

Typesetting and Layout in Multiple Directions

John Plaice

The University of New South Wales
Sydney, Australia

Visiting Scientist, Concordia University,
Montreal, Canada

October 26, 2013

Example problem: Japanese

10月27日、マイクロソフトの定番ビジネススイート「マイクロソフト・オフィス」の最新版となる「オフィス2011 (Office for Mac 2011)」が、ついに発売されました。2009年8月、マイクロソフトは新しいMac版オフィスについてのロードマップを示しました。当時、2010年のホリデーシーズンに発売する予定であることや新たに統合メールソフトの「アウトLOOK」を搭載することなどが発表されましたが、詳細はあまり明らかにされず、Macユーザは長い間その詳細の発表を心待ちにしていたわけです。

Lines are read down, from left to right. Japanese is vertical. English is rotated -90° . Numbers are vertical or horizontal.

Each hbox or vbox can be in *yoko* (normal) or *tate* direction. Hisato Hamano, *Vertical Typesetting with T_EX, TUGboat* 11(3):346–352, 1990.

組版することが可能であることを示している。もちろん、pL^AT_EXなどで実現するには、方式によって難易がある(方式Bは難しい)。具体的な方法は後述するとして、`\hbox` pL^AT_EXの柔軟性を強調してお。

(2) 本書の方式(方式D)

本書では、`\hbox` の方式(D)をそのまま使用している。すなわち、カギ括弧は、段落始めでは全角二分下がり、行頭では天ツキで出力される。すこし、長い例を次に示す。入力コードは次の通りである。

入 力

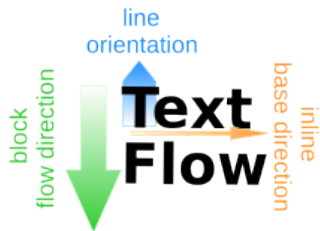
```
#begin{minipage}(30ex)
\parindent=1.2em
\let\vsurindent=\noindent
\yajikesi(30)
「段落の始まり」のカギ括弧を全角二分下がりで囲む
のは、宛直書きがはじまつてから多くなりました。
これは、多分、カギ括弧が「全角一文字分のなかに二分
下がり」ともレイアウトされた文字」(全角カギ括弧)と
して用意されていたからと推定されます。したがって、
作業の容易さから、「折り返し」は天ツキではなく二分
下がりとなります。治波所蔵などはこの方式をとって
います。 \yspar
一方、活字のときは、カギ括弧は二分で囲ままれており、
別に二分のスペースが用意されていました。この場合、
```

出力結果を次に示す。

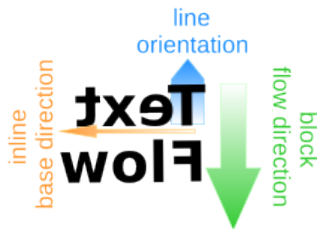
```
作業の容易さから、「段落の始まり」は全角の詰めもの
と二分のカギ括弧を組むことになるので、全角下がりと
なります。「折り返し」は空きなしで組みますから、
この方式では天ツキとなります。JIS規格はこの方式
を標準として採用しています。 \yspar
全角と半角のカギ括弧が、両方用意されているならば、
「段落の始まり」では、全角のカギ括弧にし、
「折り返し」では、半角のを使うのが簡便です。
なぜなら、「段落の始まり」で全角の空きをおく必要
がないからです。中公新書や講談社現代新書はこの
方式です。
\yajikesi(30)
\end{minipage}
```

Previous work: CSS (1)

Horizontal Text, Cascading Style Sheets (CSS) for HTML.



English, French, ...



Arabic, Hebrew, ...

Introduction of general terminology.

Previous work: CSS (2)

Vertical Text, Cascading Style Sheets (CSS) for HTML.



Japanese, Chinese, ...



Mongolian, Uighur, ...

<http://www.w3.org/TR/css3-writing-modes>

Previous work: Omega (1)

In 2001, I was working on Omega with Yannis Haralambous. We proposed an approach for multiple direction documents.

A writing direction could be defined by three components:

1. The “top” of each page.
2. The “left” of each line.
3. The “top” of each character.

John Plaice, Yannis Haralambous. Writing in multiple directions in Omega. *Fourth International Symposium on Multilingual Computing*, March 1–3, 2001, Tokyo, Japan.

