

Comparative Analysis of W3C Text Layout Requirements



Behnam Esfahbod

Virgule Typeworks / W3C i18n Activity

Internationalization & Unicode Conference 41 — October 17, 2017

Abstract

W3C's Internationalization Activity is documenting text layout and typographic needs of various writing systems and languages such as Arabic, Chinese, Ethiopic, Hebrew, Indic, and Japanese for better support in web technologies and other e-publications. Task Forces are formed by experts from around the world to document these requirements and the existing gaps between the real-world expectation and existing solutions.

In this talk, we'll have a review of the Task Forces' activities and comparative analysis of their publications, looking at problems and solutions unique to each writing system and/or language, and how common problems are addressed differently. Major requirement areas covered are: script overview, document compositions, page formats, paragraph and line compositions and adjustments, inline features, and special cases.

This talk

- Introduction
- Task Forces & Requirement Documents
- Comparative Analysis
- Conclusions & Future Work

Introduction

About me



Co-Chair of Arabic Layout Task Force @ W3C i18n Activity



Founder & Lead Engineer @ Virgule Typeworks



UNiC — Unicode and Internationalization Crates for Rust

Previously: Facebook, IRNIC, FarsiWeb

W3C Specification Maturity Levels of a Technical Report

Editor's Draft

The document as being drafted by the editors.

↳ Working Draft (WD)

Published for review by the community, including W3C Members, the public, and other technical organizations

↳ Working Group Note

Published by a chartered Working Group to indicate that work has ended on a particular topic.

↳ Candidate Recommendation (CR)

Has been widely reviewed and satisfies the Working Group's technical requirements.

↳ Proposed Recommendation (PR)

A mature technical report that, after wide review for technical soundness and implementability.

↳ W3C Recommendation (REC)

A specification or set of guidelines that has received the endorsement of W3C.

W3C i18n Activity

- Mission: formats and protocols that are usable worldwide in all languages and in all writing systems.
- Works with W3C working groups and liaises with other organizations.
 - <https://www.w3.org/International/about>
- Project RADAR
 - Layout & Typography Requirements, for languages/scripts/writing systems/cultures.
 - Developer support
 - <https://github.com/w3c/i18n-discuss>
 - <https://www.w3.org/International/tests/>
 - <https://w3c.github.io/i18n-activity/reviews/>
 - Education & outreach
 - <https://www.w3.org/International/articlelist>

Layout & Typography @ Open Web Platform

- Web pages and eBooks should look & behave as people expect around the world.
- Typography Index
 - <https://github.com/w3c/typography/>
 - <https://github.com/w3c/type-samples/>
- Requirements documents
 - Languages and Writing Systems <https://www.w3.org/International/layout#scope>
 - Arabic Mathematical notation <https://www.w3.org/TR/arabic-math/>
 - Ready-made International Counter Styles <https://github.com/w3c/predefined-counter-styles/>
- Gap analysis documents
 - Current Status of Japanese Typography Using Web Technologies
 - Mongolian variant forms <https://r12a.github.io/mongolian-variants/>
 - Additional Requirements for Bidi in HTML & CSS <https://www.w3.org/TR/html-bidi/>

Task Forces & Requirement Documents

Writing Systems

Region	Writing System	Language(s)	Task Force	Latest Document
East Asia (CJK)	Japanese	Japanese	2007	Working Group Note (2012)
	Chinese	Mandarin, ...	2015	Editor's Draft (2017)
	Hangul	Korean	pre-TF	Working Draft (2015) (inactive)
South Asia	Indic	Hindi, ...	2012	Working Draft (2017)
	Tibetan	Standard Tibetan	2017	Editor's Draft (2017)
West Asia & North Africa	Arabic	Persian, Arabic	2015	Editor's Draft (2017)
	Hebrew	Hebrew	2017	— (just started)
NE Africa	Ethiopic (Ge'ez)	Ethiopic languages	2015	Editor's Draft (2016)
Europe	Latin	European languages	(2013)	Working Draft (2014)

Japanese

- Japanese Layout Task Force (2007)
 - <https://www.w3.org/2007/02/japanese-layout/>
- Requirements for Japanese Text Layout
 - Working Group Notes: First Edition (2009), Second Edition (2012)
 - English documents, plus Japanese translations
 - <https://www.w3.org/TR/jlreq/>
- eBooks & i18n: Richer Internationalization for eBooks (Tokyo, 2013)
 - Top two concerns for Japanese participants: vertical text and ruby
 - <https://www.w3.org/2013/06/ebooks/report>
- Current Status of Japanese Typography Using Web Technologies
 - W3C Member Submission (January 2017)
 - <https://www.w3.org/Submission/2017/SUBM-CSJTUWT-20170102/>

2 Basics of Japanese Composition

2.1 Characters and the Principles of Setting them for Japanese Composition

2.1.1 Characters Used for Japanese Composition

Japanese letters used for composing Japanese text mainly consist of [ideographic \(cl-19\)](#), [hiragana \(cl-15\)](#) and [katakana \(cl-16\)](#) characters (see [Fig. 2.1](#)).

ideographic	安以宇衣於
	阿伊宇江於
hiragana	あいうえお
katakana	アイウエオ

Fig. 2.1: Kanji, hiragana and katakana.

(note 1) In addition to [ideographic \(cl-19\)](#), [hiragana \(cl-15\)](#) and [katakana \(cl-16\)](#) characters, various [punctuation marks](#) (see [Fig. 2.2](#)) as well as [Western characters \(cl-27\)](#), such as [European numerals](#), Latin letters and/or Greek letters, may be used in Japanese text. In this document these characters are classified into character classes, for which explanations are given describing their behavior in type-setting.

opening brackets	‘ “ ([[{ < 《 「 『 【
closing brackets	’ ”)]] } > 》 」 』 』
hyphens	— ~ -
dividing punctuation marks	? !
middle dots	• ; :
full stops	○ .
commas	、 ,
inseparable characters	—•••••

Fig. 2.2: Examples of punctuation marks.

