうかを示す。これらの設定は PXbase バンドルで定義されている prefercjk と類似のものである。

範囲 8<sup>J</sup> ISO 8869 1 の上半分（ラテン 1 補助）と JIS X 0208 の共通部分にある記号。この文字範囲は以下の文字で構成される：

- |                      |                     |
|----------------------|---------------------|
| • § (U+00A7, 節記号)    | • ´ (U+00B4, アクセント) |
| • ¨ (U+00A8, トレマ)    | • ¶ (U+00B6, 段落記号)  |
| • ° (U+00B0, 度)      | • × (U+00D7, 乗算記号)  |
| • ± (U+00B1, 正又は負符号) | • ÷ (U+00F7, 除算記号)  |

範囲 1<sup>A</sup> ラテン文字。一部は Adobe-Japan1-6 にも含まれている。この範囲は以下の Unicode のブロックから構成されている。ただし、範囲 8 は除く。

- |                                 |                                     |
|---------------------------------|-------------------------------------|
| • U+0080–U+00FF: ラテン 1 補助       | • U+0300–U+036F: ダイアクリティカルマーク(合成可能) |
| • U+0100–U+017F: ラテン文字拡張 A      | • U+1E00–U+1EFF: ラテン文字拡張追加          |
| • U+0180–U+024F: ラテン文字拡張 B      |                                     |
| • U+0250–U+02AF: IPA 拡張(国際音声記号) |                                     |
| • U+02B0–U+02FF: 前進を伴う修飾文字      |                                     |

範囲 2<sup>J</sup> ギリシャ文字とキリル文字。JIS X 0208 (したがって多くの和文フォント) はこれらの文字を持つ。

- |                                |                          |
|--------------------------------|--------------------------|
| • U+0370–U+03FF: ギリシア文字及びコプト文字 | • U+1F00–U+1FFF: キリル文字補助 |
| • U+0400–U+04FF: キリル文字         |                          |

範囲 3<sup>J</sup> 句読点と記号類。ブロックのリストは表 1 に示してある。

範囲 4<sup>A</sup> 通常和文フォントには含まれていない文字。この範囲は他の範囲にないほとんど全ての Unicode ブロックで構成されている。したがって、ブロックのリストを示す代わりに、範囲の定義そのものを示す：

```
\ltjdefcharrange{4}{%  
"500-"10FF, "1200-"1DFF, "2440-"245F, "27C0-"28FF, "2A00-"2AFF,
```

表 1. 文字範囲 3 に指定されている Unicode ブロック .

U+2000-U+206F	一般句読点	U+2070-U+209F	上付き・下付き
U+20A0-U+20CF	通貨記号	U+20D0-U+20FF	記号用ダイアクリティカルマーク (合成可能)
U+2100-U+214F	文字様記号	U+2150-U+218F	数字に準じるもの
U+2190-U+21FF	矢印	U+2200-U+22FF	数学記号 (演算子)
U+2300-U+23FF	その他の技術用記号	U+2400-U+243F	制御機能用記号
U+2500-U+257F	罫線素片	U+2580-U+259F	ブロック要素
U+25A0-U+25FF	幾何学模様	U+2600-U+26FF	その他の記号
U+2700-U+27BF	装飾記号	U+2900-U+297F	補助矢印 B
U+2980-U+29FF	その他の数学記号 B	U+2B00-U+2BFF	その他の記号及び矢印
U+E000-U+F8FF	私用領域 (外字領域)		

表 2. 文字範囲 6 に指定されている Unicode ブロック .

U+2460-U+24FF	囲み英数字	U+2E80-U+2EFF	CJK 部首補助
U+3000-U+303F	CJK の記号及び句読点	U+3040-U+309F	平仮名
U+30A0-U+30FF	片仮名	U+3190-U+319F	漢文用記号 (返り点)
U+31F0-U+31FF	片仮名拡張	U+3200-U+32FF	囲み CJK 文字・月
U+3300-U+33FF	CJK 互換用文字	U+3400-U+4DBF	CJK 統合漢字拡張 A
U+4E00-U+9FFF	CJK 統合漢字	U+F900-U+FAFF	CJK 互換漢字
U+FE10-U+FE1F	縦書き形	U+FE30-U+FE4F	CJK 互換形
U+FE50-U+FE6F	小字形	U+20000-U+2FFFF	(追加多言語面)

表 3. 文字範囲 7 に指定されている Unicode ブロック .

U+1100-U+11FF	ハングル字母	U+2F00-U+2FDF	康熙部首
U+2FF0-U+2FFF	漢字構成記述文字	U+3100-U+312F	注音字母 (注音符号)
U+3130-U+318F	ハングル互換字母	U+31A0-U+31BF	注音字母拡張
U+31C0-U+31EF	CJK の筆画	U+A000-U+A48F	イ文字
U+A490-U+A4CF	イ文字部首	U+A830-U+A83F	共通インド数字に準じるもの
U+AC00-U+D7AF	ハングル音節文字	U+D7B0-U+D7FF	ハングル字母拡張 B

"2C00-"2E7F, "4DC0-"4DFF, "A4D0-"A82F, "A840-"ABFF, "FB50-"FE0F,  
"FE20-"FE2F, "FE70-"FEFF, "FB00-"FB4F, "10000-"1FFFF} % non-Japanese

範囲 5<sup>A</sup> 代用符号と補助私用領域 .

範囲 6<sup>J</sup> 日本語で用いられる文字 . ブロックのリストは表 2 に示す .

範囲 7<sup>J</sup> CJK 言語で用いられる文字のうち , Adobe-Japan1-6 に含まれていないもの . ブロックのリストは表 3 に示す .

## 3.2 kanjiskip と xkanjiskip

**JAg**lue is divided into the following three categories:

- Glues/kerns specified in JFM. If `\inhibitglue` is issued around a Japanese character, this glue will be not inserted at the place.
- The default glue which inserted between two **JA**chars ( kanjiskip).
- The default glue which inserted between a **JA**char and an **AL**char (xkanjiskip).

The value (a skip) of kanjiskip or xkanjiskip can be changed as the following.

```
\ltjsetparameter{kanjiskip={Opt plus 0.4pt minus 0.4pt},
  xkanjiskip={0.25\zw plus 1pt minus 1pt}}
```

It may occur that JFM contains the data of ‘ideal width of kanjiskip’ and/or ‘ideal width of xkanjiskip’. To use these data from JFM, set the value of kanjiskip or xkanjiskip to `\maxdimen`.

