

LuaTeX-ja パッケージ

LuaTeX-ja プロジェクトチーム

2012年4月9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	pTeX からの主な変更点	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 TeX で使う	5
2.4	L ^A T _E X で使う	6
2.5	フォントの変更	7
2.6	fontspec	8
3	パラメータの変更	9
3.1	J ^A Char の範囲の設定	9
3.2	kanjiskip と xkanjiskip	11
3.3	xkanjiskip の設定の挿入	11
3.4	ベースラインの移動	12
3.5	トンボ	12
第II部	リファレンス	13
4	フォントメトリックと和文フォント	13
4.1	\jfont プリミティブ	13
4.2	psft プレフィックス	14
4.3	JFM ファイルの構造	14
4.4	数式フォントファミリ	17
4.5	コールバック	17
5	パラメータ	19
5.1	\ltjsetparameter プリミティブ	19
5.2	パラメーター一覧	19

6	その他のプリミティブ	21
6.1	互換プリミティブ	21
6.2	\inhibitglue プリミティブ	21
7	L ^A T _E X _{2ϵ} 用のコントロールシーケンス	21
7.1	NFSS2 へのパッチ	21
7.2	トンボ	23
8	拡張	23
8.1	luatexja-fontspec.sty	23
8.2	luatexja-otf.sty	23
第 III 部 実装		24
9	パラメータの保持	24
9.1	用いられる寸法レジスタ, 属性レジスタ, whatsit ノード	24
9.2	LuaT _E X-ja のスタックシステム	25
10	和文文字直後の改行	26
10.1	参考: pT _E X の動作	26
10.2	LuaT _E X-ja の動作	27
11	JFM グルーの挿入, kanjiskip と xkanjiskip	28
11.1	概要	28
11.2	「クラスタ」の定義	28
11.3	段落/hbox の先頭や末尾	30
11.4	概観と典型例: 2つの「和文 A」の場合	31
11.5	その他の場合	33
12	psft	37
参考文献		37

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-ja パッケージは、次世代標準 TeX である LuaTeX の上で、pTeX 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、「TeX 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 pTeX やその拡張物が用いられることが一般的であった。pTeX は TeX 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：pTeX 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 TeX の拡張例えば ϵ -TeX や pdfTeX 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 pTeX バイナリでは外部 UTF-8 入力が可能となり、さらに Unicode 化を推進し、pTeX の内部処理まで Unicode 化した upTeX も開発されている。また、pTeX に ϵ -TeX 拡張をマージした ϵ -pTeX も登場し、TeX Live 2011 では pLaTeX が ϵ -pTeX の上で動作するようになった。だが、pdfTeX 拡張 (PDF 直接出力や micro-typsetting) を pTeX 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ことになった。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 TeX 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を意味する。LuaTeX-ja は、このアプローチによって Lua コード・TeX 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 pTeX からの主な変更点

LuaTeX-ja は、pTeX に多大な影響を受けている。初期の開発目標は、pTeX 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、pTeX の完全な移植は不可能であり、また pTeX 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-ja は、もはや pTeX の完全な移植は目標とはしない。pTeX 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 pTeX からの主な変更点である。

- 和文フォントは(小塚明朝, IPA 明朝などの)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, pTeX 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-ja 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）、JAgglue 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 JAgglue 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{}つと   ちょ\つと
  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{}つと
```
- 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 L^AT_EX 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 | • 馬 起園 |

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のようになる :

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 の公式リリースはまだない . そのため、Git レポジトリを次のようにすることで取得する必要がある :

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

もしくは、master ブランチの HEAD のアーカイブを以下からダウンロードしてもよい :

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

master ブランチはたまにしか更新されない . 最新の開発中のコードは master ブランチには含まれていない場合がかなり多いので注意 .

2. 後者の方法でアーカイブを取得したならば、それを展開する . src/ をはじめとしたいくつかのディレクトリができるが、動作には src/ 以下の内容だけで十分 .
3. src/ の中身全てを自分の TEXMF ツリーにコピーする . シンボリックリンクが利用できる環境で、かつレポジトリを直接取得したのであれば、コピーではなくリンクを貼るのが適切だろう .
4. 必要があれば、mktexlsr を実行する .

2.2 注意点

- 原稿のソースファイルの文字コードは UTF-8 固定である . 従来日本語の文字コードとして用いられてきた EUC-JP や Shift-JIS は使用できない .

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	\tenmin	\sevenmin	\fivemin
ゴシック体	GothicBBB-Medium	\tengt	\seventgt	\fivegt

- ‘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、これらを引き続き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くデザインされている。そこで、標準では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。この 0.962216 という数値も、`pTeX` におけるスケールリングを踏襲した値である。
- `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^AT_EX で使う

L^AT_EX 2_ε L^AT_EX 2_ε 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^AT_EX 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_EX-ja 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	\mdseries	\bfseries	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^AT_EX 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、`jclasses`（pL^AT_EX 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^AT_EX では, JIS X 0208 がない 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_EX-ja 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\UTF{9DD7}外と内田百\UTF{9592}とが\UTF{9
   AD9}島屋に行く。
2
3 \CID{7652}飾区の\CID{13706}野家,
4 葛飾区の吉野家

```

森鷗外と内田百閒とが高島屋に行く。
葛飾区の吉野家, 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 pT_EX では, 特に何もしないで数式中に和文文字を記述することができた. そのため, 以下のようなソースが見られた:

```

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

```

$f_{\text{高温}}$ ($f_{\text{high temperature}}$).
 $y = (x - 1)^2 + 2$ よって $y > 0$
 $5 \in \text{素} := \{p \in \mathbb{N} : p \text{ is a prime}\}.$

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

```

$f_{\text{高温}}$ ($f_{\text{high temperature}}$).
 $y = (x - 1)^2 + 2$ よって $y > 0$
 $5 \in \text{素} := \{p \in \mathbb{N} : p \text{ is a prime}\}.$

また LuaT_EX-ja 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, したがってこの節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 T_EX plain T_EX 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^AT_EX 2_ε については, LuaT_EX-ja ではフォント選択システムを pL^AT_EX 2_ε (plfonts.dtx) の大部分を採用している.

- 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	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のエンコーディングを変更する。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を変更する。詳細は 7.1 節を参照すること。

- 和文フォントファミリの定義には `\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。

2.6 fontspec

`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>\newjfontfamily</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 \addjfontfeatures{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 JAGlue と衝突するためである (4.1 節を参照)。

注意 `luatexja-fontspec` パッケージでは、内部で `xunicode` パッケージの再読み込みを行っているが、**TeX Live 2011** 他に含まれている `xunicode` パッケージ (2011/09/09, v0.981) ではその再読み込みがうまく動作せず、例えば以下のようなパッチを当てる必要がある：

```
--- xunicode.sty.orig 2011-09-12 08:31:47.000000000 +0900
```

```
+++ xunicode.sty          2011-11-16 22:06:17.061413113 +0900
@@ -1475,7 +1475,11 @@
```

```
\newtoks\tipasavetokens
\newtoks\tipachecktokens
+
+\fi
\newif\iftipaonetoken
+\expandafter\ifx\curname ReloadXunicode\endcurname\relax
+
\def\tipalasttoken{!@! do nothing with this !@!}
\def\tipacatchonechar#1{\begingroup
\def\textipa##1{##1}% prevent recursion
```

なお、上流で v0.984 (2011/10/14) でこの問題は修正されているそうです(が、2012/03/20 現在、まだ CTAN には上がっていない)。

3 パラメータの変更

Lua_{TEX}-ja には多くのパラメータが存在する。そして Lua_{TEX} の振る舞いのために、その多くは _{TEX} のレジスタにではなく、Lua_{TEX}-ja 独自の方法で保持されている。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` を用いる必要がある。

3.1 JAchar の範囲の設定

JAchar の範囲を設定するためには、まず各文字に 0 より大きく 217 より小さい index 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漢字面 (SIP) にある全ての文字と ‘漢’ が「100 番の文字範囲」に属するように設定される。

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

この文字範囲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上書きされる。例えば、SIP は全て Lua_{TEX}-ja 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 SIP 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 Lua_{TEX}-ja 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 Lua_{TEX}-ja 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。

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)		

- 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^J ISO 8859-1 の上位領域 (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+00A8, トレマ)
- ° (U+00B0, 度)
- ± (U+00B1, 正又は負符号)
- ´ (U+00B4, アク्यूト・アクセント)
- ¶ (U+00B6, 段落記号)
- × (U+00D7, 乗算記号)
- ÷ (U+00F7, 除算記号)

範囲 1^A 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+0100–U+017F: ラテン文字拡張 A
- 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
- U+0300–U+036F: ダイアクリティカルマーク (合成可能)
- U+1E00–U+1EFF: ラテン文字拡張追加

範囲 2^J 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+0400–U+04FF: キリル文字
- U+1F00–U+1FFF: キリル文字補助

範囲 3^J 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^A 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 :