Chinese

- Chinese Layout Task Force (2015)
 - <https://www.w3.org/International/groups/chinese-layout/>

- Requirements for Chinese Text Layout
 - Editor's Draft (2017)
 - English and Traditional/Simplified Chinese bilingual document
 - <https://w3c.github.io/clreq/>

2.2 Typefaces for Chinese Characters | 中文排版常用字体 | 中文排版常用字體

2.2.1 Four frequently-used Typefaces for Chinese Characters | 中文排版经常使用的四种字体 | 中文排版經常使用的四種字體

There are four main typefaces in use for Chinese characters:

中文排版时，主要使用的四种字体为：

中文排版時，主要使用的四種字體為：

- Song | 宋体 | 宋體
- Kai | 楷体 | 楷體
- Hei | 黑体 | 黑體
- Fangsong | 仿宋体 | 仿宋體

The following sections describe common practice and contexts for the use of these four typefaces.

这四种字体常用于书籍排版中，下列各节分别介绍其使用情境。

這四種字體常用於書籍排版中，下列各節分別介紹其使用情境。

2.2.2 Song | 宋体 | 宋體

關關雎鳩、在河之洲；
窈窕淑女、君子好逑。

宋體-繁

关关雎鸠、在河之洲；
窈窕淑女、君子好逑。

宋体-简

關關雎鳩、在河之洲；
窈窕淑女、君子好逑。

Adobe 明體

关关雎鸠、在河之洲；
窈窕淑女、君子好逑。

方正大标宋

Figure 1: Song | 宋体 | 宋體

Hangul (Korean)

- There has never been a Korean Layout Task Force. A group of volunteer experts produced the requirement document in late 2010's.
- Requirements for Hangul Text Layout and Typography (Working Draft, 2015)
 - Working Draft (2015)
 - English and Korean bilingual document
 - <https://www.w3.org/TR/klreq/>

2.3.1 Arrangement of 'Letter Face Position in Character Frame' for Full Width Parentheses | 전각 괄호의 '글자틀 내 글자면 위치' 지정

In horizontal writing, the letter face of a full width opening parenthesis is placed on the right end of the character frame, and the left space is considered a user controlled area. In vertical writing, the letter face of a full width opening parenthesis is placed on the bottom end of character frame, and the space is considered a user controlled area.

가로짜기의 경우 전각 열기 괄호의 글자면은 글자들의 오른쪽(가로짜기)·아랫쪽(세로짜기) 끝에 두고, 왼쪽 여백은 사용자 조정 영역으로 간주한다. 세로짜기의 경우 전각 열기 괄호의 글자면은 글자들의 아랫쪽 끝에 두고, 여백은 사용자 조정 영역으로 간주한다.

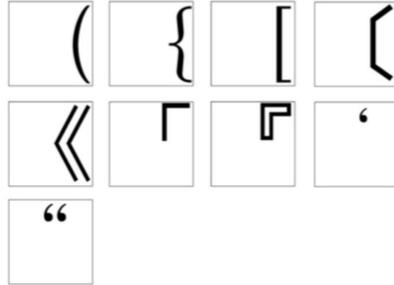


Fig. 4: Full-width opening parentheses and quotation marks positioned in the character frame ({ [[⎵ ' ' ' ' ' ' (U+FF08, U+FF5B, U+FF3B, U+3014, U+300A, U+300C, U+300E, U+ 02BB, U+201C) | 전각 열기 괄호와 굵은 열기 따옴표의 글자틀 내 위치 ({ [⌠⎵' ' ' ' ' ' (U+FF08, U+FF5B, U+FF3B, U+3014, U+300A, U+300C, U+300E, U+ 02BB, U+201C)

In horizontal writing, the letter face of a full width closing parenthesis is placed on the left end of the character frame, and the space is considered a user controlled area. In vertical writing, the letter face of a full width closing parenthesis is placed on the top end of the character frame, and the space is considered a user controlled area.

가로짜기의 경우 전각 닫기 괄호의 글자면은 글자들의 왼쪽(가로짜기)·윗쪽(세로짜기) 끝에 두고, 오른쪽 여백은 사용자 조정 영역으로 간주한다. 세로짜기의 경우 전각 닫기 괄호의 글자면은 글자들의 윗쪽 끝에 두고, 여백은 사용자 조정 영역으로 간주한다.

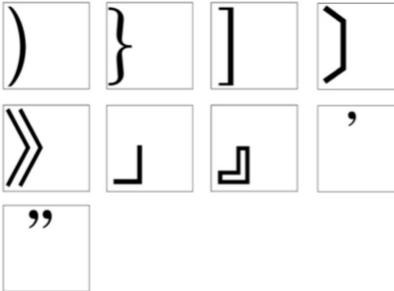


Fig. 5: Full-width closing parentheses and quotation marks positioned in the character frame) } } } ' ' ' ' ' ' (U+FF09, U+FF5D, U+FF3D, U+3015, U+300B, U+300D, U+300F, U+02BC, U+201D) | 전각 닫기 괄호와 굵은 닫기 따옴표의 글자틀 내 위치) } } } ' ' ' ' ' ' (U+FF09, U+FF5D, U+FF3D, U+3015, U+300B, U+300D, U+300F, U+02BC, U+201D)

Indic

- Indic Layout Task Force (2012)
 - <https://www.w3.org/International/groups/indic-layout/>

- Indic Layout Requirements
 - Working Draft (2017)
 - English document
 - Focused largely on the Devanagari script, with the goal to expand to 22 constitutionally recognized languages and 12 scripts
 - <https://www.w3.org/TR/ilreq/>

3.2 Line breaking

When inline-level content is laid out into lines, it is broken across line boxes. Such a break is called a line break. In most writing systems, in the absence of hyphenation a line break occurs only at word boundaries. Many writing systems use spaces or punctuation to explicitly separate words, and line break opportunities can be identified by these characters. Line breaking, also known as word wrapping, is the process of breaking a section of text into lines such that it will fit in the available width of a page, window or other display area.

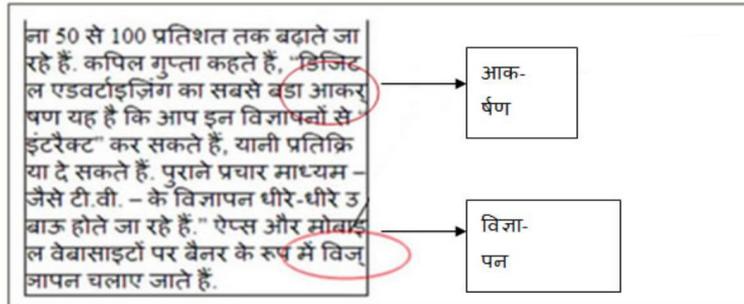
3.2.1 Hyphenation

There are different cases of hyphenation, some of the cases are given below :

Case 1 : Hyphens are commonly used in Copulative compounds words in Hindi language. Hindi has both prefix- and suffixes which are joined to words with a hyphen.