### 3.3 xkanjiskip の設定の挿入

It is not desirable that xkanjiskip is inserted between every boundary between **J**Achars and **A**Lchars. For example, xkanjiskip should not be inserted after opening parenthesis (*e.g.*, compare ‘(あ’ and ‘( あ’).

LuaTeX-ja can control whether xkanjiskip can be inserted before/after a character, by changing jaxspmode for **J**Achars and alxspmode parameters **A**Lchars respectively.

```
1 \ltjsetparameter{jaxspmode={`あ,preonly},
  alxspmode={`!\!,postonly}}          p あq!う
2 p あ q !う
```

The second argument `preonly` means ‘the insertion of xkanjiskip is allowed before this character, but not after’. the other possible values are `postonly`, `allow` and `inhibit`. For the compatibility with pTeX, natural numbers between 0 and 3 are also allowed as the second argument<sup>\*1</sup>.

If you want to enable/disable all insertions of kanjiskip and xkanjiskip, set `autospacing` and `autoxspacing` parameters to `false`, respectively.

### 3.4 ベースラインの移動

To make a match between a Japanese font and an alphabetic font, sometimes shifting of the baseline of one of the pair is needed. In pTeX, this is achieved by setting `\ybaselineshift` to a non-zero length (the baseline of alphabetic fonts is shifted below). However, for documents whose main language is not Japanese, it is good to shift the baseline of Japanese fonts, but not that of alphabetic fonts. Because of this, LuaTeX-ja can independently set the shifting amount of the baseline of alphabetic fonts (`yalbaselineshift` parameter) and that of Japanese fonts (`yjabaselineshift` parameter).

```
1 \vrule width 150pt height 0.4pt depth 0pt\
  hskip-120pt
2 \ltjsetparameter{yjabaselineshift=0pt,
  yalbaselineshift=0pt}abc あいう          _____ abc あいう
3 \ltjsetparameter{yjabaselineshift=5pt,
  yalbaselineshift=2pt}abc あいう
```

Here the horizontal line in above is the baseline of a line.

There is an interesting side-effect: characters in different size can be vertically aligned center in a line, by setting two parameters appropriately. The following is an example (beware the value is not well tuned):

```
1 xyz 漢字
2 {\scriptsize
3 \ltjsetparameter{yjabaselineshift=-1pt,
4 yalbaselineshift=-1pt}          xyz 漢字 XYZ ひらがな abc かな
5 XYZ ひらがな
6 }abc かな
```

---

<sup>\*1</sup> But we don’t recommend this: since numbers 1 and 2 have opposite meanings in jaxspmode and alxspmode.

## 3.5 トンボ

Cropmark is a mark for indicating 4 corners and horizontal/vertical center of the paper. In Japanese, we call cropmark as tomo(w). pL<sup>A</sup>T<sub>E</sub>X and this LuaT<sub>E</sub>X-ja support ‘tombow’ by their kernel. The following steps are needed to typeset cropmark:

1. First, define the banner which will be printed at the upper left of the paper. This is done by assigning a token list to `\@bannertoken`.

For example, the following sets banner as ‘filename (2012-01-01 17:01)’:

```
\makeatletter

\hour\time \divide\hour by 60 \@tempcnta\hour \multiply\@tempcnta 60\relax
\minute\time \advance\minute-\@tempcnta
\@bannertoken{%
  \jobname\space(\number\year-\two@digits\month-\two@digits\day
  \space\two@digits\hour:\two@digits\minute)}%
```

2. ...

## 第 II 部

# リファレンス

## 4 フォントメトリックと日本語フォント

### 4.1 `\jfont` プリミティブ

To load a font as a Japanese font, you must use the `\jfont` primitive instead of `\font`, while `\jfont` admits the same syntax used in `\font`. LuaT<sub>E</sub>X-ja automatically loads `luaotfload` package, so TrueType/OpenType fonts with features can be used for Japanese fonts:

```
1 \jfont\tradgt={file:ipaexg.ttf:script=latn;%
2 +trad;-kern;jfm=ujis} at 14pt
3 \tradgt{}当 / 体 / 医 / 区
```

當 / 體 / 醫 / 區

Note that the defined control sequence (`\tradgt` in the example above) using `\jfont` is not a `font_def` token, hence the input like `\fontname\tradgt` causes a error. We denote control sequences which are defined in `\jfont` by `<jfont_cs>`.

**Prefix psft** Besides `file:` and `name:` prefixes, `psft:` can be used a prefix in `\jfont` (and `\font`) primitive. Using this prefix, you can specify a ‘name-only’ Japanese font which will be not embedded to PDF. Typical use of this prefix is to specify the ‘standard’ Japanese fonts, namely, ‘Ryumin-Light’ and ‘GothicBBB-Medium’. For kerning or other information, that of Kozuka Mincho Pr6N Regular (this is a font by Adobe Inc., and included in Japanese Font Packs for Adore Reader) will be used.

**JFM** As noted in Introduction, a JFM has measurements of characters and glues/kerns that are automatically inserted for Japanese typesetting. The structure of JFM will be described in the next subsection. At the calling

表 4. Differences between JFMs shipped with LuaTeX-ja

	jfm-ujis.lua	jfm-jis.lua	jfm-min.lua
Example 1	ある日モモちゃん がお使いで迷 子になって泣き ました。	ある日モモちゃん がお使いで迷 子になって泣き ました。	ある日モモちゃん がお使いで迷 子になって泣き ました。
Example 2	ちょっと！何	ちょっと！何	ちょっと！何
Bounding Box			

of `\jfont` primitive, you must specify which JFM will be used for this font by the following keys:

`jfm=<name>` Specify the name of JFM. A file named `jfm-<name>.lua` will be searched and/or loaded.

The followings are JFMs shipped with LuaTeX-ja:

`jfm-ujis.lua` A standard JFM in LuaTeX-ja. This JFM is based on `upnmlminr-h.tfm`, a metric for UTF/OTF package that is used in `upTeX`. When you use the `luatexja-otf` package, please use this JFM.

`jfm-jis.lua` A counterpart for `jis.tfm`, ‘JIS font metric’ which is widely used in `pTeX`. A major difference of `jfm-ujis.lua` and this `jfm-jis.lua` is that most characters under `jfm-ujis.lua` are square-shaped, while that under `jfm-jis.lua` are horizontal rectangles.

`jfm-min.lua` A counterpart for `min10.tfm`, which is one of the default Japanese font metric shipped with `pTeX`. There are notable difference between this JFM and other 2 JFMs, as shown in Table 4.