```
\ljdefcharrange{4}{%
  "500-"10FF, "1200-"1DFF, "2440-"245F, "27C0-"28FF, "2A00-"2AFF,
  "2C00-"2E7F, "4DC0-"4DFF, "A4D0-"A82F, "A840-"ABFF, "FB50-"FE0F,
  "FE20-"FE2F, "FE70-"FEFF, "FB00-"FB4F, "10000-"1FFFFF} % non-Japanese
```

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	(追加漢字面)

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

範囲 5^A 代用符号と補助私用領域 .

範囲 6^J 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^J 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 glue は以下の 3 つのカテゴリに分類される :

- JFM で指定されたグルー / カーン . もし `\inhibitglue` が和文文字の周りで発行されていれば , このグルーは挿入されない .
- デフォルトで 2 つの JAchar の間に挿入されるグルー (kanjiskip) .
- デフォルトで JAchar と ALchar の間に挿入されるグルー (xkanjiskip) .

kanjiskip や xkanjiskip の値は以下のようにして変更可能である .

```
\ltjsetparameter{kanjiskip={0pt plus 0.4pt minus 0.4pt},
xkanjiskip={0.25\zw plus 1pt minus 1pt}}
```

JFM は「望ましい kanjiskip の値」や「望ましい xkanjiskip の値」を持っていることがある . これらのデータを使うためには , kanjiskip や xkanjiskip の値を `\maxdimen` の値に設定すればよい .

3.3 xkanjiskip の設定の挿入

xkanjiskip がすべての JAchar と ALchar の境界に挿入されるのは望ましいことではない . 例えば , xkanjiskip は開き括弧の後には挿入されるべきではない (‘(あ’ と ‘(あ’ を比べてみよ) . LuaTeX-ja では xkanjiskip をある文字の前 / 後に挿入するかどうかを , JAchar に対しては jaxspmode を , ALchar に対しては alxspmode をそれぞれ変えることで制御することができる .

```

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

```

2 目目の引数の `preonly` は「`xkanjiskip` の挿入はこの文字の前でのみ許され、後では許さない」ことを意味する。他に指定可能な値は `postonly`, `allow`, `inhibit` である。

なお、現行の仕様では、`jaxspmode`, `alxspmode` はテーブルを共有しており、上のソースの 1 行目を次のように変えても同じことになる：

```
\ltjsetparameter{alxspmode={`あ,preonly}, jaxspmode={`\!,postonly}}
```

また、これら 2 パラメータには数値で値を指定することもできる（第 II 部を参照）。

もし全ての `kanjiskip` と `xkanjiskip` の挿入を有効化 / 無効化したければ、それぞれ `autospaceing` と `autoxspacing` を `true/false` に設定すればよい。

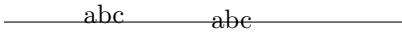
3.4 ベースラインの移動

和文フォントと欧文フォントを合わせるためには、時々どちらかのベースラインの移動が必要になる。pTeX ではこれは `\ybaselineshift` を非零の長さに設定することでなされていた（欧文フォントのベースラインが下がる）。しかし、日本語が主ではない文書に対しては、欧文フォントではなく和文フォントのベースラインを移動した方がよい。このため、LuaTeX-ja では欧文フォントのベースラインのシフト量（`yabaselineshift` パラメータ）と和文フォントのベースラインのシフト量（`yjabaselineshift` パラメータ）を独立に設定できるようになっている。

```

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

```



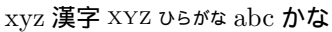
上の例において引かれている水平線がベースラインである。

この機能には面白い使い方があり：2 つのパラメータを適切に設定することで、サイズの異なる文字を中心線に揃えることができるのだ。以下は一つの例である（値はあまり調整されていないことに注意）：

```

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

```



3.5 トンボ

トンボは用紙の四つ角と水平 / 垂直方向の中心を表す印である。pLaTeX と LuaTeX-ja ではトンボの出力をサポートしている。トンボを出力するためには以下の手順が必要である：

1. まず、用紙の左上に印刷されるバナーを定義する。これは `\@bannertoken` にトークンリストを与えることでなされる。

例えば、以下はバナーとして ‘filename (YYYY-MM-DD hh:mm)’ を設定する:

```
\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和文フォント

4.1 \jfont プリミティブ

フォントを和文フォントとして読み込むためには、`\jfont` プリミティブを `\font` プリミティブの代わりに用いる。`\jfont` プリミティブの文法は `\font` と同じである。`LuaTeX-j` は `luaotfload` パッケージを自動的に読み込むので、TrueType/OpenType フォントに `feature` を指定したものを和文フォントとして用いることができる:

```
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

なお、`\jfont` で定義されたコントロールシーケンス (上の例だと `\tradgt`) は `font_def` トークンではないので、`\fontname\tradgt` のような入力はエラーとなることに注意する。以下では `\jfont` で定義されたコントロールシーケンスを `⟨jfont_cs⟩` で表す。

JFM 「はじめに」の節で述べたように、JFM は文字と和文組版で自動的に挿入されるグルーノカーンの寸法情報を持っている。JFM の構造は次の小節で述べる。`\jfont` プリミティブの呼び出しの際には、どの JFM を用いるのかを以下のキーで指定する必要がある:




`jfm=⟨name⟩` JFM の名前を指定する。もし以前に指定された JFM が読み込まれていなければ、`jfm-⟨name⟩.lua` を読み込む。

以下の JFM が `LuaTeX-j` には同梱されている:

`jfm-ujis.lua` `LuaTeX-j` の標準 JFM である。この JFM は `upTeX` で用いられる UTF/OTF パッケージ用のメトリックである `upnmlminr-h.tfm` を元に行っている。`luatexja-otf` パッケージを使うときはこの JFM を指定するべきである。

`jfm-jis.lua` `pTeX` で広く用いられている「JIS フォントメトリック」`jis.tfm` に相当する JFM である。`jfm-ujis.lua` とこの `jfm-jis.lua` の主な違いは、`jfm-ujis.lua` ではほとんどの文字が正形状であるのに対し、`jfm-jis.lua` では横長の長形状である。

表 4. LuaTeX-ja に同梱されている JFM の違い

	jfm-ujis.lua	jfm-jis.lua	jfm-min.lua
例 1* 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。
例 2	ちょっと！何	ちょっと！何	ちょっと！何
Bounding Box			

jfm-min.lua pTeX に同梱されているデフォルトの和文フォントメトリックである min10.tfm に相当する JFM である。この JFM と他の 2 つの JFM の間には表 4 に示すような特筆すべき違いがある。

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

注意：kern feature いくつかのフォントはグリフ間のスペースについての情報を持っている。しかし、この情報は LuaTeX-ja とはあまり相性がよくない。具体的には、この情報に基づいて挿入されるカーニングスペースは JAgLue の挿入過程の前に挿入され、JFM に基づくグルー / カーンも挿入される場合には 2 文字間の意図しないスペースの原因となる。

- `script=...` といった feature を使いたい場合には、`\jfont` プリミティブに `-kern` を指定するべきである。
- もしプロポーショナル幅の和文フォントをそのフォントの情報に基づいて使いたいならば、`jfm-prop.lua` を JFM として指定し、.....

TODO: kanjiskip?

4.2 psft プレフィックス

`file:` と `name:` のプレフィックスに加えて、`\jfont` プリミティブ (と `\font` プリミティブ) では `psft:` プレフィックスを用いることができる。このプレフィックスを用いることで、PDF には埋め込まれない「名前だけの」和文フォントを指定することができる。典型的な使い方は「標準的な」和文フォント、つまり `'Ryumin-Light'` と `'GothicBBB-Medium'` を指定することである。この場合、カーニング他の情報は小塚明朝 Pr6N Regular (Adobe 社によるフォントで、Adobe Reader の日本語フォントパックに含まれている) が用いられる。

`cid key cid key, ...`

4.3 JFM ファイルの構造

JFM ファイルはただ一つの関数呼び出しを含む Lua スクリプトである：

*1 from: 乙部巖己, min10 フォントについて。 <http://argent.shinshu-u.ac.jp/~otobe/tex/files/min10.pdf>.

luatexja.jfont.define_jfm { ... }

実際のデータは上で { ... } で示されたテーブルの中に格納されている．以下ではこのテーブルの構造について記す．なお，JFM ファイル中の長さは全て design-size を単位とする浮動小数点数であることに注意する．

dir=*<direction>* (必須)

JFM の書字方向．現時点では 'yoko' のみがサポートされる．

zw=*<length>* (必須)

「全角幅」の長さ．

zh=*<length>* (必須)

「全角高さ」(height + depth) の長さ．

kanjiskip={*<natural>*, *<stretch>*, *<shrink>*} (任意)

「理想的な」kanjiskip の量を指定する．3.2 節で述べたように，もし kanjiskip が \maxdimen の値ならば，このフィールドで指定された値が実際には用いられる（もしこのフィールドが JFM で指定されていないならば，0pt であるものとして扱われる）．*<stretch>* と *<shrink>* のフィールドも design-size が単位であることに注意せよ．

xkanjiskip={*<natural>*, *<stretch>*, *<shrink>*} (任意)

kanjiskip フィールドと同様に，xkanjiskip の「理想的な」量を指定する．

上記のフィールドに加えて，JFM ファイルはそのインデックスが自然数であるいくつかのサブテーブルを持つ．インデックスが $i \in \omega$ であるテーブルは「文字クラス」 i の情報を格納する．少なくとも，文字クラス 0 は常に存在するので，JFM ファイルはインデックスが [0] のサブテーブルを持たなければならない．それぞれのサブテーブル（そのインデックスを i で表わす）は以下のフィールドを持つ：

chars={*<character>*, ...} (文字クラス 0 を除いて必須)

このフィールドは文字クラス i に属する文字のリストである．このフィールドは $i = 0$ の場合には必須ではない．なぜならば，文字クラス 0 には，0 以外の文字クラスに属するものを除いた全ての JAchar が属するからである（よって，文字クラス 0 はほとんどの JAchar を含む）．このリストでは，文字はその文字コードを用いて，もしくは文字それ自体（長さ 1 の文字列）によって指定される．さらに，このリストで指定される「仮想的な文字」も存在する．これらについては後に記す．

width=*<length>*, height=*<length>*, depth=*<length>*, italic=*<length>* (必須)

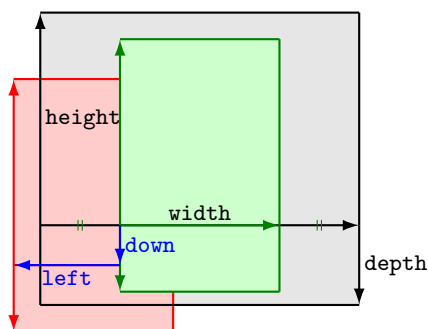
文字クラス i に属する文字の幅，高さ，深さ，イタリック補正の量を指定する．文字クラス i に属する全ての文字は，その幅，高さ，深さがこのフィールドで指定した値であるものとして扱われる．しかし，例外が一つある：もし 'prop' が width フィールドに指定された場合，文字の幅はその「実際の」グリフの幅となる．

left=*<length>*, down=*<length>*, align=*<align>*

これらのフィールドは「実際の」グリフの位置を調整するためにある．align フィールドに指定できる値は 'left', 'middle', 'right' のいずれかである．もしこれら 3 つのフィールドのうちの 1 つが省かれた場合，left と down は 0，align フィールドは 'left' であるものとして扱われる．これら 3 つのフィールドの意味については図 1 で説明する．

多くの場合，left と down は 0 である一方，align フィールドが 'middle' や 'right' であることは珍しいことではない．例えば，align フィールドを 'right' に指定することは，文字クラスが開き括弧類であるときに実際必要である．

kern=[*j*]=*<kern>*, ...}



align フィールドの値が 'middle' である和文文字を含むノードを考えよう。

- 黒色の長方形はノードの枠である。その幅，高さ，深さは JFM によって指定される。
- align フィールドは middle なので，「実際の」グリフは水平方向の中心に配置される（緑色の長方形）。
- さらに，グリフは left と down の値に従ってシフトされる。最終的な実際のグリフの位置は赤色の長方形で示された位置になる。

図 1. 「実際の」グリフの位置。

`glue={ [j] = { <width>, <stretch>, <shrink> }, ... }`

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。以下でその一覧を述べる：

'lineend' 行の終端を表す。

'diffmet' JFM やサイズが異なる 2 つの JAchar の間の境界として用いられる。

'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。

-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

表 5. 和文数式フォントに対するプリミティブ.

	和文フォント	欧文フォント
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>

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

```

「 二分下がり
【 全角下がり
{ 全角二分下がり

4.4 数式フォントファミリ

TeX は数式フォントを 16 のファミリ^{*2}で管理し、それぞれのファミリは 3 つのフォントを持っている：`\textfont`, `\scriptfont` そして `\scriptscriptfont` である.

LuaTeX-ja の数式中での和文フォントの扱いも同様である. 表 5 は数式フォントファミリに対する TeX のプリミティブと対応するものを示している. `\fam` と `\jfam` の値の間には関係はなく、適切な設定の下では `\fam` と `\jfam` の両方に同じ値を設定することができる.

4.5 コールバック

LuaTeX 自体のものに加えて, LuaTeX-ja もコールバックを持っている. これらのコールバックには, 他のコールバックと同様に `luatexbase.add_to_callback` 関数などを用いることでアクセスすることができる.

`luatexja.load_jfm` コールバック このコールバックを用いることで JFM を上書きすることができる. このコールバックは新しい JFM が読み込まれるときに呼び出される.

```

1 function (<table> jfm_info, <string> jfm_name)
2   return <table> new_jfm_info
3 end

```

引数 `jfm_info` は JFM ファイルのテーブルと似たものが格納されるが, クラス 0 を除いた文字のコードを含んだ `chars` フィールドを持つ点が異なる.

このコールバックの使用例は `ltjarticle` クラスにあり, `jfm-min.lua` 中の `'parbdd'` を強制的にクラス 0 に割り当てている. このコールバックは LuaTeX-ja のコードを書き換えない.