नर-नारी, लाम-हानि, माता-पिता, ऊंच-नीच

Case 2: Single word can breaks at the end of the line follow Indic orthographic syllable using hyphen. Following example shows correct representation of word आकर्षण and विज्ञापन using hyphen :



Tibetan

- Tibetan Layout Task Force (2017)
 - <https://w3c.github.io/i18n-activity/tlreq/>

- Requirements for Tibetan Text Layout and Typography
 - Editor's Draft (2017)
 - English document
 - <https://w3c.github.io/tlreq/>

2.3 Tibetan Syllables | 藏文音节

Word boundaries within a section are not indicated, only 'syllables', known as tsheg-bar /tsek bar/. Syllable boundaries are usually separated by the tsek character, U+0F0B TIBETAN MARK INTER-SYLLABIC TSHEG །.

The pronunciation of Tibetan words is typically much simpler than the orthography, which involves patterns of consonants. These patterns reduce ambiguity and can affect pronunciation and tone.

The following diagram shows characters in all of the syllabic positions, and lists the characters that can appear in each of the non-root locations. The two-syllable word in the example is འགྲེམས་སྟོན་ 'grem-s-ton /djem-ton/ (exhibition).



Figure 6: Syllable composition in Tibetan

Arabic Script

- Arabic Layout Task Force (2015)
 - Two languages: Persian and Arabic
 - <https://w3c.github.io/i18n-activity/alreq/>

- Text Layout Requirements for the Arabic Script
 - Editor's Draft (2017)
 - English only, with Persian and Arabic glossary
 - <https://w3c.github.io/alreq/>

2.3 Direction

Arabic script is written from right to left. Numbers, even Arabic numbers, are written from left to right, as is text in a script that is normally left-to-right.

When the main script is Arabic, the layout and structure of pages and documents are also set from right to left.

Unicode Bidirectional Algorithm (or ‘bidi algorithm’, for short) [BIDI] details an algorithm for rendering right-to-left text and covers a myriad of situations in mixing different kinds of characters. A simpler explanation of the basics of the algorithm exists in the W3C article [Unicode Bidirectional Algorithm basics](#). [UBA-BASICS] You can refer to these documents for more information about Unicode’s bidirectional algorithm.

A brief overview of the [bidirectional algorithm](#) follows, because the direction is an essential part of how Arabic script is used.

The characters of a text are digitally stored and transferred in the same order that they are typed by a user. This is the order in which the text is read and pronounced by people and held in memory by software applications, as shown in [Figure 1](#) for a sample text.

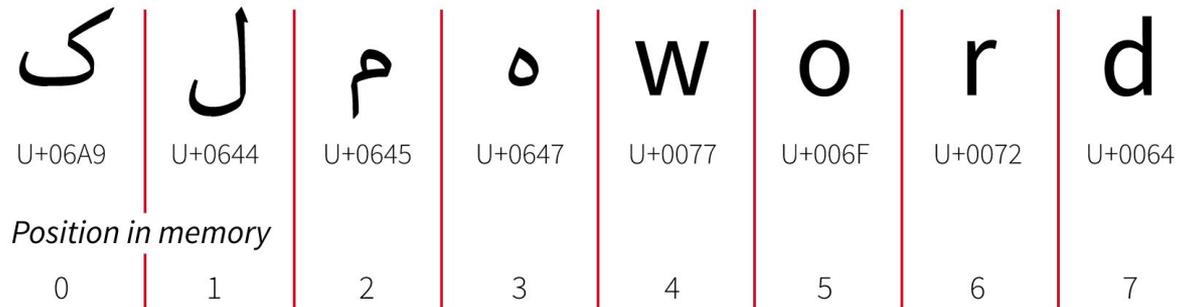


Figure 1: The order of characters in memory

Hebrew

- Hebrew Layout Task Force (2017)
 - Early stages
 - <https://w3c.github.io/i18n-activity/hlreq/>

Ethiopic (Ge'ez)

- Ethiopic Layout Task Force (2015)
 - <https://www.w3.org/International/groups/ethiopic-layout/>

- Ethiopic Layout Requirements
 - Editor's Draft (2016)
 - <https://w3c.github.io/elreq/>

2.3.5 Optical Balancing

Foreign language words or phrases are regularly found inline within a paragraph of Ethiopic text, often bounded within enclosing punctuation such as brackets and quotation marks (e.g. []()""«»«»). This practice is most often observed in news articles on international topics. The weight of the enclosing punctuation may found as matching either the Ethiopic or Latin weight. The preference of stakeholders must be determined here. Comparative samples follow:

ካገላበጥኳቸው ስራዎች ውስጥ ግን ታዩ አሰፋ እና ሽፈራው በቀለ የተባሉ ምሁራን በጋራ ያሳተሙት «The Study of Amharic Literature: An Overview» የሚለው ጥናት አጥጋቢ መልስ የሚሰጥ ሆኖ አግኝቼዋለሁ።

Figure 10: Latin text enclosed with Latin weight guillemets.

ካገላበጥኳቸው ስራዎች ውስጥ ግን ታዩ አሰፋ እና ሽፈራው በቀለ የተባሉ ምሁራን በጋራ ያሳተሙት «The Study of Amharic Literature: An Overview» የሚለው ጥናት አጥጋቢ መልስ የሚሰጥ ሆኖ አግኝቼዋለሁ።

Figure 11: Latin text enclosed with Ethiopic weight guillemets.

ካገላበጥኳቸው ስራዎች ውስጥ ግን ታዩ አሰፋ እና ሽፈራው በቀለ የተባሉ ምሁራን በጋራ ያሳተሙት “The Study of Amharic Literature: An Overview” የሚለው ጥናት አጥጋቢ መልስ የሚሰጥ ሆኖ አግኝቼዋለሁ።

Figure 12: Latin text enclosed with Latin weight quotes.

ካገላበጥኳቸው ስራዎች ውስጥ ግን ታዩ አሰፋ እና ሽፈራው በቀለ የተባሉ ምሁራን በጋራ ያሳተሙት “The Study of Amharic Literature: An Overview” የሚለው ጥናት አጥጋቢ መልስ የሚሰጥ ሆኖ አግኝቼዋለሁ።

Figure 13: Latin text enclosed with Ethiopic weight quotes.