`jfmvar=<string>` Sometimes there is a need that

**Note: kern feature** Some fonts have information for inter-glyph spacing. However, this information is not well-compatible with LuaTeX-ja. More concretely, this kerning space from this information are inserted *before* the insertion process of **JAgglue**, and this causes incorrect spacing between two characters when both a glue/kern from the data in the font and it from JFM are present.

- You should specify `-kern` in `\jfont` primitive, when you want to use other font features, such as `script=...`
- If you want to use Japanese fonts in proportional width, and use information from this font, use `jfm-prop.lua` for its JFM, and ...  
TODO: kanjiskip?

## 4.2 JFM ファイルの構造

A JFM file is a Lua script which has only one function call:

```
luatexja.jfont.define_jfm { ... }
```

Real data are stored in the table which indicated above by `{ ... }`. So, the rest of this subsection are devoted to describe the structure of this table. Note that all lengths in a JFM file are floating-point numbers in design-size unit.

`dir`= $\langle direction \rangle$  (required)

The direction of JFM. At the present, only 'yoko' is supported.

`zw`= $\langle length \rangle$  (required)

The amount of the length of the 'full-width'.

`zh`= $\langle length \rangle$  (required)

`kanjiskip`={ $\langle natural \rangle$ ,  $\langle stretch \rangle$ ,  $\langle shrink \rangle$ } (optional)

This field specifies the 'ideal' amount of `kanjiskip`. As noted in Subsection 3.2, if the parameter `kanjiskip` is `\maxdimen`, the value specified in this field is actually used (if this field is not specified in JFM, it is regarded as 0pt). Note that  $\langle stretch \rangle$  and  $\langle shrink \rangle$  fields are in design-size unit too.

`xkanjiskip`={ $\langle natural \rangle$ ,  $\langle stretch \rangle$ ,  $\langle shrink \rangle$ } (optional)

Like the `kanjiskip` field, this field specifies the 'ideal' amount of `xkanjiskip`.

Besides from above fields, a JFM file have several sub-tables those indices are natural numbers. The table indexed by  $i \in \omega$  stores information of 'character class'  $i$ . At least, the character class 0 is always present, so each JFM file must have a sub-table whose index is [0]. Each sub-table (its numerical index is denoted by  $i$ ) has the following fields:

`chars`={ $\langle character \rangle$ , ...} (required except character class 0)

This field is a list of characters which are in this character type  $i$ . This field is not required if  $i = 0$ , since all **J**Achar which are not in any character class other than 0 (hence, the character class 0 contains most of **J**Achars). In the list, a character can be specified by its code number, or by the character itself (as a string of length 1). Moreover, there are 'imaginary characters' which specified in the list. We will describe these later.

`width`= $\langle length \rangle$ , `height`= $\langle length \rangle$ , `depth`= $\langle length \rangle$ , `italic`= $\langle length \rangle$  (required)

Specify width of characters in character class  $i$ , height, depth and the amount of italic correction. All characters in character class  $i$  are regarded that its width, height and depth are as values of these fields. But there is one exception: if 'prop' is specified in `width` field, width of a character becomes that of its 'real' glyph

`left`= $\langle length \rangle$ , `down`= $\langle length \rangle$ , `align`= $\langle align \rangle$

These fields are for adjusting the position of the 'real' glyph. Legal values of `align` field are 'left', 'middle' and 'right'. If one of these 3 fields are omitted, `left` and `down` are treated as 0, and `align` field is treated as 'left'. The effects of these 3 fields are indicated in Figure 1.

In most cases, `left` and `down` fields are 0, while it is not uncommon that the `align` field is 'middle' or 'right'. For example, setting the `align` field to 'right' is practically needed when the current character class is the class for opening delimiters'.

`kern`=[ $j$ ]= $\langle kern \rangle$ , ...}

`glue`=[ $j$ ]={ $\langle width \rangle$ ,  $\langle stretch \rangle$ ,  $\langle shrink \rangle$ }, ...}

上で説明した通り, `chars` フィールド中にはいくつかの「特殊文字」も指定可能である. これらは, 大半が pTeX の JFM グルーの挿入処理ではみな「文字クラス 0 の文字」として扱われていた文字であり, その結果として pTeX より細かい組版調整ができるようになっている. 以下のその一覧を述べる:

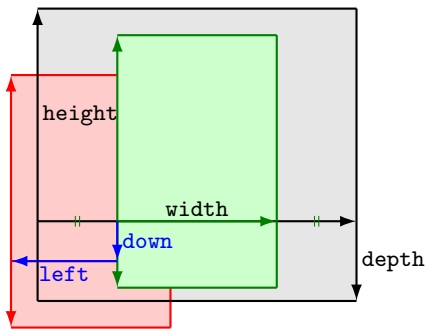
'lineend' 行の終端を表す.

'diffmet'

'boxbdd' hbox の先頭と末尾, 及びインデントされていない (`\noindent` で開始された) 段落の先頭を表す.

'parbdd' 通常の (`\noindent` で開始されていない) 段落の先頭.

'jcharbdd' 和文文字と「その他のもの」(欧文文字, `glue`, `kern` 等) との境界.



Consider a node containing Japanese character whose value of the `align` field is 'middle'.

- The black rectangle is a frame of the node. Its width, height and depth are specified by JFM.
- Since the `align` field is 'middle', the 'real' glyph is centered horizontally (the green rectangle).
- Furthermore, the glyph is shifted according to values of fields `left` and `down`. The ultimate position of the real glyph is indicated by the red rectangle.

図 1. The position of the 'real' glyph.

#### -1 行中数式と地の文との境界 .

pTeX 用和文フォントメトリックの移植 以下に, pTeX 用和文フォントメトリックを LuaTeX-ja 用に移植する場合の注意点を挙げておく .

- 実際に出力される和文フォントのサイズが design size となる . このため, 例えば `1zw` が design size の 0.962216 倍である JIS フォントメトリック等を移植する場合は,
  - JFM 中の全ての数値を 1/0.962216 倍しておく .
  - TeX ソース中で使用するところで, サイズ指定を 0.962216 倍にする . LaTeX でのフォント宣言なら, 例えば次のように :

```
\DeclareFontShape{JY3}{mc}{m}{n}{<-> s*[0.962216] psft:Ryumin-Light:jfm=jis}{}
```

- 上に述べた特殊文字は, 'boxbdd' を除き文字クラスを全部 0 とする (JFM 中に単に書かなければよい).
- 'boxbdd' については, それのみで一つの文字クラスを形成し, その文字クラスに関しては glue/kern の設定はしない .

これは, pTeX では, hbox の先頭・末尾とインデントされていない (`\noindent` で開始された) 段落の先頭には JFM グルーは入らないという仕様を実現させるためである .

