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

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

Lines are read down, from right to left. Mongolian is like a Middle-Eastern script, rotated 90°. Chinese is vertical. English is rotated -90°. Numbers are horizontal.
Each hbox or vbox can be in *yoko* (normal) or *tate* direction. Hisato Hamano, *Vertical Typesetting with TeX*, *TUGboat* 11(3):346–352, 1990.
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.


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 \( \text{LTT}_R \), LR scripts \( \text{LTR} \), RL scripts \( \text{RTL} \), (RBR), and MU scripts \( \text{TRT} \), (RTL).

Horizontal: LR and CJK scripts use TLT, and RL and MU scripts "TRT əə" (TRT).

Vertical MU: MU and RL scripts \( \text{LTT}_L \), LR scripts \( \text{LTR} \) and CJK scripts \( \text{LTT}_L \).

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.
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 *pneumonoultramicroscopicsilicovolcanoconiosis* is not well known.

- 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 \( \TeX \text{-X\TeX} \) bidirectional paragrapher can be used.
Embedding a baseline text in an axial text (1)

- If the embedded text is rotated, it can be rotated $-90^\circ$ or $90^\circ$. 
Embedding a baseline text in an axial text (2)

| K | K | volcano- | I |
| N | N | coniosis | A |
| O | O |         | M |
| W | W |         | G |
| N | N |         | L |
| ° | ° |         | A |
| T | T |         | D |
| W | W | pneumo- | N |
| E | E | noultra- | U |
| L | L | microsco- | P |
| L | L | picsilico-|

- If the embedded text is not rotated, its width must be stated.
- An inner paragrapher is required within the outer paragrapher.
Embedding an axial text in a baseline text (1)

I am glad you can't, so is not well known.

▶ If the embedded text is rotated, it can be rotated $90^\circ$ or $-90^\circ$. 
Embedding an axial text in a baseline text (2)

I am glad

\[
\begin{array}{cccccccc}
P & O & T & C & O & I & V \\
N & N & R & R & P & L & O \\
E & O & A & O & I & I & L \\
U & U & M & S & C & C & C \\
M & L & I & C & S & O & A \\
N & I & O & O \\
C & S & I & O \\
N & S \\
\end{array}
\]

is not well known.

- 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)

I am glad is not well known.

▶ The embedded text can be inserted as LR text or RL text.
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選択的に適用することがで
きる（選択的とは,どんなフィルタをいつでも使ったり止めた
りできるということである.たとえば言語ごと,構成要素ごと
に違うフィルタを適用することができる).この能力は, NLP
（自然言語処理）の手法を用いて,組版の新たな地平を拓くこと
となった. 実際,文書の加工に特別な自然言語向けソフトウェア
を必要とする場合があるだろう. 単にその言語の単語のコーパス
で済む場合も, 本当の形態素解析を要する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
The importance of line orientation (3)

Example from talk of NAKANO Ken.
The importance of line orientation (4)

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.
Adapting existing \TeX\ 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 p\TeX\ and Xe\TeX\ can be extended in these ways.
- We need to add a paragger within the paragger.
- New kinds of embeddings can be defined.
- Adding multiple interacting streams is non-trivial.