`luatexja.define_font` コールバック このコールバックと次のコールバックは組をなしており, Unicode 中に固定された文字コード番号を持たない文字を非零の文字クラスに割り当てることができる. このコールバックは新しい和文フォントが読み込まれたときに呼び出される.

```

1 function (<table> jfont_info, <number> font_number)

```

^{*2} Omega, Aleph, LuaTeX, そして ϵ - (u) pTeX では 256 の数式ファミリを扱うことができるが, これをサポートするために plain TeX と L^ATeX では外部パッケージを読み込む必要がある.

```

2   return <table> new_jfont_info
3 end

```

`jfont_info` は以下のフィールドを持つ：

`jfm` JFM のインデックス番号．

`size` スケールド・ポイント (= 2^{-16} pt) を単位としたフォントのサイズ．

`var` `\jfont` の呼び出しの際に `jfmvar=...` で指定された値．

戻り値の `new_jfont_info` テーブルもこれら 3 つのフィールドを含まなければならない．`font_number` はフォント番号である．

これと次のコールバックの良い使用例は `luatexja-otf` パッケージであり，JFM 中で Adobe-Japan1 CID の文字を "AJ1-xxx" の形で指定するために用いられている．このコールバックは LuaTeX-ja のコードを書き換えない．

`luatexja.find_char_class` コールバック このコールバックは LuaTeX-ja が `chr_code` の文字がどの文字クラスに属するかを決定しようとする際に呼び出される．このコールバックで呼び出される関数は次の形をしていなければならない：

```

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

```

引数 `char_class` は LuaTeX-ja のデフォルトルーチンか，このコールバックの直前の関数呼び出しの結果を含んでおり，したがってこの値は 0 ではないかもしれない．さらに，戻り値の `new_char_class` は `char_class` が非零のときには `char_class` の値と同じであるべきで，そうでないときは LuaTeX-ja のデフォルトルーチンを書き換えることになる．

このコールバックは LuaTeX-ja のコードを書き換えない．

`luatexja.set_width` コールバック このコールバックは LuaTeX-ja が `JAchar` の寸法と位置を調節するためにその `glyph_node` をカプセル化しようとする際に呼び出される．

```

1 function (<table> shift_info, <table> jfont_info, <number> char_class)
2   return <table> new_shift_info
3 end

```

引数 `shift_info` と戻り値の `new_shift_info` は `down` と `left` のフィールドを持ち，これらの値は文字の下 / 左へのシフト量 (スケールド・ポイント単位) である．

良い例が `test/valign.lua` である．このファイルが読み込まれた状態では，JFM 内で規定された文字クラス 0 の文字における (高さ) : (深さ) の比になるように，実際のフォントの出力上下位置が自動調整される．例えば，

- JFM 側の設定 : (高さ) = $88x$, (深さ) = $12x$ (和文 OpenType フォントの標準値)
 - 実フォント側の数値 : (高さ) = $28y$, (深さ) = $5y$ (和文 TrueType フォントの標準値)
- となっていたとする．すると，実際の文字の出力位置は，

$$\frac{88x}{88x + 12x}(28y + 5y) - 28y = \frac{26}{825}y = 0.0315y$$

だけ上にずらされることになる．

5 パラメータ

5.1 `\ltjsetparameter` プリミティブ

先に述べたように、`\ltjsetparameter` と `\ltjgetparameter` は LuaTeX-ja のほとんどのパラメータにアクセスするためのプリミティブである。LuaTeX-ja が pTeX のような文法 (例えば、`\prebreakpenalty``) =10000) を採用しない理由の一つは、LuaTeX のソースにおける `\hpack_filter` コールバックの位置にある。9 節を見よ。

`\ltjsetparameter` と `\ltjglobalsetparameter` はパラメータを指定するためのプリミティブである。これらは $\langle key \rangle = \langle value \rangle$ のリストを引数としてとる。許されるキーは次の節に記述する。`\ltjsetparameter` と `\ltjglobalsetparameter` の違いはスコープの違いのみである。`\ltjsetparameter` はローカルな指定、`\ltjglobalsetparameter` はグローバルな指定を行う。これらは他のパラメータ指定と同様に `\globaldefs` の値に従う。

`\ltjgetparameter` はパラメータの値を取得するためのプリミティブであり、常にパラメータの名前を第一引数にとる。そして、いくつかの場合には加えてさらに引数 (例えば文字コード) をとる。

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

`\ltjgetparameter` の戻り値は常に文字列である。これは `tex.write()` によって出力しているためで、スペース ‘ ’ (U+0020) を除いた文字のカテゴリコードは全て 12 (other) となる。一方、スペースのカテゴリコードは 10 (space) である。

5.2 パラメータ一覧

以下は `\ltjsetparameter` に指定することができるパラメータの一覧である。`[\cs]` は pTeX における対応物を示す。また、それぞれのパラメータの右上にある記号には次の意味がある：

- 記号なし：段落や `hbox` の終端での値がその段落 / `hbox` 全体で用いられる。
- ‘*’：ローカルなパラメータであり、段落 / `hbox` 内のどこでも値を変えることができる。
- ‘†’：指定は常にグローバルになる。

`jcharwidowpenalty = \langle penalty \rangle` [`\jcharwidowpenalty`]

パラグラフの最後の字が孤立して改行されるのを防ぐためのペナルティの値。このペナルティは (日本語の) 句読点として扱われない最後の `JChar` の直後に挿入される。

`kcatcode = \{ \langle chr_code \rangle, \langle natural number \rangle \}`

文字コードが `\langle chr_code \rangle` の文字を持つ付加的な属性値 (attribute)。現在のバージョンでは、`\langle natural number \rangle` の最下位ビットが、その文字が句読点とみなされるかどうかを表している (上の `jcharwidowpenalty` の記述を見よ)。

`prebreakpenalty = \{ \langle chr_code \rangle, \langle penalty \rangle \}` [`\prebreakpenalty`] 文字コード `\langle chr_code \rangle` の `JChar` 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={⟨chr_code⟩,⟨penalty⟩}` [`\postbreakpenalty`] 文字コード `⟨chr_code⟩` の JAchar 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={⟨jfam⟩,⟨jfont_cs⟩} [TeX の \textfont]
```

```
jascriptfont={⟨jfam⟩,⟨jfont_cs⟩} [TeX の \scriptfont]
```

```
jascriptscriptfont={⟨jfam⟩,⟨jfont_cs⟩} [TeX の \scriptscriptfont]
```

```
yjabaselineshift=⟨dimen⟩*
```

```
yabaselineshift=⟨dimen⟩* [\ybaselineshift]
```

`jaxspmode={⟨chr_code⟩,⟨mode⟩}` 文字コードが `⟨chr_code⟩` の JAchar の前/後ろに xkanjiskip の挿入を許すかどうかの設定。以下の `⟨mode⟩` が許される：

- 0, inhibit xkanjiskip の挿入は文字の前/後ろのいずれでも禁止される。
- 1, preonly xkanjiskip の挿入は文字の前では許されるが、後ろでは許されない。
- 2, postonly xkanjiskip の挿入は文字の後ろでは許されるが、前では許されない。
- 3, allow xkanjiskip の挿入は文字の前/後ろのいずれでも許される。これがデフォルトの値である。このパラメータは pTeX の `\inhibitxspcode` プリミティブと似ているが、互換性はない。

```
alxspmode={⟨chr_code⟩,⟨mode⟩} [\xspcode]
```

文字コードが `⟨chr_code⟩` の ALchar の前/後ろに xkanjiskip の挿入を許すかどうかの設定。以下の `⟨mode⟩` が許される：

- 0, inhibit xkanjiskip の挿入は文字の前/後ろのいずれでも禁止される。
- 1, preonly xkanjiskip の挿入は文字の前では許されるが、後ろでは許されない。
- 2, postonly xkanjiskip の挿入は文字の後ろでは許されるが、前では許されない。
- 3, allow xkanjiskip の挿入は文字の前/後ろのいずれでも許される。これがデフォルトの値である。

`jaxspmode` と `alxspmode` は共通のテーブルを用いているため、これら 2 つのパラメータは互いの異名となっていることに注意する。

```
autospacing=⟨bool⟩* [\autospacing]
```

```
autoxspacing=⟨bool⟩* [\autoxspacing]
```

```
kanjiskip=⟨skip⟩ [\kanjiskip]
```

```
xkanjiskip=⟨skip⟩ [\xkanjiskip]
```

`differentjfm=⟨mode⟩†` JFM（もしくはサイズ）が異なる 2 つの JAchar の間にグルー/カーンをどのように入れるかを指定する。許される値は以下の通り：