Latin Writing System

- Digital Publishing Interest Group
 - Which became part of Publishing@W3C, born out of the merger of IDPF and W3C (2017)
 - <https://www.w3.org/publishing/>

- Requirements for Latin Text Layout and Pagination
 - Working Draft (2014)
 - <https://www.w3.org/TR/dpub-latinreq/>

7. Initial Capitals

Large, decorative letters have been used to start new sections of text since long before printing. In fact, their use predates lowercase letters entirely.

7.1 Drop caps

A **drop cap** is a larger-than-usual letter at the start of a paragraph, with a baseline at least one line lower than the first baseline of the paragraph. The size of drop caps is usually indicated by how many lines they occupy—two-line and three-line drop caps are the most common.

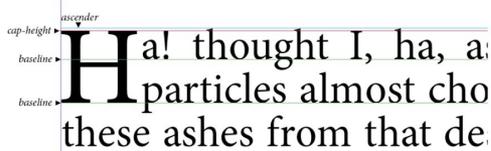


Fig. 7 Two-line drop cap

Aligning the letter vertically is a challenge. The cap height of the letter should align with the cap height of the first line of text. The baseline of the letter should fall on the baseline of one of the following lines (the second for a 2-line drop cap, etc.).

Call me Ishmael. Some years ago—never mind how long precisely—having little or no money in my purse, and nothing particular to interest me on shore, I thought I would sail about a little and see the watery part of the world. It is a way I have of driving off the spleen and regulating the circulation. Whenever I find myself growing

Fig. 8 Three-line drop cap

The horizontal position of the drop cap and the surrounding text is also an issue, as variations in glyph shapes may require increasing or decreasing space to the right of the drop cap, and in some cases separate adjustments may be required for each line adjacent to the drop cap.

A day or two passed, and there was great activity aboard the Pequod. Not only were the old sails being mended, but new sails were coming on board, and bolts of canvas, and coils of rigging; in short, everything betokened that the ship's preparations were hurrying to a close. Captain Peleg seldom or never went ashore, but sat in his wigwam keeping a sharp look-out upon the hands:

Fig. 9 Drop cap without runaround

A day or two passed, and there was great activity aboard the Pequod. Not only were the old sails being mended, but new sails were coming on board, and bolts of canvas, and coils of rigging; in short, everything betokened that the ship's preparations were hurrying to a close. Captain Peleg seldom or never went ashore, but sat in his wigwam

Fig. 10 Drop cap with runaround

The position of a drop cap in relation to the left margin may also need to be adjusted. Letters like "C" may need to move left slightly to visually align with the left margin.

A drop cap may be desired on a paragraph which starts with a punctuation mark, most often a quotation mark. In this case, one option is to delete the quotation mark entirely.

Internationalization support on the Web (HTML/CSS)

- International text layout and typography index
 - Entry point for browser implementers and specification developers, pointing to relevant information in specifications, tests, and useful articles and papers
 - [Editor's Draft \(2017\)](#)
 - <https://w3c.github.io/typography/>

- Gap analysis checklist
 - Documents issues for scripts/languages in terms of support in specifications or user agents
 - [Draft Article \(2017\)](#)
 - <https://w3c.github.io/typography/gap-analysis/>
 - <https://w3c.github.io/typography/gap-analysis/language-matrix.html>

script	language	Level	Encoding	Fonts	Font styles	Glyph control	Cursive text	Char transforms	Numbers & digits	Boundaries & selection	Quotation	Inter-char spacing	Ruby	Text decoration	Emphasis & highlight	Bidi text	Other inline	Line breaking	Hyphenation	Justification & alignment	Lists, counters	Initial letter style	Baselines, inline align	Other paragraph	Bidi layout	Vertical text	Notes, footnotes	Page numbers, running heads	Other pagination
Arab	Arabic*	0.15	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
	Persian*	0.15	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Deva	Hindi*	0.89	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Ethi	Amharic*	0.55	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
	Tigrīña*	0.55	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Hang	Korean*	0.67	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Hani	SChinese*	0.5	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
	TChinese*	0.5	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Hebr	Hebrew*	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Japn	Japanese*	0.5	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Latn	English*	0.86	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
	Spanish*	0.91	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Tibt	Tibetan*	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
	Dzongkha*	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?

■ 3 All needs covered, or - Not applicable
■ 2 Basic needs covered, but work needed for advance publishing
■ 1 Can create interoperable web pages, but work still needed for basic features
■ 0 Something prevents interoperable or effective use of the language in webpages
 * next to a language name indicates a tentative score, pending validation by experts

Comparative Analysis

Characters & Phrases

Encoding Considerations

Region	Writing System	Status
East Asia (CJK)	Japanese	Ideographic (Kanji), Hiragana, Katakana, European letters, common punctuations, and European numerals.
	Chinese	Ideographic, European letters, common punctuations, and European numerals.
	Hangul	Ideographic (Hanja), Hangul Syllables, Hangul Jamo*, European letters, common punctuations, and European numerals.
South Asia	Indic	Devanagari & Devanagari Extended.
	Tibetan	Tibetan consonant, vowels, numerals, punctuations, and head marks.
W Asia & N Africa	Arabic	(Semantical) Arabic letters and diacritics, Arabic-Indic and Eastern Arabic-Indic numerals, joining behavior and joining control, bidirectional rendering.
NE Africa	Ethiopic (Ge'ez)	Ethiopic letters, numerals, punctuations,
Europe	Latin	—

Writing and Font Styles

Region	Writing System	Status
East Asia (CJK)	Japanese	—
	Chinese	Four main typefaces: Song, Kai, Hei, Fangsong.
	Hangul	Both fixed width and proportional width fonts. (Same codepoint sets.)
South Asia	Indic	—
	Tibetan	—
W Asia & N Africa	Arabic	Different Writing Styles: Kufi, Thuluth, Naskh, Ruq'a, Ta'liq, Diwani, Nasta'liq, Maghribi.
NE Africa	Ethiopic (Ge'ez)	—
Europe	Latin	—

關關雎鳩、在河之洲；
窈窕淑女、君子好逑。

思源黑體 (繁)

關關雎鳩、在河之洲；
窈窕淑女、君子好逑。

蘋方-繁

關關雎鳩、在河之洲；
窈窕淑女、君子好逑。

方正蘭亭黑 (繁)