Yannis Haralambous, John Plaice. The Omega Typesetting and Document Processing System (in Japanese). *BIT* 4:173–152, 2001.

Previous work: Omega (2)

TLT — Left-right (LR) scripts, horizontal CJK.

TRT — Right-left (RL) scripts.

RTT — Vertical CJK, upright LR scripts in vertical CJK.

LTL — Mongolian and Uighur (MU).

RTL — MU scripts in vertical CJK.

RTR — Rotated LR scripts in vertical CJK.

LTR — Rotated LR scripts in MU.

LTT — Vertical CJK in MU.

Previous work: Omega (3)

Vertical CJK: CJK scripts use $\begin{matrix} \text{use} \\ \text{R} \\ \text{TLT} \\ \text{T} \end{matrix}$, LR scripts use $\begin{matrix} \text{use} \\ \text{RTL} \\ \text{R} \\ \text{T} \end{matrix}$, RL scripts use $\begin{matrix} \text{use} \\ \text{LTL} \\ \text{L} \\ \text{T} \end{matrix}$ (RBR), and MU scripts use $\begin{matrix} \text{use} \\ \text{RTL} \\ \text{R} \\ \text{T} \end{matrix}$ (RTL).

Horizontal: LR and CJK scripts use TLT, and RL and MU scripts “ТЯТ эиТ” (TRT).

Vertical MU: MU and RL scripts use $\begin{matrix} \text{use} \\ \text{LTL} \\ \text{L} \\ \text{T} \end{matrix}$ (LTL), LR scripts use $\begin{matrix} \text{use} \\ \text{RTL} \\ \text{R} \\ \text{T} \end{matrix}$ and CJK scripts use $\begin{matrix} \text{use} \\ \text{TLT} \\ \text{T} \\ \text{T} \end{matrix}$.

図2. さまざまな方向のテキストの例

テキスト全体は左から右へ流れているが、違う向きに書かれるいろいろな部分を含んでいる。各行は、書字方向が混在するもっともありふれた場合の構造を説明している。

Each writing direction required different fonts.

My thesis: All this can be simplified

- ▶ A writing system should be invariant to rotation.
- ▶ Writing should not be defined by the box holding it.
- ▶ For example, typesetting along a curve requires no boxes.
- ▶ The different kinds of writing: *baseline* (*LR* and *RL*) and *axial*.
- ▶ Baseline writing is naturally horizontal: glyphs are hung from or placed on the baseline, the direction of writing is perpendicular to glyph orientation.
- ▶ Axial writing is naturally vertical: glyphs are attached to the axis, the direction of writing is opposite to glyph orientation.
- ▶ The line orientation can be *positive* or *negative*.

The three different kinds of writing

This is baseline left-to-right writing.

This is baseline right-to-left writing.

W R I T I N G °
T H I S I S A X I A L

Embeddings

In the following six slides, text A will be embedded in text B :

A. **pneumonoultramicroscopicsilicovolcanoconiosis**

B. I am glad . . . is not well known.

Texts A and B will be using different combinations of writing style.
The choice of text A : long, with many possible hyphenation points.

Embedding a baseline text in another baseline text

I am glad **pneumonoultramicroscopicsilicovol-
canoconiosis** is not well known.

-lovocilizicqpscovicimertluonumeng bslg ms l
.nwonk llw ton zi **canoconiosis**

pneumonoultramicroscopicsilicovol- bslg ms l
.nwonk llw ton zi **canoconiosis**

- ▶ Depending on the fonts, vertical adjustment of the baseline of the embedded text may be necessary.
- ▶ In mixed-mode, the embedded text can be mirrored or the $\text{T}_{\text{E}}\text{X}-\text{X}_{\text{E}}\text{T}$ bidirectional paragrapher can be used.

Embedding a baseline text in an axial text (1)

K **N** **O** **W** **N** **°**
osis **I** **S** **N** **O** **T** **W** **E** **L** **L**
ultramicroscopic**silico****volcano****coni-**
I **A** **M** **G** **L** **A** **D** **pneumono-**

K **N** **O** **W** **N** **°**
osis **I** **S** **N** **O** **T** **W** **E** **L** **L**
ultramicroscopic**silico****volcano****coni-**
I **A** **M** **G** **L** **A** **D** **pneumono-**

- ▶ If the embedded text is rotated, it can be rotated -90° or 90° .