```
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

以下のプリミティブは p_TEX との互換性のために実装されている：

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

6.2 \inhibitglue プリミティブ

\inhibitglue プリミティブは JAg_{lue} の挿入を抑制する。以下は、ボックスの始めと‘あ’の間、‘あ’と‘う’の間にグルーが入る特別な JFM を用いた例である。

```
1 \jfont\g=psft:Ryumin-Light:jfm=test \g
2 \fbox{\hbox{あう\inhibitglue う}}
3 \inhibitglue\par\noindent あ1
4 \par\inhibitglue\noindent あ2
5 \par\noindent\inhibitglue あ3
6 \par\inhibitglue\hrule あ off\inhibitglue
   ice
```

あ	うあう
あ	1
あ	2
あ	3
あ	office

この例を援用して、\inhibitglue の仕様について述べる。

- \inhibitglue の垂直モード中での呼び出しは次の段落の始めで効力を持つ。これは \everypar のハックによって実現されている。
- \inhibitglue の（制限された）水平モード中での呼び出しはその場でのみ有効であり、段落の境界を乗り越えない。さらに、\inhibitglue は上の例の 4 行目のようにリガチャとカーニングを打ち消す。
- \inhibitglue を数式モード中で呼び出した場合はただ無視される。

7 L^AT_EX_{2_ε} 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_TEX-ja simply adopted pl_fonts.dtx in p_LA_TEX_{2_ε} 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 Lua_TEX-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変更の命令 `<al-cmd>` (`\mathrm` 等) と，和文数式用フォントファミリ変更の命令 `<ja-cmd>` (`\mathmc` 等) 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>` で和文側も変更させるようにするのが一般的と思われる．

本命令は

```
<unified-cmd>{<arg>} → ((<al-cmd> を 1 段展開したもの){<ja-cmd> を 1 段展開したもの){<arg>}}
```

と定義を行うので，使用には注意が必要である：

- `<al-cmd>`, `<ja-cmd>` は既に定義されていなければならない．`\reDeclareMathAlphabet` 後に両命令の内容を再定義しても，`<unified-cmd>` の内容にそれは反映されない．
- `<al-cmd>`, `<ja-cmd>` に `\@mathrm` などと `@` をつけた命令を指定した時の動作は保証できない．

```
\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 IAT_{EX} 2_ε, 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 $\langle fam \rangle$ under the encoding $\langle ja-enc \rangle$ is already defined by `\DeclareKanijFamily`.
- A font definition named $\langle enc \rangle \langle ja-enc \rangle .fd$ (the file name is all lowercase) exists.
- Let current encoding scheme for Japanese fonts be $\langle al-enc \rangle$. 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 $\langle family \rangle$ doesn't seem to be defined neither under the encoding $\langle ja-enc \rangle$, nor under $\langle al-enc \rangle$.
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^AT_EX.

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 LuaTeX-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. This `\jH` is a synonym of `\jQ`.

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

`\tj@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.

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

`\tj@charclass` (attribute) The character class of Japanese *glyph_node*.

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

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

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

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

`\tj@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 (\backslash). This distinction of origins of glues (from explicit $\backslash\text{kern}$, or from \backslash) 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.

`\tj@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, LuaTeX-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 LuaTeX-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 **JAg**lues.

9.2 LuaTeX-ja のスタックシステム

Background LuaTeX-ja has its own stack system, and most parameters of LuaTeX-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      ふがふが. ほ げほ げ. びよびよ
   =5pt}ほげほげ}
3 \box0. びよびよ\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 5 pt. However, by the implementation method of LuaTeX, this '5 pt' 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 '5 pt' 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^{*3}.

^{*3} [Dev-luatex] `tex.currentgrouplevel`, a post at 2008/8/19 by Jonathan Sauer.

These are two T_EX count registers for maintaining information: `\ltj@stack` for the stack level, and `\ltj@group@level` for the T_EX'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, LuaT_EX-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 T_EX'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 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: pT_EX の動作

欧文では文章の改行は単語間でしか行わない。そのため、T_EX 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、pT_EX 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、pT_EX が T_EX 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。pT_EX の入力処理部は、T_EX 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。

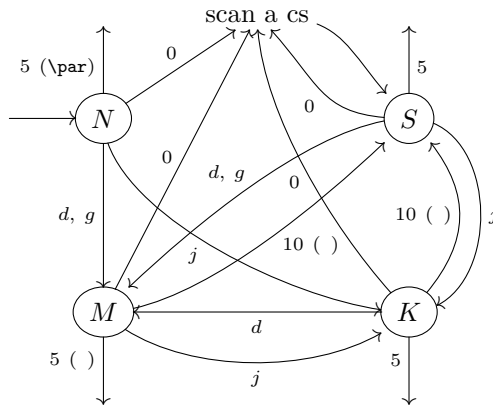
- State N : 行の開始。
- State S : 空白読み飛ばし。
- State M : 行中。
- State K : 行中 (和文文字の後)。

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 T_EX 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、T_EX 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。



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

- 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.

10.2 LuaTeX-ja の動作

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^{*4}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：

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

この仕様は、前節で述べた pTeX の仕様に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：

^{*4} 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。

```

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） $\langle char_node \rangle$ を追加する過程で行われる。
- 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。

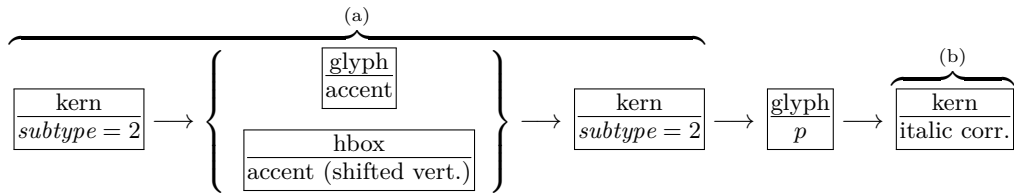
しかし、LuaTeX-ja では、hbox へのパッケージングや行分割前に全ての JAgglue、即ち JFM グルー・xkanjiskip・kanjiskip の 3 種類を一度に挿入することになっている。これは、LuaTeX において欧文の合字・カーニング処理がノードベースになったことに対応する変更である。

LuaTeX-ja における JAgglue 挿入処理では、次節で定義する「クラスタ」を単位にして行われる。大雑把にいうと、「クラスタ」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 つのクラスタの間には、ペナルティ、\vadjust、whatsit など、行組版には関係しないものがある。

11.2 「クラスタ」の定義

定義 1. A *cluster* is a list of consecutive 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. The *id* is *id_math*.
3. A *glyph_node p* with nodes which relate with it:
 - (1) A kern for the italic correction of *p*.
 - (2) An accent attached to *p* by $\backslash\text{accent}$.



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, a vbox, a rule (`\vrule`) and an *unset_node*. 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*.

以降は日本語 .

id の意味 *Np.id* の意味を述べるとともに、「先頭の文字」を表す *glyph_node* *Np.head* と、「最後の文字」を表す *glyph_node* *Np.tail* を次のように定義する . 直感的に言うと , *Np* は *Np.head* で始まり *Np.tail* で終わるような単語 , と見做すことができる . これら *Np.head*, *Np.tail* は説明用に準備した概念であって , 実際の Lua コード中にそのように書かれているわけではないことに注意 .

id_jglyph 和文文字 .

Np.head, *Np.tail* は , その和文文字を表している *glyph_node* そのものである .

id_glyph 和文文字を表していない *glyph_node* *p* .

多くの場合 , *p* は欧文文字を格納しているが , ‘ffi’ などの合字によって作られた *glyph_node* である可能性もある . 前者の場合 , *Np.head*, *Np.tail* = *p* である . 一方 , 後者の場合 ,

- *Np.head* は , 合字の構成要素の先頭 → (その *glyph_node* における) 合字の構成要素の先頭 → と再帰的に検索していったどり着いた *glyph_node* である .
- *Np.last* は , 同様に末尾 → 末尾 → と検索してたどり着いた *glyph_node* である .

id_math インライン数式 .

便宜的に , *Np.head*, *Np.tail* ともに「文字コード -1 の欧文文字」とおく .

id_hlist 縦方向にシフトされていない hbox .