- pTeX の組版を再現させようというのが目的であれば以上の注意を守れば十分である .  
ところで, pTeX では通常の段落の先頭に JFM グルーが残るという仕様があるので, 段落先頭の開き括弧は全角二分下がりになる . 全角下がりを実現させるには, 段落の最初に手動で `\inhibitglue` を追加するか, あるいは `\everypar` の hack を行い, それを自動化させるしかなかった .

一方, LuaTeX-ja では, 'parbdd' によって, それが JFM 側で調整できるようになった . 例えば, LuaTeX-ja 同梱の JFM のように, 'boxbdd' と同じ文字クラスに 'parbdd' を入れれば全角下がりとなる .

```
1 \jfont\g=psft:Ryumin-Light:jfm=test \g
2 \parindent1\zw\noindent{}
3 \par{}「 二分下がり
4 \par{}【 全角下がり
5 \par{}{ 全角二分下がり
```

### 4.3 数式フォントファミリ

TeX handles fonts in math formulas by 16 font families<sup>\*2</sup>, and each family has three fonts: `\textfont`, `\scriptfont` and `\scriptscriptfont`.

<sup>\*2</sup> Omega, Aleph, LuaTeX and  $\varepsilon$ (u)pTeX can handles 256 families, but an external package is needed to support this in plain TeX and LaTeX.



表 5. Primitives for Japanese math fonts.

	Japanese fonts	alphabetic fonts
font family	<code>\jfam ∈ [0, 256)</code>	<code>\fam</code>
text size	<code>jatextfont={⟨jfam⟩, ⟨jfont_cs⟩}</code>	<code>\textfont⟨fam⟩=⟨font_cs⟩</code>
script size	<code>jascriptfont={⟨jfam⟩, ⟨jfont_cs⟩}</code>	<code>\scriptfont⟨fam⟩=⟨font_cs⟩</code>
scriptscript size	<code>jascriptscriptfont={⟨jfam⟩, ⟨jfont_cs⟩}</code>	<code>\scriptscriptfont⟨fam⟩=⟨font_cs⟩</code>

LuaTeX-jā's handling of Japanese fonts in math formulas is similar; Table 5 shows counterparts to TeX's primitives for math font families. There is no relation between the value of `\fam` and that of `\jfam`; with appropriate settings, you can set both `\fam` and `\jfam` to the same value.

#### 4.4 コールバック

Like LuaTeX itself, LuaTeX-jā also has callbacks. These callbacks can be accessed via `luatexbase.add_to_callback` function and so on, as other callbacks

**luatexja.load\_jfm callback** With this callback you can overwrite JFMs.

```
function (<table> jfm_info, <string> jfm_name)
  return <table> new_jfm_info
end
```

The argument `jfm_info` contains a table similar to the table in a JFM file, except this argument has `chars` field which contains character codes whose character class is not 0.

An example of this callback is the `ltjarticle` class, with forcefully assigning character class 0 to 'parbdd' in the JFM `jfm-min.lua`. This callback doesn't replace any code of LuaTeX-jā.

**luatexja.define\_font callback** This callback and the next callback form a pair, and you can assign letters which don't have fixed code points in Unicode to non-zero character classes. This `luatexja.define_font` callback is called just when new Japanese font is loaded.

```
function (<table> jfont_info, <number> font_number)
  return <table> new_jfont_info
end
```

You may assume that `jfont_info` has the following fields:

`jfm` The index number of JFM.

`size` Font size in a scaled point (=  $2^{-16}$  pt).

`var` The value specified in `jfmvar=...` at a call of `\jfont`.

The returned table `new_jfont_info` also should include these three fields. The `font_number` is a font number.

A good example of this and the next callbacks is the `luatexja-otf` package, supporting "AJ1-xxx" form for Adobe-Japan1 CID characters in a JFM. This callback doesn't replace any code of LuaTeX-jā.

**luatexja.find\_char\_class callback** This callback is called just when LuaTeX-jā inready to determine which character class a character `chr_code` belongs. A function used in this callback should be in the following form:

```
1 function (<number> char_class, <table> jfont_info, <number> chr_code)
```

```

2   if char_class~=0 then return char_class
3   else
4     ....
5     return (<number> new_char_class or 0)
6   end
7 end

```

The argument `char_class` is the result of LuaTeX-ja's default routine or previous function calls in this callback, hence this argument may not be 0. Moreover, the returned `new_char_class` should be as same as `char_class` when `char_class` is not 0, otherwise you will overwrite the LuaTeX-ja's default routine. This callback doesn't replace any code of LuaTeX-ja.

## 5 パラメータ

### 5.1 `\ltjsetparameter` プリミティブ

As noted before, `\ltjsetparameter` and `\ltjgetparameter` are primitives for accessing most parameters of LuaTeX-ja. One of the main reason that LuaTeX-ja didn't adopted the syntax similar to that of pTeX (*e.g.*, `\prebreakpenalty`)=10000`) is the position of `hpack_filter` callback in the source of LuaTeX, see Section 9.

`\ltjsetparameter` and `\ltjglobalsetparameter` are primitives for assigning parameters. These take one argument which is a *<key>=<value>* list. Allowed keys are described in the next subsection. The difference between `\ltjsetparameter` and `\ltjglobalsetparameter` is only the scope of assignment; `\ltjsetparameter` does a local assignment and `\ltjglobalsetparameter` does a global one. They also obey the value of `\globaldefs`, like other assignment.

`\ltjgetparameter` is the primitive for acquiring parameters. It always takes a parameter name as first argument, and also takes the additional argument—a character code, for example—in some cases.

```

1 \ltjgetparameter{differentjfm},
2 \ltjgetparameter{autospacing},           average, 1, 10000.
3 \ltjgetparameter{prebreakpenalty}{`})}.

```

*The return value of `\ltjgetparameter` is always a string.* This is outputted by `tex.write()`, so any character other than space ' ' (U+0020) has the category code 12 (other), while the space has 10 (space).

### 5.2 パラメーター一覧

The following is the list of parameters which can be specified by the `\ltjsetparameter` command. `[<cs>]` indicates the counterpart in pTeX, and symbols beside each parameter has the following meaning:

- No mark: values at the end of the paragraph or the hbox are adopted in the whole paragraph/hbox.
- '\*': local parameters, which can change everywhere inside a paragraph/hbox.
- '†': assignments are always global.

`jcharwidowpenalty=<penalty> [jcharwidowpenalty]`

Penalty value for suppressing orphans. This penalty is inserted just after the last **J**Achar which is not regarded as a (Japanese) punctuation mark.