关关雎鸠、在河之洲；
窈窕淑女、君子好逑。

思源黑体 (简)

关关雎鸠、在河之洲；
窈窕淑女、君子好逑。

苹方-简

关关雎鸠、在河之洲；
窈窕淑女、君子好逑。

方正兰亭黑 (简)

关关雎鸠、在河之洲；
窈窕淑女、君子好逑。

冬青黑体简体中文

Fig. 3: Hei

關關雎鳩、在河之洲；
窈窕淑女、君子好逑。

華康仿宋體

关关雎鸠、在河之洲；
窈窕淑女、君子好逑。

华文仿宋

Fig. 4: Fangsong (Imitation Song)

關關雎鳩、在河之洲；
窈窕淑女、君子好逑。

宋體-繁

關關雎鳩、在河之洲；
窈窕淑女、君子好逑。

Adobe 明體

关关雎鸠、在河之洲；
窈窕淑女、君子好逑。

宋体-简

关关雎鸠、在河之洲；
窈窕淑女、君子好逑。

方正大标宋

Fig. 1: Song

關關雎鳩、在河之洲；
窈窕淑女、君子好逑。

楷體-繁

關關雎鳩、在河之洲；
窈窕淑女、君子好逑。

教育部標準楷書

关关雎鸠、在河之洲；
窈窕淑女、君子好逑。

楷体-简

Fig. 2: Kai



Figure 16: Kufi example [Source].



Figure 18: Nask example [Source].



Figure 20: Ta'liq example [Source].



Figure 22: Nasta'liq example [Source].



Figure 17: Thuluth example [Source].

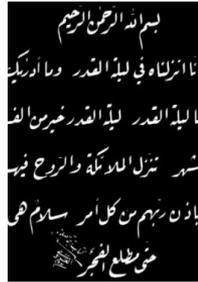


Figure 19: Ruq'a example [Source].



Figure 21: Diwani example [Source].

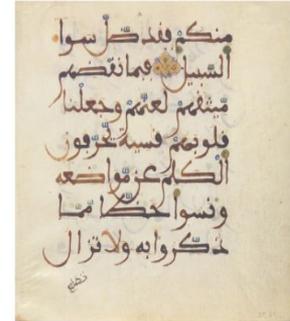
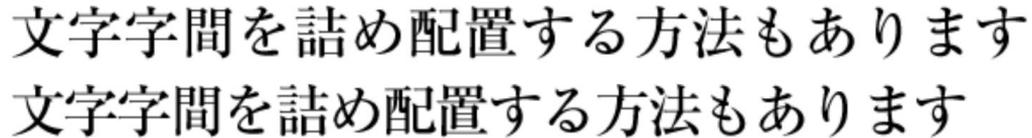


Figure 23: Maghribi example [Source].

Kerning & Ligature

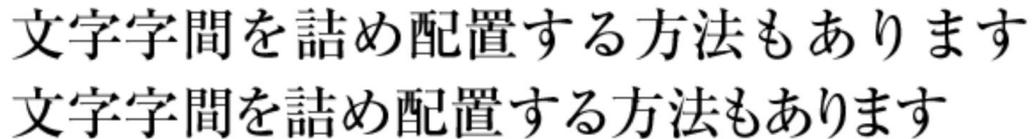
Region	Writing System	Status
East Asia (CJK)	Japanese	Tsumegumi (kerning/tracking): Text is set with negative inter-character space by arranging characters so that a portion of two character frames overlap each other. In the main text of books, the most reader-friendly approach is to use solid setting.
	Chinese	—
	Hangul	Hangul font kerning is adjusted by considering the inner space and contour region of Hangul syllables that are composed of Hangul Jamos.
South Asia	Indic	—
	Tibetan	—
W Asia & N Africa	Arabic	Almost all the writing styles of Arabic script use a special shape when letters lam and alef are joined. Most Arabic fonts include mandatory ligatures for this combination.
NE Africa	Ethiopic (Ge'ez)	Any number of kerning pairs and ligatures are possible for Ethiopic typography that would lead to better visual quality of printed literature.
Europe	Latin	—

- C. **Tsumegumi** (kerning / tracking): Text is set with negative inter-character space by arranging characters so that a portion of two character frames overlap each other. This is divided further into two types, depending on the methods used for inter-character space reduction. One method involves reducing by the same amount of inter-character space (**even tsumegumi** or tracking, see [Fig. 2.9](#)) and the other involves determining the amount of space to reduce based on the distance between the two letter faces of adjacent characters (**face tsumegumi** or letter face kerning, see [Fig. 2.10](#)).



文字字間を詰め配置する方法もあります
文字字間を詰め配置する方法もあります

Fig. 2.9: Example of even tsumegumi in horizontal writing mode. (The 1st line is the same text with solid setting, for comparison.)



文字字間を詰め配置する方法もあります
文字字間を詰め配置する方法もあります

Fig. 2.10: Example of face tsumegumi in horizontal writing mode. (The 1st line is the same text with solid setting, for comparison.)

In the **main text** of books, the most reader-friendly approach is to use **solid setting**. However, if the character size is larger, the distance between characters may become unbalanced, and tsumegumi will be applied. For example, there are books where tsumegumi is used with headings set in large character sizes. These methods are rarely used in books, since ease of reading is very important, but in magazines or advertisements there are many more examples of tsumegumi. Magazines tend to use type to differentiate themselves from others, and so devices like this are sometimes used for that purpose.

2.4.2 Group Kerning for Hangul Fonts

Group kerning can be applied for efficient adjustment of kerning for the 11,172 Hangul syllables. In order to apply Hangul group kerning, kerning groups are defined first, then groups are paired.

Set Kerning Group for Hangul Fonts	group name	characters	outline form	sample characters
	group 1	49		고 고 고 ... 흥 흥 흥 고 고 고 ... 흥 흥 흥
	group 2	108		노 노 노 ... 썸 썸 썸 노 노 노 ... 썸 썸 썸
	group 3	615		거 거 거 ... 힉 힉 힉 거 거 거 ... 힉 힉 힉
	group 4	125		가 가 가 ... 꺾 꺾 꺾 가 가 가 ... 꺾 꺾 꺾
	group 5	802		나 나 나 ... 힉 힉 힉 나 나 나 ... 힉 힉 힉
	group 6	132		나 내 냐 ... 여 여 여 나 내 냐 ... 여 여 여
	group 7	200		고 과 과 ... 휴 휴 휴 고 과 과 ... 휴 휴 휴
	group 8	33		거 거 거 ... 처 처 처 거 거 거 ... 처 처 처
	...			