この場合 , *Np.head*, *Np.tail* はそれぞれ *p* の内容を表すリストの , 先頭・末尾のノードである .

- 状況によっては , TeX ソースで言うと

$$\backslash\hbox{\backslash\hbox{abc}\dots\backslash\hbox{\lower1pt\hbox{xyz}}}$$
のように , *p* の内容が別の hbox で開始・終了している可能性も十分あり得る . そのような場合 , *Np.head*, *Np.tail* の算出は , 垂直方向にシフトされていない hbox の場合だけ内部を再帰的に探索する . 例えば上の例では , *Np.head* は文字 ‘a’ を表すノードであり , 一方 *Np.tail* は垂直方向にシフトされた hbox , $\backslash\lower1pt\hbox{xyz}$ に対応するノードである .
- また , 先頭にアクセント付きの文字がきたり , 末尾にイタリック補正用の kern が来ることもあり得る . この場合は , クラスタの定義のところにもあったように , それらは無視して算出を行う .
- 最初・最後のノードが合字によって作られた *glyph_node* のときは , それぞれに対して *id_glyph* と同様に再帰的に構成要素をたどっていく .

id_pbox 「既に処理された」ノードのリストであり , これらのノードが二度処理を受けないためにまとめて 1 つのクラスタとして取り扱うだけである . *id_hlist* と同じ方法で *Np.head*, *Np.tail* を算出する ,

id_disc discretionary break (`\discretionary{pre}{post}{nobreak}`).

id_hlist と同じ方法で *Np.head*, *Np.tail* を算出するが, 第 3 引数の *nobreak* (行分割が行われない時の内容) を使う. 言い換えれば, ここで行分割が発生した時の状況は全く考慮に入れない.

id_box_like *id_hlist* とならない box や, rule .

この場合は, *Np.head*, *Np.tail* のデータは利用されないので, 2 つの算出は無意味である. 敢えて明示するならば, *Np.head*, *Np.tail* は共に nil 値である.

他 以上がない *id* に対しても, *Np.head*, *Np.tail* の算出は無意味.

クラスタの別の分類 さらに, JFM グルー挿入処理の実際の説明により便利なように, *id* とは別のクラスタの分類を行っておく. 挿入処理では 2 つの隣り合ったクラスタの間に空白等の実際の挿入を行うことは前に書いたが, ここでの説明では, 問題にしているクラスタ *Np* は「後ろ側」のクラスタであるとする. 「前側」のクラスタについては, 以下の説明で *head* が *last* に置き換わることに注意すること.

和文 A リスト中に直接出現している和文文字. *id* が *id_jglyph* であるか,

id が *id_pbox* であって *Np.head* が JAchar であるとき.

和文 B リスト中の hbox の中身の先頭として出現した和文文字. 和文 A との違いは, これの前に JFM グルーの挿入が行われない (`xkanjiskip`, `kanjiskip` は入り得る) ことである.

id が *id_hlist* か *id_disc* であって *Np.head* が JAchar であるとき.

欧文 リスト中に直接/hbox の中身として出現している欧文文字. 次の 3 つの場合が該当:

- *id* が *id_glyph* である.
- *id* が *id_math* である.
- *id* が *id_pbox* か *id_hlist* か *id_disc* であって, *Np.head* が ALchar .

箱 box, またはそれに類似するもの. 次の 2 つが該当:

- *id* が *id_pbox* か *id_hlist* か *id_disc* であって, *Np.head* が *glyph_node* でない.
- *id* が *id_box_like* である.

11.3 段落/hbox の先頭や末尾

先頭部の処理 まず, 段落/hbox の一番最初にあるクラスタ *Np* を探索する. hbox の場合は何の問題もないが, 段落の場合では以下のノード達を事前に読み飛ばしておく:

`\parindent` 由来の hbox (*subtype* = 3), 及び *subtype* が 44 (*user_defined*) でないような whatsit .

これは, `\parindent` 由来の hbox がクラスタを構成しないようにするためである.

次に, *Np* の直前に空白 *g* を必要なら挿入する:

1. この処理が働くような *Np* は和文 A である.
2. 問題のリストが字下げありの段落 (`\parindent` 由来の hbox あり) の場合は, この空白 *g* は「文字コード 'parbdd' の文字」と *Np* の間に入る glue/kern である.
3. そうでないとき (`noindent` で開始された段落や, hbox) は, *g* は「文字コード 'boxbdd' の文字」と *Np* の間に入る glue/kern である.

ただし, もし *g* が glue であった場合, この挿入によって *Np* による行分割が新たに可能になるべきではない. そこで, 以下の場合には, *g* の直前に `\penalty10000` を挿入する:

- 問題にしているリストが段落であり，かつ
- N_p の前には予めペナルティがなく， g は glue .

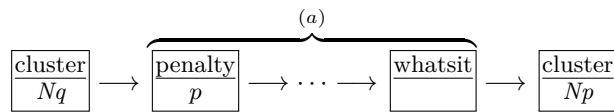
末尾の処理 末尾の処理は，問題のリストが段落のものか hbox のものかによって異なる．後者の場合は容易い：最後のクラスタを N_q とおくと， N_q と「文字コード 'boxbddd' の文字」の間に入る glue/kern を， N_q の直後に挿入するのみである．

一方．前者（段落）の場合は，リストの末尾は常に `\penalty10000` と，`\parfillskip` 由来のグルーが存在する．よって，最後のクラスタ N_p はこの `\parfillskip` 由来のグルーとなり，実質的な中身の最後はその 1 つ前のクラスタ N_q となる．

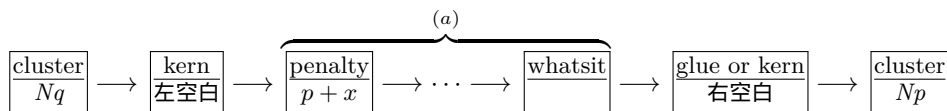
1. まず N_q の直後に（後に述べる）line-end [E] によって定まる空白を挿入する．
2. 次に，段落の最後の「通常の和文文字 + 句点」が独立した行となるのを防ぐために，`jcharwidowpenalty` の値の分だけ適切な場所のペナルティを増やす．
ペナルティ量を増やす場所は，`head` が JAchar であり，かつその文字の `kcatcode` が偶数であるような最後のクラスタの直前にあるものたちである*5．

11.4 概観と典型例：2 つの「和文 A」の場合

先に述べたように，2 つの隣り合ったクラスタ， N_q と N_p 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．模式的に表すと，



のようになっている．間の (a) に相当する部分には，何のノードもない場合ももちろんあり得る．そうして，JFM グルー挿入後には，この 2 クラスタ間は次のようになる：



以後，典型的な例として，クラスタ N_q と N_p が共に和文 A である場合を見ていこう，この場合が全ての場合の基本となる．

「右空白」の算出 まず，「右空白」にあたる量を算出する．通常はこれが，隣り合った 2 つの和文文字間に入る空白量となる．

JFM 由来 [M] JFM の文字クラス指定によって入る空白を以下によって求める．この段階で空白量が未定義（未指定）だった場合，デフォルト値 `kanjiskip` を採用することとなるので，次へ．

1. もし両クラスタの間で `\inhibitglue` が実行されていた場合（証として `whatsit` ノードが自動挿入される），代わりに `kanjiskip` が挿入されることとなる．次へ．
2. N_q と N_p が同じ JFM・同じ `jfmvar` キー・同じサイズの和文フォントであったならば，共通に使っている JFM 内で挿入される空白 (glue or kern) が決まっているか調べる．

*5 大雑把に言えば，`kcatcode` が奇数であるような JAchar を約物として考えていることになる．`kcatcode` の最下位ビットはこの `jcharwidowpenalty` 用にも利用される．

3. 1. でも 2. でもない場合は、 Nq と Np が違う JFM/jfmvar/サイズである。この場合、まず

$$gb := (Nq \text{ と「文字コードが'diffmet' の文字」との間に入る glue/kern})$$

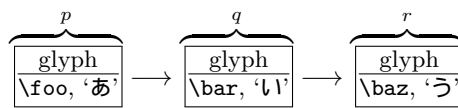
$$ga := (\text{「文字コードが'diffmet' の文字」と } Np \text{ との間に入る glue/kern})$$