Embedding a baseline text in an axial text (2)

K	volcano-	I
N	coniosis	A
O	I	M
W	S	G
N	N	L
°	O	A
	T	D
	W	pneumo-
	E	noultra-
	L	microscop-
	L	icsilico-

- ▶ If the embedded text is not rotated, its width must be stated.
- ▶ An inner paragraph is required within the outer paragraph.

Embedding an axial text in a baseline text (1)

I am glad P N E U M O N O S U L T R A
M I C R O S C O P I C S I L I C A
V O L C A N O C O N I O S I S is
not well known.

I am glad A R T L I U N O S M U R E N I P
O S C I L I S I C O N P O S C A S R O I P M
S I S I O N I C O P O N A C L O R V is
not well known.

- ▶ If the embedded text is rotated, it can be rotated 90° or -90° .

Embedding an axial text in a baseline text (2)

		P	O	T	C	O	I	V
		N	N	R	R	P	L	O
I am glad		E	O	A	O	I	I	L
		U	U	M	S	C	C	C
		M	L	I	C	S	O	A
N	I							
O	O							
C	S	is not well known.						
O	I							
N	S							

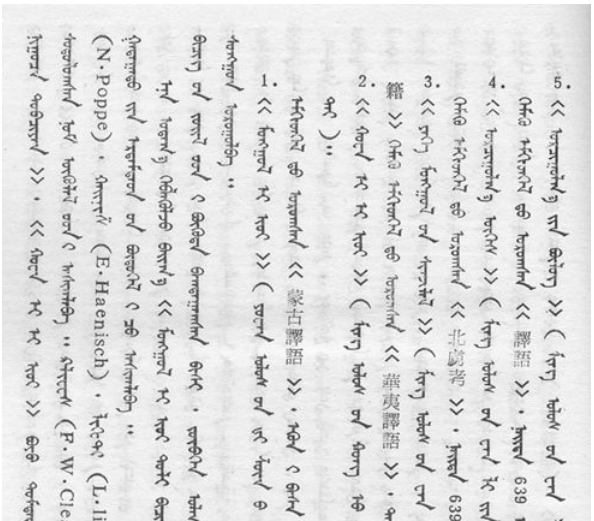
- ▶ If the embedded text is not rotated, its height must be stated.
- ▶ An inner paragrapher is required within the outer paragrapher.

Embedding an axial text in a baseline text (3)

		V	I	O	C	T	O	P
		O	L	P	R	R	N	N
I am glad		L	I	I	O	A	O	E
		C	C	C	S	M	U	U
		A	O	S	C	I	L	M
I	N							
O	O							
S	C	is not well known.						
I	O							
S	N							

- ▶ The embedded text can be inserted as LR text or RL text.

The importance of line orientation (1)



The line orientation of the Mongolian is negative, and of the Chinese and English is positive. Annotations should go to the left of the lines.

The importance of line orientation (2)

は、組版する文書にフィルタを^{せん たくてき}選択的に^{てきよう}適用することが
できる（^{せん たくてき}選択的とは、どんなフィルタをいつでも使ったり止めたり
できるといことである。たとえば^{げん ご}言語ごと、^{こう せい よう そ}構成要素ごと
に^{ちが}違うフィルタを^{てきよう}適用することができる）。この能力は、NLP
（^{し ぜん げん ご しょり}自然言語処理）の手法を用いて、^{しゆ ほう}組版の^{くみ はん}新たな^{あら}地平^{ち へい}を^{ひら}拓くこと
となった。実際、^{じっさい}文書の^{ぶん しょ}加工^{か こう}に特別な^{とく べつ}自然言語向け^{し ぜん げん ご む}ソフトウェア
を^{ひつ よう}必要とする^{ばあい}場合があるだろう。単にその^{たん}言語の^{げん ご}単語の^{たん ご}コーパス
で^す済む場合も、^{ほん とう}本当の^{けい たい}形態素^{そ かい}解析^{せき}を^{よう}要する^{ばあい}場合もあるだろうけ
れど。

Yannis Haralambous and John Plaice. Automatic language and typesetting processing with Omega (in French). *Cahiers GUTenberg* 39–40:139–166, May 2001.