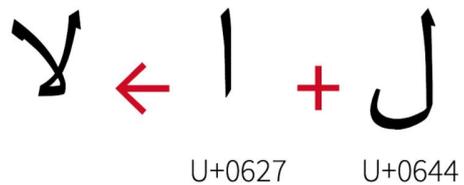
Fig. 8: Defining kerning groups for Hangul fonts

Apply Kerning Group for Hangul Fonts (unit: em/1000)



Fig. 9: Pairing kerning groups for Hangul fonts

Correct



Wrong

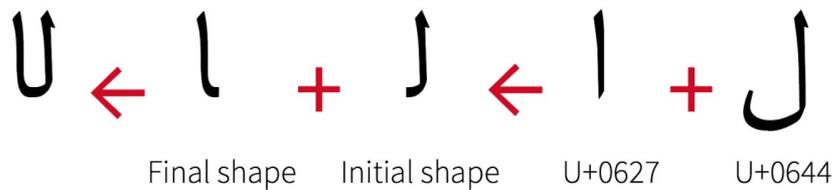


Figure 14: Correct and wrong ways of rendering letter lam followed by letter alef

Without Kerning	With Kerning
ኢትዮጵያ	ኢትዮጵያ
ምሥጋና	ምሥጋና
ምሥራች	ምሥራች
ቤተመንግሥት	ቤተመንግሥት
ሥራ	ሥራ
ዋጋ	ዋጋ
ዜና	ዜና

Figure 23: Examples of Common Words With and Without Kerning Applied

Decomposed Sequence	Ligature
ኤል	ኤል
ግዚ	ግዚ
ብሔ	ብሔ
ብር	ብር

Figure 24: Examples of Common Ligatures found in Calligraphical Manuscripts

Issues/Questions:

1. Is defining kerning pairs of interest to your organization?
2. Is defining ligatures of interest to your organization?

Lines & Paragraphs

Justification & Line-End Alignment

Region	Writing System	Status
East Asia (CJK)	Japanese	The Japanese "single line alignment method" is a process for setting alignment for a run of text that is shorter than a given line length. The intent of widow adjustment of paragraphs is to avoid that the last line of a paragraph contains less than a given number of characters.
	Chinese	(ditto)
	Hangul	(ditto)
South Asia	Indic	—
	Tibetan	There are two alternative methods of justification: Inter-character spacing, and Tsek padding.
W Asia & N Africa	Arabic	There are a number of different ways to produce justified text in Arabic. In some cases several of these methods may be combined. In other cases, certain methods are disallowed.
NE Africa	Ethiopic (Ge'ez)	Justification When SPACE is the Word Delimiter: like Western. Justification When ETHIOPIC WORDSPACE is the Word Delimiter: "word bound" and "centered" styles.
Europe	Latin	Hyphenation and Justification (TBD)

- Approaches to full justification <https://w3c.github.io/i18n-drafts/articles/typography/justification.en>

●centering

企画編集
原稿編集
本のデザイン
組版と校正
刷版の作製と印刷
製本

line head

line end

●line head alignment

企画編集
原稿編集
本のデザイン
組版と校正
刷版の作製と印刷
製本

line head

line end

●line end alignment

企画編集
原稿編集
本のデザイン
組版と校正
刷版の作製と印刷
製本

line head

line end

●even space distribution

企 画 編 集
原 稿 編 集
本 の デ ザ イ ン
組 版 と 校 正
刷 版 の 作 製 と 印 刷
製 本

line head

line end

Fig. 3.101: Single line alignment processing.

to be avoided

heading that causes page break

第五章 原稿整理

原稿が著者の手から出版社に渡ると、編集部ではまず、原稿の内容を吟味する。そのうえで、最終的に出版することが決まると、原稿の内容を細かく点検し、体系を整え、さらに原稿の用字用語や体裁などを整理し、統一する仕事を行う。

原稿内容の吟味では、入手した原稿が出版に通しているかどうかを検討する。テーマや範囲、表現のスタイル、内容のレベル、構成、原稿の分量は問題ないだろうか、そして全巻毀損や著作権といった法規上の問題が含まれていないかを吟味し、刊行するかどうかを検討・決定する。

この後、「原稿整理」または「原稿編集」(copy-editing)とよばれる編集作業が行われる。内容や形式について、細かく点検し整理していく仕事である。

この原稿整理においては、本の内容により読者に読みやすくする工夫を講ずることも必要になる。一般に本で伝える情報を豊富にし、読みやすくするためには、情報を論理的な構造をもったものとして分けて示すこと大切である。本の内容は、通常、入れ子になっており、適切なレベルの显出して区分し、適切な言葉で内容を示す。逆に、注にするのも一つの方法である。

本の論理的な構造は「目次」として表現されるので、「目次」で内容を確認することは、とても

trim size

Fig. 3.103: Example of just one character on a page just before a page break (to be avoided).

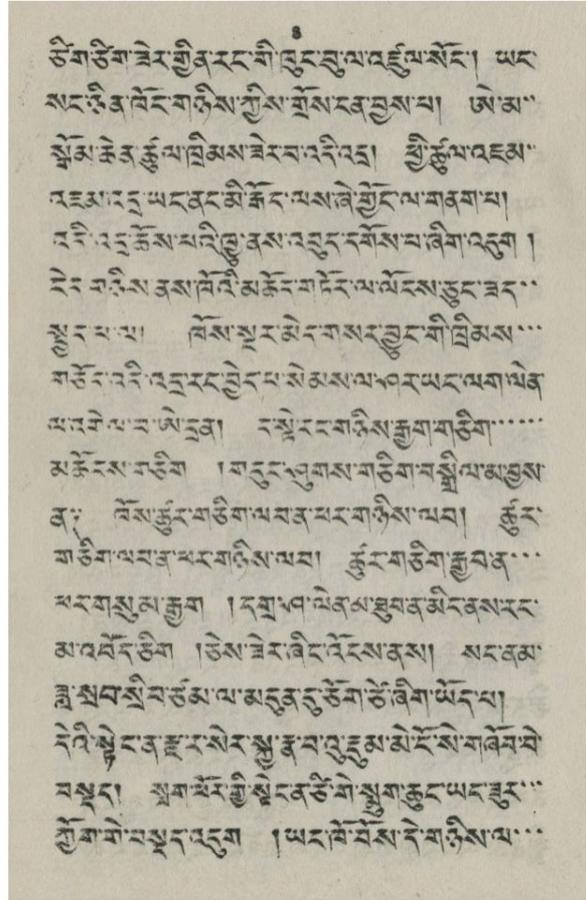


Figure 21: A page of a booklet showing tsek padding.