として、左側由来・右側由来の空白 (glue/kern) を (それぞれの JFM から) 求める。 ga と gb のどちらか片方が未定義であるならば、定義されている側の値をそのまま採用する。もし ga と gb が両方決まっているならば、両者の値を平均*6した値を採用する。

例えば、

```
\font\foo=psft:Ryumin-Light:jfm=ujis
\font\bar=psft:GothicBBB-Medium:jfm=ujis
\font\baz=psft:GothicBBB-Medium:jfm=ujis;jfmvar=piyo
```

という 3 フォントを考え、



という 3 ノードを考える (それぞれ単独でクラスタをなす)。この場合、 p と q の間は、実フォントが異なるにもかかわらず (2) の状況となる一方で、 q と r の間は (実フォントが同じなのに) jfmvar キーの内容が異なるので (3) の状況となる。

kanjiskip [K] 上の [M] において空白が定まらなかった場合、kanjiskip の値を以下で定め、それを「右空白」として採用する。この段階においては、\inhibitglue は効力を持たないため、結果として、2 つの和文文字間には常に何らかの glue/kern が挿入されることとなる。

1. 両クラスタ (厳密には $Nq.tail$, $Np.head$) の中身の文字コードに対する autospacing パラメタが両方とも false だった場合は、長さ 0 の glue とする。
2. ユーザ側から見た kanjiskip パラメタの自然長が $\backslashmaxdimen = (2^{30} - 1) \text{ sp}$ でなければ、kanjiskip パラメタの値を持つ glue を採用する。
3. 2. でない場合は、 Nq , Np で使われている JFM に指定されている kanjiskip の値を用いる。どちらか片方のクラスタだけが和文文字 (和文 A・和文 B) のときは、そちらのクラスタで使われている JFM 由来の値だけを用いる。もし両方で使われている JFM が異なった場合は、上の [M] 3. と同様の方法を用いて調整する。

「左空白」の算出とそれに伴う補正 次に、「左空白」にあたる量を算出する：

line-end [E] Nq と Np の間で行分割が起きたときに、 Nq と行末の間に入る空白である。ぶら下げ組の組版などに用いられることを期待している。

1. 既に算出した「右空白」が kern である場合は、「左空白」は挿入されない。
2. 「右空白」が glue か未定義 (長さ 0 の glue とみなす) の場合は、「左空白」は Nq と「文字コード 'lineend' の文字」との間に入る kern として、JFM から決定される。
3. 2. で決まった「左空白」の長さが 0 でなければ、その分だけ先ほど算出した「右空白」の自然長を引く。

禁則用ペナルティの挿入 まず、

$$a := (Nq^{*7} \text{ の文字に対する postbreakpenalty の値}) + (Np^{*8} \text{ の文字に対する prebreakpenalty の値})$$

*6 differentjfm パラメタの値によって、「大きい方」「小さい方」「合計」に変えることができる。

表 6. Summary of JFM glues

$Np \downarrow$	和文 A	和文 B	欧文	箱	glue	kern
和文 A	$\frac{E \quad M \rightarrow K}{PN}$	$\frac{O_A \rightarrow K}{PN}$	$\frac{O_A \rightarrow X}{PN}$	$\frac{O_A}{PA}$	$\frac{O_A}{PN}$	$\frac{O_A}{PS}$
和文 B	$\frac{E \quad O_B \rightarrow K}{PA}$	$\frac{K}{PS}$	$\frac{X}{PS}$			
欧文	$\frac{E \quad O_B \rightarrow X}{PA}$	$\frac{X}{PS}$				
箱	$\frac{E \quad O_B}{PA}$					
glue	$\frac{E \quad O_B}{PN}$					
kern	$\frac{E \quad O_B}{PS}$					

Here $\frac{E \quad M \rightarrow K}{PN}$ means that

1. To determine the ‘right-space’, LuaTeX-ja first attempts by the method ‘JFM-origin [M]’. If this attempt fails, LuaTeX-ja use the method ‘kanjiskip [K]’.
2. The ‘left space’ between Nq and Np is determined by the method ‘line-end [E]’.
3. LuaTeX-ja adopts the method ‘P-normal [PN]’ to adjust the penalty between two clusters for *kinsoku shori*.

とおく、ペナルティは通常 $[-10000, 10000]$ の整数値をとり、また ± 10000 は正負の無限大を意味することになっているが、この a の算出では単純な整数の加減算を行う。

a は禁則処理用に Nq と Np の間に加えられるべきペナルティ量である。

P-normal [PN] Nq と Np の間の (a) 部分にペナルティ (*penalty_node*) があれば処理は簡単である：それらの各ノードにおいて、ペナルティ値を (± 10000 を無限大として扱いつつ) a だけ増加させればよい。また、 $10000 + (-10000) = 0$ としている。

少々困るのは、(a) 部分にペナルティが存在していない場合である。直感的に、補正すべき量 a が 0 でないとき、その値をもつ *penalty_node* を作って「右空白」の (もし未定義なら Np の) 直前に挿入…… ということになるが、実際には僅かにこれより複雑である。

- 「右空白」が kern であるとき、それは「 Nq と Np の間で改行は許されない」ことを意図している。そのため、この場合は $a \neq 0$ であってもペナルティの挿入はしない。
- 「左空白」が kern としてきちんと定義されている時 (このとき、「右空白」は kern でない)、この「左空白」の直後での行分割を許容しないとイケないので、 $a = 0$ であっても *penalty_node* を作って挿入する。
- 以上のどれでもないときは、 $a \neq 0$ ならば *penalty_node* を作って挿入する。

11.5 その他の場合

本節の内容は表 6 にまとめてある。

*8 厳密にはそれぞれ $Nq.tail$, $Np.head$ 。

和文 A と欧文の間 N_q が和文 A で、 N_p が欧文の場合、JFM グルー挿入処理は次のようにして行われる。

- 「右空白」については、まず以下に述べる Boundary-B [O_B] により空白を決定しようと試みる。それが失敗した場合は、`xkanjiskip [X]` によって定める。
- 「左空白」については、既に述べた line-end [E] をそのまま採用する。それに伴う「右空白」の補正も同じ。
- 禁則用ペナルティも、以前述べた P-normal [PN] と同じである。

Boundary-B [O_B] 和文文字と「和文でないもの」との間に入る空白を以下によって求め、未定義でなければそれを「右空白」として採用する。JFM-origin [M] の変種と考えて良い。これによって定まる空白の典型例は、和文の閉じ括弧と欧文文字の間に入る半角アキである。

1. もし両クラスタの間で `\inhibitglue` が実行されていた場合（証として `whatsit` ノードが自動挿入される）、次へ。
2. そうでなければ、 N_q と「文字コードが `'jcharbdd'` の文字」との間に入る `glue/kern` として定まる。
`xkanjiskip [X]` この段階では、`kanjiskip [K]` のときと同じように、`xkanjiskip` の値を以下で定め、それを「右空白」として採用する。この段階で `\inhibitglue` は効力を持たないのも同じである。
1. 以下のいずれかの場合は、`xkanjiskip` の挿入は抑止される。しかし、実際には行分割を許容するために、長さ 0 の `glue` を採用する：
 - 両クラスタにおいて、それらの中身の文字コードに対する `autoxspacing` パラメタが共に `false` である。
 - N_q の中身の文字コードについて、「直後への `xkanjiskip` の挿入」が禁止されている（つまり、`jaxspmode` (or `alxspmode`) パラメタが 2 以上）。
 - N_p の中身の文字コードについて、「直前への `xkanjiskip` の挿入」が禁止されている（つまり、`jaxspmode` (or `alxspmode`) パラメタが偶数）。
2. ユーザ側から見た `xkanjiskip` パラメタの自然長が `\maxdimen = (2^{30} - 1)sp` でなければ、`xkanjiskip` パラメタの値を持つ `glue` を採用する。
3. 2. でない場合は、 N_q , N_p (和文 A/和文 B なのは片方だけ) で使われている JFM に指定されている `xkanjiskip` の値を用いる。