`kcatcode={<chr_code>,<natural number>}`

An additional attributes having each character whose character code is  $\langle chr\_code \rangle$ . At the present version, the lowermost bit of  $\langle natural\ number \rangle$  indicates whether the character is considered as a punctuation mark (see the description of `jcharwidowpenalty` above).

`prebreakpenalty={ $\langle chr\_code \rangle$ , $\langle penalty \rangle$ }` [`\prebreakpenalty`] 文字コード  $\langle chr\_code \rangle$  の **JA**char が行頭にくることを抑止するために、この文字の前に挿入/追加されるペナルティの量を指定する。

例えば閉じ括弧「`】`」は絶対に行頭にきてはならないので、標準で読み込まれる `luatexja-kinsoku.tex` において

```
\ltjsetparameter{prebreakpenalty={`】,10000}}
```

と、最大値の 10000 が指定されている。他にも、小書きのカナなど、絶対禁止というわけではないができれば行頭にはきて欲しくない場合に、0 と 10000 の間の値を指定するのも有用であろう。

```
\ltjsetparameter{prebreakpenalty={`か,150}}
```

`postbreakpenalty={ $\langle chr\_code \rangle$ , $\langle penalty \rangle$ }` [`\postbreakpenalty`] 文字コード  $\langle chr\_code \rangle$  の **JA**char が行末にくることを抑止するために、この文字の後に挿入/追加されるペナルティの量を指定する。

pTeX では、`\prebreakpenalty`、`\postbreakpenalty` において、

- 一つの文字に対して、`pre`、`post` どちらか一つしか指定することができなかった（後から指定した方で上書きされる）。
- `pre`、`post` 合わせて 256 文字分の情報を格納することしかできなかった。

という制限があったが、`LuaTeX-ja` ではこれらの制限は解消されている。

`jatextfont={ $\langle jfam \rangle$ , $\langle jfont\_cs \rangle$ }` [`\textfont` in TeX]

`jascriptfont={ $\langle jfam \rangle$ , $\langle jfont\_cs \rangle$ }` [`\scriptfont` in TeX]

`jascriptscriptfont={ $\langle jfam \rangle$ , $\langle jfont\_cs \rangle$ }` [`\scriptscriptfont` in TeX]

`yjabaselineshift= $\langle dimen \rangle$ *`

`yalbaselineshift= $\langle dimen \rangle$ *` [`\ybaselineshift`]

`jaxspmode={ $\langle chr\_code \rangle$ , $\langle mode \rangle$ }` [`\inhibitxspcode`]

Setting whether inserting `xkanjiskip` is allowed before/after a **JA**char whose character code is  $\langle chr\_code \rangle$ .

The followings are allowed for  $\langle mode \rangle$ :

0, `inhibit` Insertion of `xkanjiskip` is inhibited before the character, nor after the character.

2, `preonly` Insertion of `xkanjiskip` is allowed before the character, but not after.

1, `postonly` Insertion of `xkanjiskip` is allowed after the character, but not before.

3, `allow` Insertion of `xkanjiskip` is allowed before the character and after the character. This is the default value.

`alxspmode={ $\langle chr\_code \rangle$ , $\langle mode \rangle$ }` [`\xspcode`]

Setting whether inserting `xkanjiskip` is allowed before/after a **AL**char whose character code is  $\langle chr\_code \rangle$ .

The followings are allowed for  $\langle mode \rangle$ :

0, `inhibit` Insertion of `xkanjiskip` is inhibited before the character, nor after the character.

1, `preonly` Insertion of `xkanjiskip` is allowed before the character, but not after.

2, `postonly` Insertion of `xkanjiskip` is allowed after the character, but not before.

3, `allow` Insertion of `xkanjiskip` is allowed both before the character and after the character. This is the default value.

Note that parameters `jaxspmode` and `alxspmode` use a common table.

`autospacing= $\langle bool \rangle$ *` [`\autospacing`]

`autoxspacing= $\langle bool \rangle$ *` [`\autoxspacing`]

`kanjiskip= $\langle skip \rangle$`  [`\kanjiskip`]

`xkanjiskip= $\langle skip \rangle$`  [`\xkanjiskip`]

`differentjfm=mode`<sup>†</sup> Specify how glues/kerns between two **J**Achars whose JFM (or size) are different. The allowed arguments are the followings:

```
average
both
large
small
```

`jacharrange=ranges*`

`kansujichar={digit, chr_code} [\kansujichar]`

## 6 その他のプリミティブ

### 6.1 互換プリミティブ

The following primitives are implemented for compatibility with p<sub>T</sub>E<sub>X</sub>:

```
\kuten
\jis
\euc
\sjis
\ucs
\kansuji
```

### 6.2 \inhibitglue プリミティブ

The primitive `\inhibitglue` suppresses the insertion of **J**Aglue. The following is an example, using a special JFM that there will be a glue between the beginning of a box and ‘あ’, and also between ‘あ’ and ‘ウ’.

1 <code>\jfont\g=psft:Ryumin-Light:jfm=test \g</code>	あ	ウあウ
2 <code>あウあ\inhibitglue{}ウ\inhibitglue\par</code>	あ	
3 <code>あ\par\inhibitglue{}あ</code>	あ	
4 <code>\par\inhibitglue\hrule{}あoff\inhibitglue ice</code>	あ	office

With the help of this example, we remark the specification of `\inhibitglue`:

- The call of `\inhibitglue` in the (internal) vertical mode is effective at the beginning of the next paragraph. This is realized by hacking `\everypar`.
- The call of `\inhibitglue` in the (restricted) horizontal mode is only effective on the spot; does not get over boundary of paragraphs. Moreover, `\inhibitglue` cancels ligatures and kernings, as shown in line 4 of above example.
- The call of `\inhibitglue` in math mode is just ignored.

## 7 L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> 用のコントロールシーケンス

### 7.1 NFSS2 へのパッチ

As described in Subsection 2.4, Lua<sub>T</sub>E<sub>X</sub>-ja simply adopted `plfonts.dtx` in p<sub>L</sub>A<sub>T</sub>E<sub>X</sub> 2<sub>ε</sub> for the Japanese patch for NFSS2. For an convenience, we will describe commands which are not described in Subsection 2.5.

`\DeclareYokoKanjiEncoding{⟨encoding⟩}{⟨text-settings⟩}{⟨math-settings⟩}`

In NFSS2 under LuaTeX-ja, distinction between alphabetic font families and Japanese font families is only made by its encoding. For example, encodings OT1 and T1 are for alphabetic font families, and a Japanese font family cannot have these encodings. This command defines a new encoding scheme for Japanese font family (in horizontal direction).