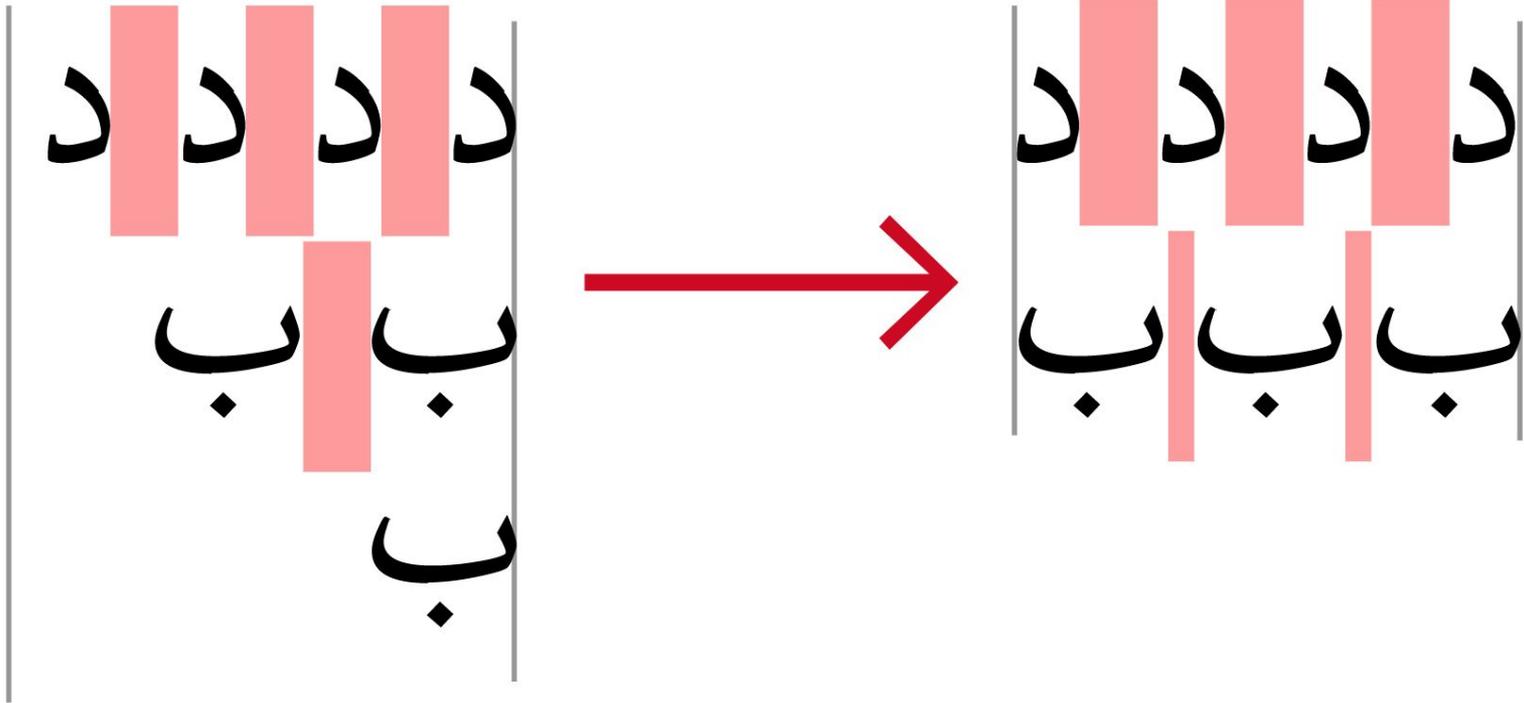


Figure 29: Aligning lines by increasing and decreasing spaces between the words.

Arabic



Figure 30: Altering intra-word spaces between unjoined letters.

ك
ك
ك

default final form Kaf
horizontal scaling Kaf
real stretching Kaf

ك	ك
ك	ك
ك	ك
ك	ك
ك	ك
ك	ك
ك	ك

ا	ا
س	س
س	س
س	س
س	س
س	س
س	س
س	س

Figure 31: Alternative shapes for changing the width of the text.

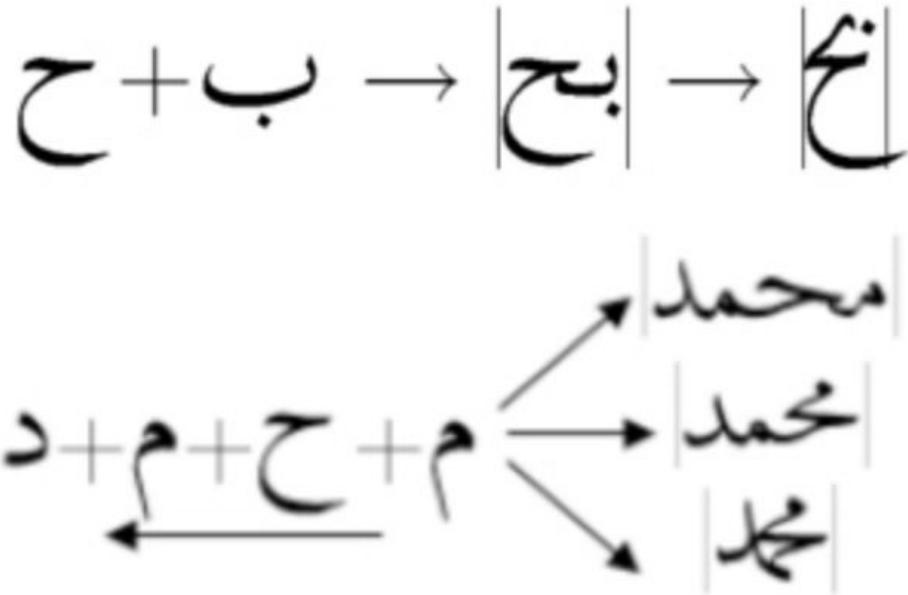


Figure 32: Various ligatures reducing the widths of the words

مراكز ← مراكز
أرسلان ← أرسلان

Figure 33: Two words extended with kashida.