欧文と和文 A の間 N_q が欧文で、 N_p が和文 A の場合、JFM グルー挿入処理は上の場合とほぼ同じである。和文 A のクラスタが逆になるので、Boundary-A [O_A] の部分が変わるだけ。

- 「右空白」については、まず以下に述べる Boundary-A [O_A] により空白を決定しようと試みる。それが失敗した場合は、`xkanjiskip [X]` によって定める。
- N_q が和文でないので、「左空白」は算出されない。
- 禁則用ペナルティは、以前述べた P-normal [PN] と同じである。

Boundary-A [O_A] 「和文でないもの」と和文文字との間に入る空白を以下によって求め、未定義でなければそれを「右空白」として採用する。JFM-origin [M] の変種と考えて良い。これによって定まる空白の典型例は、欧文文字と和文の開き括弧との間に入る半角アキである。

1. もし両クラスタの間で `\inhibitglue` が実行されていた場合（証として `whatsit` ノードが自動挿入される）、次へ。
2. そうでなければ、「文字コードが `'jcharbdd'` の文字」と N_p との間に入る `glue/kern` として定まる。

和文 A と箱 $\cdot \text{glue} \cdot \text{kern}$ の間 N_q が和文 A で、 N_p が箱 $\cdot \text{glue} \cdot \text{kern}$ のいずれかであった場合、両者の間に挿入される JFM グルーについては同じ処理である。しかし、ここでの行分割に対する仕様が異なるので、ペナルティの挿入処理は若干異なったものとなっている。

- 「右空白」については、既に述べた Boundary-B [O_B] により空白を決定しようと試みる。それが失敗した場合は、「右空白」は挿入されない。
- 「左空白」については、既に述べた line-end [E] の算出方法をそのまま採用する。それに伴う「右空白」の補正も同じ。
- 禁則用ペナルティの処理は、後ろのクラス N_p の種類によって異なる。なお、 $N_p.head$ は無意味であるから、「 $N_p.head$ に対する prebreakpenalty の値」は 0 とみなされる。言い換えれば、

$$a := (N_q^{*9} \text{の文字に対する } \text{postbreakpenalty} \text{ の値}).$$

箱 N_p が箱であった場合は、両クラスの間での行分割は（明示的に両クラスの間には $\backslash\text{penalty}10000$ があつた場合を除き）いつも許容される。そのため、ペナルティ処理は、後に述べる P-allow [PA] が P-normal [PN] の代わりに用いられる。

glue N_p が glue の場合、ペナルティ処理は P-normal [PN] を用いる。

kern N_p が kern であった場合は、両クラスの間での行分割は（明示的に両クラスの間にはペナルティがあつた場合を除き）許容されない。ペナルティ処理は、後に述べる P-suppress [PS] を使う。

これらの P-normal [PN]、P-allow [PA]、P-suppress [PS] の違いは、 N_q と N_p の間（以前の図だと (a) の部分）にペナルティが存在しない場合にのみ存在する。

P-allow [PA] N_q と N_p の間の (a) 部分にペナルティがあれば、P-normal [PN] と同様に、それらの各ノードにおいてペナルティ値を a だけ増加させる。

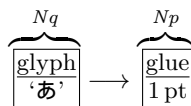
(a) 部分にペナルティが存在していない場合、 $\text{LuaTeX-j}\ddot{a}$ は N_q と N_p の間での行分割を可能にしようとする。そのために、以下の場合に a をもつ penalty_node を作って「右空白」の（もし未定義なら N_p の）直前に挿入する：

- 「右空白」が glue でない（kern か未定義）であるとき。
- 「左空白」が kern としてきちり定義されている時。

P-suppress [PS] N_q と N_p の間の (a) 部分にペナルティがあれば、P-normal [PN] と同様に、それらの各ノードにおいてペナルティ値を a だけ増加させる。

(a) 部分にペナルティが存在していない場合、 N_q と N_p の間での行分割は元々不可能のはずだったのであるが、 $\text{LuaTeX-j}\ddot{a}$ はそれをわざわざ行分割可能にはしない。そのため、「右空白」が glue であれば、その直前に $\backslash\text{penalty}10000$ を挿入する。

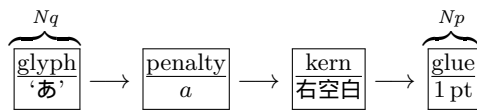
なお、「右空白」は kern、「左空白」は未定義の



のような状況を考える。このとき、 a 、即ち「あ」の postbreakpenalty がいかなる値であっても、この 2 クラス間には最終的に



となり、 a 分のペナルティは挿入されないことに注意して欲しい。postbreakpenalty は (a は) 殆どの場合が非負の値と考えられ、そのような場合では (1) と



との間に差異は生じない*10。

箱・glue・kern と和文 A の間 Np が箱・glue・kern のいずれかで、 Np が和文 A であった場合は、すぐ上の (Nq と Np の順序が逆になっている) 場合とほぼ同じであるが、「左空白」がなくなることにのみ注意。

- 「右空白」については、既に述べた Boundary-A [O_A] により空白を決定しようとする。それが失敗した場合は、「右空白」は挿入されない。
- Nq が和文でないので、「左空白」は算出されない。
- 禁則用ペナルティの処理は、 Nq の種類によって異なる。 $Nq.tail$ は無意味なので、

$$a := (Np^{*11} \text{の文字に対する prebreakpenalty の値}).$$

- 箱 Nq が箱の場合は、P-allow [PA] を用いる。
- glue Nq が glue の場合は、P-normal [PN] を用いる。
- kern Nq が kern の場合は、P-suppress [PS] を用いる。

和文 A と和文 B の違い 先に述べたように、和文 B は hbox の中身の先頭 (or 末尾) として出現している和文文字である。リスト内に直接ノードとして現れている和文文字 (和文 A) との違いは、

- 和文 B に対しては、JFM の文字クラス指定から定まる空白 JFM-origin [M], Boundary-A [O_A], Boundary-B [O_B] の挿入は行われない。「左空白」の算出も行われない。例えば、
 - 片方が和文 A、もう片方が和文 B のクラスタの場合、Boundary-A [O_A] または Boundary-B [O_B] の挿入を試み、それがダメなら kanjiskip [K] の挿入を行う。
 - 和文 B の 2 つのクラスタの間には、kanjiskip [K] が自動的に入る。
- 和文 B と箱・glue・kern が隣接したとき (どちらが前かは関係ない)、間に JFM グルー・ペナルティの挿入は一切しない。
- 和文 B と和文 B、また和文 B と欧文とが隣接した時は、禁則用ペナルティ挿入処理は P-suppress [PS] が用いられる。
- 和文 B の文字に対する prebreakpenalty, postbreakpenalty の値は使われず、0 として計算される。

次が具体例である：

1 あ .\inhibitglue A\\	あ .A
2 \hbox{あ .}A\\	あ .A
3 あ .A	あ .A

- 1 行目の \inhibitglue は Boundary-B [O_B] の処理のみを抑制するので、ピリオドと「A」の間には xkanjiskip (四分アキ) が入ることに注意。
- 2 行目のピリオドと「A」の間においては、前者が和文 B となる (hbox の中身の末尾として登場しているから) ので、そもそも Boundary-B [O_B] の処理は行われない。よって、xkanjiskip が入ることとなる。

*10 kern→glue が 1 つの行分割可能点 (行分割に伴うペナルティは 0) であるため、たとえ $a = 10000$ であっても、 Nq と Np の間で行分割を禁止することはできない。

- 3 行目では , ピリオドの属するクラスは和文 A である . これによって , ピリオドと「A」の間には Boundary-B [O_B] 由来の半角アキが入ることになる .

12 psft

参考文献

- [1] Victor Eijkhout, *TEX by Topic, A T_EXnician's Reference*, Addison-Wesley, 1992.