`\DeclareKanjiEncodingDefaults{⟨text-settings⟩}{⟨math-settings⟩}`

`\DeclareKanjiSubstitution{⟨encoding⟩}{⟨family⟩}{⟨series⟩}{⟨shape⟩}`

`\DeclareErrorKanjiFont{⟨encoding⟩}{⟨family⟩}{⟨series⟩}{⟨shape⟩}{⟨size⟩}`

The above 3 commands are just the counterparts for `\DeclareFontEncodingDefaults` and others.

`\reDeclareMathAlphabet{⟨unified-cmd⟩}{⟨al-cmd⟩}{⟨ja-cmd⟩}`

和文・欧文の数式用フォントファミリーを一度に変更する命令を作成する．具体的には，欧文数式用フォントファミリー変更の命令 `⟨al-cmd⟩` と，和文数式用フォントファミリー変更の命令 `⟨ja-cmd⟩` の2つを同時に行う命令として `⟨unified-cmd⟩` を（再）定義する．実際の使用では `⟨unified-cmd⟩` と `⟨al-cmd⟩` に同じものを指定する，すなわち，`⟨al-cmd⟩` に和文側も変更させるようにするのが一般的と思われる．

本コマンドの使用については，pL<sup>A</sup>T<sub>E</sub>X 配布中の `plfonts.dtx` に詳しく注意点が述べられているので，そちらを参照されたい．

`\DeclareRelationFont{⟨ja-encoding⟩}{⟨ja-family⟩}{⟨ja-series⟩}{⟨ja-shape⟩}`  
`{⟨al-encoding⟩}{⟨al-family⟩}{⟨al-series⟩}{⟨al-shape⟩}`

いわゆる「従属欧文」を設定するための命令である．前半の4引数で表される和文フォントファミリーに対して，そのフォントに対応する「従属欧文」フォントファミリーを後半の4引数により与える．

`\SetRelationFont`

This command is almost same as `\DeclareRelationFont`, except that this command does a local assignment, where `\DeclareRelationFont` does a global assignment.

`\userelfont`

Change current alphabetic font encoding/family/... to the ‘accompanied’ alphabetic font family with respect to current Japanese font family, which was set by `\DeclareRelationFont` or `\SetRelationFont`. Like `\fontfamily`, `\selectfont` is required to take an effect.

`\adjustbaseline`

...

`\fontfamily{⟨family⟩}`

As in L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>, this command changes current font family (alphabetic, Japanese, or both) to `⟨family⟩`. Which family will be changed is determined as follows:

- Let current encoding scheme for Japanese fonts be `⟨ja-enc⟩`. Current Japanese font family will be changed to `⟨family⟩`, if one of the following two conditions is met:
  - The family `⟨fam⟩` under the encoding `⟨ja-enc⟩` is already defined by `\DeclareKanijFamily`.
  - A font definition named `⟨enc⟩⟨ja-enc⟩.fd` (the file name is all lowercase) exists.
- Let current encoding scheme for Japanese fonts be `⟨al-enc⟩`. For alphabetic font family, the criterion as above is used.
- There is a case which none of the above applies, that is, the font family named `⟨family⟩` doesn’t seem to be defined neither under the encoding `⟨ja-enc⟩`, nor under `⟨al-enc⟩`.

In this case, the default family for font substitution is used for alphabetic and Japanese fonts. Note that current encoding will not be set to  $\langle family \rangle$ , unlike the original implementation in L<sup>A</sup>T<sub>E</sub>X.

As closing this subsection, we shall introduce an example of `SetRelationFont` and `\userelfont`:

```
1 \gtfamily{}あいうabc
2 \SetRelationFont{JY3}{gt}{m}{n}{OT1}{pag}{m}{
  n}          あいう abc あいう abc
3 \userelfont\selectfont{}あいうabc
```

## 7.2 トンボ

# 8 拡張

## 8.1 luatexja-fontspec.sty

## 8.2 luatexja-otf.sty

This optional package supports typesetting characters in Adobe-Japan1. `luatexja-otf.sty` offers the following 2 low-level commands:

`\CID{ $\langle number \rangle$ }` Typeset a character whose CID number is  $\langle number \rangle$ .

`\UTF{ $\langle hex\_number \rangle$ }` Typeset a character whose character code is  $\langle hex\_number \rangle$  (in hexadecimal). This command is similar to `\char"  $\langle hex\_number \rangle$` , but please remind remarks below.

Remarks Characters by `\CID` and `\UTF` commands are different from ordinary characters in the following points:

- Always treated as **J**Achars.
- Processing codes for supporting OpenType features (*e.g.*, glyph replacement and kerning) by the `luaotfload` package is not performed to these characters.

Additionally Syntax of JFM `luatexja-otf.sty` extends the syntax of JFM; the entries of `chars` table in JFM now allows a string in the form 'AJ1-xxx', which stands for the character whose CID number in Adobe-Japan1 is xxx.

## 第 III 部

# 実装

## 9 パラメータの保持

### 9.1 用いられる寸法レジスタ, 属性レジスタ, whatsit ノード

Here the following is the list of dimensions and attributes which are used in LuaT<sub>E</sub>X-ja.

`\jQ` (dimension) As explained in Subsection 2.3, `\jQ` is equal to  $1\text{Q} = 0.25\text{mm}$ , where 'Q' (also called '級') is a unit used in Japanese phototypesetting. So one should not change the value of this dimension.

`\jH` (dimension) There is also a unit called ‘`齒`’ which equals to 0.25 mm and used in Japanese phototypesetting. The dimension `\jH` stores this length, similar to `\jQ`.

`\ltj@zw` (dimension) A temporal register for the ‘full-width’ of current Japanese font.

`\ltj@zh` (dimension) A temporal register for the ‘full-height’ (usually the sum of height of imaginary body and its depth) of current Japanese font.

`\jfam` (attribute) Current number of Japanese font family for math formulas.

`\ltj@curjfnt` (attribute) The font index of current Japanese font.

`\ltj@charclass` (attribute) The character class of Japanese *glyph\_node*.

`\ltj@yablshift` (attribute) The amount of shifting the baseline of alphabetic fonts in scaled point ( $2^{-16}$  pt).

`\ltj@ykblshift` (attribute) The amount of shifting the baseline of Japanese fonts in scaled point ( $2^{-16}$  pt).

`\ltj@autospc` (attribute) Whether the auto insertion of `kanjiskip` is allowed at the node.

`\ltj@autoxspc` (attribute) Whether the auto insertion of `xkanjiskip` is allowed at the node.