The importance of line orientation (3)

萬延元年三月

八

Example from talk
of NAKANO Ken.

鷹司殿・近衛殿・三條殿等御眞御落飾被遊候様取計、其他諸大夫始何一ツ罪科無之ものを召

捕、關東^(一)に差下し、夫々非道之所置いたし、専虎狼之威ヲ以天下を屏息せしめ、畿内之開港

并邪教寺取建等本條約差許候得^(二)も、青蓮院宮様御英邁を奉忌御失徳有之様申觸、御寺務取放

奉幽閉候所業、乍恐 玉躰も奉迫候趣顯然^(三)にふ、北條・足利之暴横之均く、不共戴天之國

賊と云へし、嗚呼此儘に打過な^(四)、赫々たる神州、一兩年を不出、内地之奸民邪教を靡キ、彼

が勢焰を助け

皇國之奸賊平身低頭して彼か正朔を奉する事、掌の上^(五)に視るか如し、苟も人心有之もの、實

に痛哭長大息に不堪事ならずや、雖然、

東照宮之德澤未地に墜^(六)ず、御三家御一門^(七)を尾張殿・水戸殿・一橋殿・越前家・阿波家・因

州家之如き、徳川家輔佐之良將も有之、外諸侯も薩州・仙臺・福岡・佐賀・長州・土州・

宇和島・柳川等天下之爲忠憤之念日夜不息、有名之諸侯も不少候得去、内^(八)に則御加門方將

軍家を奉補佐専内政を脩め、外^(九)に則有名之諸侯一意に忠力を盡し武備を整な^(一〇)、神州之恥辱

を一洗して、叡慮を奉要候事天地神明に誓ひ疑有まし、依之當今事態々概略を記して天下

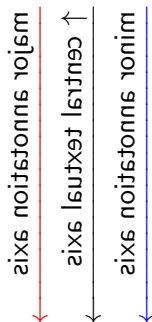
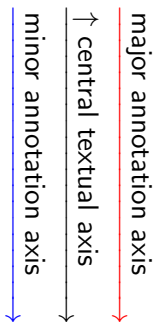
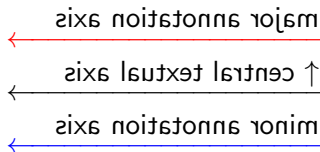
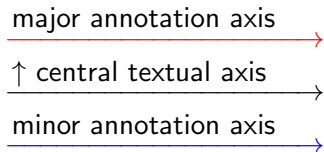
之公論折衷を待、左袒して天下を興起せんと欲する所也、周之衰る婦人すら不恤諱して周家

The importance of line orientation (4)

北條・足利之暴横均く、
(義時、承久三年後鳥羽上皇ノ軍ヲ破ル)
(尊氏、後醍醐天皇ニ叛ク)〔し〕

Note that there are two annotation lines, one to the right, one to the left.
Magnification of previous example.

Multiple parallel axes and line orientation (1)



Positive orientation

Negative orientation

Multiple parallel axes and line orientation (2)

- ▶ A simple running text is a set of multiple interacting streams.
- ▶ A more general setting is discussed in: Blanca Mancilla, Jarryd P. Beck, John Plaice: Typesetting multiple interacting streams. *ACM Symposium on Document Engineering 2012*: 149-152, Paris, France.

Adapting existing T_EX engines

- ▶ Limited number of primitives:
 - ▶ mirror text;
 - ▶ rotate clockwise;
 - ▶ rotate counterclockwise.
- ▶ Limited number of parameters:
 - ▶ numeric: shifting the base line;
 - ▶ Boolean: mirror or not, rotation or not, clockwise or counterclockwise;
 - ▶ direction: text, mathematics, paragraph, page body, page.
- ▶ Most of the Omega code is still relevant.
- ▶ Both pT_EX and XeT_EX can be extended in these ways.
- ▶ We need to add a paragrapher within the paragrapher.
- ▶ New kinds of embeddings can be defined.
- ▶ Adding multiple interacting streams is non-trivial.