`\ltj@icflag` (attribute) An attribute for distinguishing ‘kinds’ of a node. One of the following value is assigned to this attribute:

- italic* (1) Glues from an italic correction (`\/`). This distinction of origins of glues (from explicit `\kern`, or from `\/`) is needed in the insertion process of `xkanjiskip`.
- packed* (2)
- kinsoku* (3) Penalties inserted for the word-wrapping process of Japanese characters (*kinsoku*).
- from\_jfm* (4) Glues/kerns from JFM.
- line\_end* (5) Kerns for ...
- kanji\_skip* (6) Glues for `kanjiskip`.
- xkanji\_skip* (7) Glues for `xkanjiskip`.
- processed* (8) Nodes which is already processed by ...
- ic\_processed* (9) Glues from an italic correction, but also already processed.
- boxbdd* (15) Glues/kerns that inserted just the beginning or the ending of an hbox or a paragraph.

`\ltj@kcati` (attribute) Where *i* is a natural number which is less than 7. These 7 attributes store bit vectors indicating which character block is regarded as a block of **J**Achars.

Furthermore, Lua $\TeX$ -ja uses several ‘user-defined’ whatsit nodes for typesetting. All those nodes store a natural number (hence the node’s `type` is 100).

- 30111 Nodes for indicating that `\inhibitglue` is specified. The `value` field of these nodes doesn’t matter.
- 30112 Nodes for Lua $\TeX$ -ja’s stack system (see the next subsection). The `value` field of these nodes is current group.
- 30113 Nodes for Japanese Characters which the callback process of `luaotfload` won’t be applied, and the character code is stored in the `value` field. Each node having this `user_id` is converted to a ‘`glyph_node`’ after the callback process of `luaotfload`.

These whatsits will be removed during the process of inserting **J**Aglues.

## 9.2 Lua $\TeX$ -ja のスタックシステム

**Background** Lua $\TeX$ -ja has its own stack system, and most parameters of Lua $\TeX$ -ja are stored in it. To clarify the reason, imagine the parameter `kanjiskip` is stored by a skip, and consider the following source:

```

1 \ltjsetparameter{kanjiskip=0pt}ふがふが.%
2 \setbox0=\hbox{\ltjsetparameter{kanjiskip=5pt
   }ほげほげ}      ふがふが. ほ げ ほ げ. ぴよぴよ
3 \box0. ぴよぴよ\par

```

As described in Part II, the only effective value of `kanjiskip` in an `hbox` is the latest value, so the value of `kanjiskip` which applied in the entire `hbox` should be 5pt. However, by the implementation method of `LuaTeX`, this ‘5pt’ cannot be known from any callbacks. In the `tex/packaging.w` (which is a file in the source of `LuaTeX`), there are the following codes:

```

void package(int c)
{
    scaled h;          /* height of box */
    halfword p;        /* first node in a box */
    scaled d;          /* max depth */
    int grp;
    grp = cur_group;
    d = box_max_depth;
    unsave();
    save_ptr -= 4;
    if (cur_list.mode_field == -hmode) {
        cur_box = filtered_hpack(cur_list.head_field,
                                cur_list.tail_field, saved_value(1),
                                saved_level(1), grp, saved_level(2));
        subtype(cur_box) = HLIST_SUBTYPE_HBOX;
    }
}

```

Notice that `unsave` is executed *before* `filtered_hpack` (this is where `hpack_filter` callback is executed): so ‘5pt’ in the above source is orphaned at `+unsave+`, and hence it can’t be accessed from `hpack_filter` callback.

The method The code of stack system is based on that in a post of Dev-luatex mailing list<sup>\*3</sup>.

These are two `TeX` count registers for maintaining information: `\ltj@@stack` for the stack level, and `\ltj@@group@level` for the `TeX`’s group level when the last assignment was done. Parameters are stored in one big table named `charprop_stack_table`, where `charprop_stack_table[i]` stores data of stack level  $i$ . If a new stack level is created by `\ltjsetparameter`, all data of the previous level is copied.

To resolve the problem mentioned in ‘Background’ above, `LuaTeX-ja` uses another thing: When a new stack level is about to be created, a `whatsit` node whose type, subtype and value are 44 (`user_defined`), 30112, and current group level respectively is appended to the current list (we refer this node by `stack_flag`). This enables us to know whether assignment is done just inside a `hbox`. Suppose that the stack level is  $s$  and the `TeX`’s group level is  $t$  just after the `hbox` group, then:

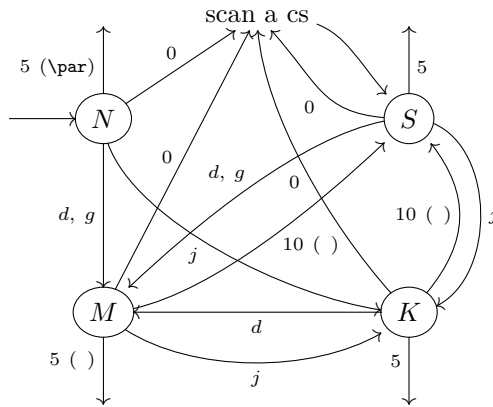
- If there is no `stack_flag` node in the list of `hbox`, then no assignment was occurred inside the `hbox`. Hence values of parameters at the end of the `hbox` are stored in the stack level  $s$ .
- If there is a `stack_flag` node whose value is  $t + 1$ , then an assignment was occurred just inside the `hbox` group. Hence values of parameters at the end of the `hbox` are stored in the stack level  $s + 1$ .
- If there are `stack_flag` nodes but all of their values are more than  $t + 1$ , then an assignment was occurred in the box, but it is done in ‘more internal’ group. Hence values of parameters at the end of the `hbox` are stored in the stack level  $s$ .

Note that to work this trick correctly, assignments to `\ltj@@stack` and `\ltj@@group@level` have to be local

---

<sup>\*3</sup> [Dev-luatex] `tex.currentgrouplevel`, a post at 2008/8/19 by Jonathan Sauer.





$d := \{3, 4, 6, 7, 8, 11, 12, 13\}$ ,  $g := \{1, 2\}$ ,  $j := (\text{Japanese characters})$

- Numbers represent category codes.
- Category codes 9 (ignored), 14 (comment) and 15 (invalid) are omitted in above diagram.

図 2. State transitions of pTeX's input processor.

always, regardless the value of `\globaldefs`. This problem is resolved by using `\directlua{tex.globaldefs=0}` (this assignment is local).

## 10 和文文字直後の改行

### 10.1 参考: pTeX の挙動

欧文では文章の改行は単語間でしか行わない。そのため、TeX では、(文字の直後の)改行は空白文字と同じ扱いとして扱われる。一方、和文ではほとんどどこでも改行が可能のため、pTeX では和文文字の直後の改行は単純に無視されるようになっている。

このような動作は、pTeX が TeX からエンジンとして拡張されたことによって可能になったことである。pTeX の入力処理部は、TeX におけるそれと同じように、有限オートマトンとして記述することができ、以下に述べるような 4 状態を持っている。

- State  $N$ : 行の開始。
- State  $S$ : 空白読み飛ばし。
- State  $M$ : 行中。
- State  $K$ : 行中 (和文文字の後)。

また、状態遷移は、図 2 のようになっており、図中の数字はカテゴリーコードを表している。最初の 3 状態は TeX の入力処理部と同じであり、図中から状態  $K$  と「 $j$ 」と書かれた矢印を取り除けば、TeX の入力処理部と同じものになる。

この図から分かることは、

行が和文文字 (とグループ境界文字) で終わっていれば、改行は無視される

ということである。

## 10.2 LuaTeX-ja の挙動

LuaTeX の入力処理部は TeX のそれと全く同じであり、callback によりユーザがカスタマイズすることはできない。このため、改行抑制の目的でユーザが利用できそうな callback としては、`process_input_buffer` や `token_filter` に限られてしまう。しかし、TeX の入力処理部をよく見ると、後者も役には経たないことが分かる：改行文字は、入力処理部によってトークン化される時に、カテゴリコード 10 の 32 番文字へと置き換えられてしまうため、`token_filter` で非標準なトークン読み出しを行おうとしても、空白文字由来のトークンと、改行文字由来のトークンは区別できないのだ。

すると、我々のとれる道は、`process_input_buffer` を用いて LuaTeX の入力処理部に引き渡される前に入力文字列を編集するというものしかない。以上を踏まえ、LuaTeX-ja における「和文文字直後の改行抑制」の処理は、次のようになっている：

各入力行に対し、その入力行が読まれる前の内部状態で以下の 2 条件が満たされている場合、LuaTeX-ja は U+FFFFF 番の文字<sup>\*4</sup>を末尾に追加する。よって、その場合に改行は空白とは見做されないこととなる。

1. 改行文字（文字コード 13 番）のカテゴリコードが 5 (end-of-line) である。
2. 入力行は次の「正規表現」にマッチしている：

$$(\text{any char})^*(\mathbf{JA}\text{char})(\{\text{catcode} = 1\} \cup \{\text{catcode} = 2\})^*$$

この仕様は、前節で述べた pTeX の仕様にてできるだけ近づけたものとなっている。最初の条件は、`verbatim` 系環境などの日本語対応マクロを書かなくてすませるためのものである。しかしながら、完全に同じ挙動が実現できたわけではない。差異は、次の例が示すように、和文文字の範囲を変更した行の改行において見られる：

```
1 \ltjsetparameter{autoxspacing=false}
2 \ltjsetparameter{jacharrange={-6}}x あ          xyzあ u
3 y\ltjsetparameter{jacharrange={+6}}z あ
4 u
```

もし pTeX とまったく同じ挙動を示すならば、出力は「x yzあu」となるべきである。しかし、実際には上のように異なる挙動となっている。

- 2 行目は「あ」という和文文字で終わる（2 行目を処理する前の時点では、「あ」は和文文字扱いである）ため、直後の改行文字は無視される。
- 3 行目は「あ」という欧文文字で終わる（2 行目を処理する前の時点では、「あ」は欧文文字扱いである）ため、直後の改行文字は空白に置き換わる。

このため、トラブルを避けるために、和文文字の範囲を `\ltjsetparameter` で編集した場合、その行はそこで改行するようにした方がいいだろう。

## 11 JFM グルーの挿入、`kanjiskip` と `xkanjiskip`

### 11.1 概要

LuaTeX-ja における和文処理グルーの挿入方法は、pTeX のそれとは全く異なる。pTeX では次のような仕様であった：

- JFM グルーの挿入は、和文文字を表すトークンを元に水平リストに（文字を表す）`\char_node` を追加する過程で行われる。

<sup>\*4</sup> この文字はコメント文字として扱われるように LuaTeX-ja 内部で設定をしている。

- xkanjiskip の挿入は、hbox へのパッケージングや行分割前に行われる。
- kanjiskip はノードとしては挿入されない。パッケージングや行分割の計算時に「和文文字を表す 2 つの  $\langle char\_node \rangle$  の間には kanjiskip がある」ものとみなされる。

しかし、LuaTeX-ja では、hbox へのパッケージングや行分割前に全ての JAglue、即ち JFM グルー・xkanjiskip・kanjiskip の 3 種類を一度に挿入することになっている。これは、LuaTeX において欧文の合字・カーニング処理がノードベースになったことに対応する変更である。

LuaTeX-ja における JAglue 挿入処理では、下の図??のように「塊」を単位にして行われる。大雑把にいうと、「塊」は文字とそれに付随するノード達（アクセント位置補正用の kern や、イタリック補正）をまとめたものであり、2 つの塊の間には、ペナルティ、\vadjust、whatsit など、行組版には関係しないものがある。そのため、.....

## 11.2 「クラスタ」の定義

**定義 1.** A *cluster* is a list of nodes in one of the following forms, with the *id* of it:

1. Nodes whose value of  $\backslash\text{tj@icflag}$  is in  $[3, 15)$ . These nodes come from a hbox which is already packaged, by unpackaging ( $\backslash\text{unhbox}$ ). The *id* is *id\_pbox*.
2. A inline math formula, including two *math\_nodes* at the boundary of it: HOGE The *id* is *id\_math*.
3. A *glyph\_node* with nodes which relate with it: HOGE The *id* is *id\_jglyph* or *id\_glyph*, according to whether the *glyph\_node* represents a Japanese character or not.
4. An box-like node, that is, an hbox, an vbox and an rule ( $\backslash\text{vrule}$ ). The *id* is *id\_hlist* if the node is an hbox which is not shifted vertically, or *id\_box\_like* otherwise.
5. A glue, a kern whose subtype is not 2 (*accent*), and a discretionary break. The *id* is *id\_glue*, *id\_kern* and *id\_disc*, respectively.

We denote a cluster by  $Np$ ,  $Nq$  and  $Nr$ .

Internally, a cluster is represented by a table  $Np$  with the following fields.

*first*, *last* The first/last node of the cluster.

*id* The *id* in above definition.

*nuc*

*auto\_kspc*, *auto\_xspc*

*xspc\_before*, *xspc\_after*

*pre*, *post*

*char*

*class*

*lend*

*met*, *var*