كان يدرس مع صديقي من حميمي، وتعاها ثلاثتهم علم على أن
يساعد من يؤاتيه الحظ الأخرين، وهما ما كان، فلم
أصبح صديقه نظام الملك وزيراً للسلطان ألب أرسلان
ثم لحفيده ملكشاه، حُصَّصَ له مائتين وألف مثقال
يتقاضاه من بيت المال كل عام، من خزينة نيسابور
فضمن له العيش في رفاهة مما ساعدته على
التفرغ للبحث والدراسة. وقد عاش معظم حياته في
نيسابور وسمرقند. وكان يتنقل بين مراكز العلم الكبرى مثل
بخارى وبلخ وأصفهان رغبة من في التزود من
العلم وتبادل الأفكار مع العلماء. وهكذا صار لعمرو
بن الخيام الوقت الكافي للتفكير بأمر وأسرار الحياة،
بعد أن توقفت له أسباب المعيشة، وكان صديقه المثلث
هو الشاعر حسن الصبر صاحب مؤسس طائفة الحشاشين، وهو

Figure 34: Unpleasant result of excessive use of kashida.

የኒል፡ ወንዝ፡ ስላለ፡ ዝናም፡ የማይዘንምበት፡ የም
ስር፡ መሬት፡ ብዙ፡ ዓይነት፡ ፍሬ፡ የሚያፈራ፡ ፍሬያም፡
መሬት፡ ይሆናል። በሐምሌ፡ ያ፡ የወንዙ፡ ውሀ፡ ሊነሳ፡
ይጀምራል፡ ከፍ፡ እያለ፡ ሂደም፡ በመስከረም፡ አገሩን፡
በሞላው፡ ሐይቅ፡ ያደርገዋል። ውሀው፡ ከሚቻለው፡ ስፍራ፡

Figure 26: Ethiopic justification in word bound style (Erikson, 1921 (1913 EC))

እኔም ፡ መለስሁና ፤ «ቢሆን፡ ቢሆን ፡ በሰብአዊ፡
 ርገራጌ ፡ ለፈጸማችሁት ፡ በጎ ፡ ተግባር ፡ ዋጋ ፡ መጠ
 የቅ ፡ ባልተገባችሁ ፡ ነበር ። ዋጋ ፡ የምትጠይቁስ ፡
 ከሆነ ፡ ያሳደጋችሁበትን ፡ ዋጋ ፡ እየሠራሁ ፡ እንድ
 ከፍል ፡ ጠይቁኝ ፡ እንጂ ፤ ለማሳደጊያ ፡ ዋጋ ፡ የመን
 ፈሴን ፡ ነፃነት ፡ እንድከፍላችሁ ፡ ትፈልጋላችሁ ፡ ?
 በፍጹም ፡ መጥፎ ፡ ጥያቄ ፡ ነው ። በወዳጅነትም ፡
 ሆነ ፡ በውለታ ፡ አስጠግቶ ፡ የሰውን ፡ መንፈስ ፡ ለመ
 ያዝ ፡ መሞከር ፡ ታላቅ ፡ ስሕተት ፡ ነው ።» አልሁት ።

Figure 27: Ethiopic justification in centered style (Gubenyā, 1973 (1966 EC)).

قالا شهدنا الاحمر وسوا الابل
شبهه بها الابل من قدر حشاها من السبر بسطه ابن السمران بها الابل
كان يدهن القاع القوق ايدي خوار بها خوار القوق

فصبه ياصرا يدي الابل يبا صرا يدي ليو اوى

ووال
ابوعبد محمد السبعي عليه الصلوات تكبر والابل
والصنور ومن اسقى مكبارة فهو اجرا لثوبين
ان سارة ما وردة معها كما غنا من مير والحد ثنا ه
هنسيه فالاحمرنا معده عن ابره عن ابي هريره عن
السبعي عليه السلام قوله مكبارة يعني لاقية او البقرة
او انا لثابة الة ود كوي القين في كوعها سعي جني
فيه وجمع ابا ما فلون الحلبه واكلا القصر يجمعين
الما وجمعهم فاعل منه كويث الماء وكويثه

قال الاغلب
معناه انه راى غلاما قوما احمره والسبعي من غلامه
رأى غلاما مائة كويثه وفكرته ماء القمار يجمعون ستره

وهالهداما كبره مفضوه والعبد من الابرص

باربمراء كبره وودته سيمله خاب كبره
من

ويف
منه سميت الصرارة كانهما مائة اجتمع و كان
يكنى الناس يرا اولها في الصرارة من كتاب الابل
وليسهما من الة عتي لو كان من الة الابل
مكرودة وما حارة ان يقال لة لة البعر والصرارة
الصرارة لا يكون الا لابل وحدث اخذ انه
نهي عن بيع اللبنة و قال انما حياه فالقيلة هي
الاصحارة يجمعها والاحمر من الابل عن سليمان التيمي

Arabic

Lists & Counter Styles

Region	Writing System	Status
East Asia (CJK)	Japanese	— (tracking)
	Chinese	—
	Hangul	—
South Asia	Indic	—
	Tibetan	Tibetan numerals or European numerals
W Asia & N Africa	Arabic	TBD
NE Africa	Ethiopic (Ge'ez)	In Ethiopic ordered lists a number of symbols are used for the counter suffix. For example: "፲" ; "፳" ; "፵" ; "፶" and even "፷" (Ethiopic Wordspace). Ethiopic corpus will present lists with two styles of alignment. These are a left side alignment at the list counter, or alignment along the counter suffix.
Europe	Latin	TBD



- 生芋丸こんにやくの製造工程**
1. 仕込む前に生芋を1/4ぐらいの大きさにカット。
 2. すり潰した生芋
 3. 昔ながらのバケット式の定量練り、通称バタ練り機
 4. バタ練り機にすり潰した生芋を入れた後、国産の唐辛子を入れます。ミキシングで粘度調整を行うことで、味染みが良く、食感が通います。
 5. 成型するホイール機に入れて圧縮し、熱湯に入れます。



生芋の中は真っ白。皮のまますりつぶすのでこんにやく独特の色がでます。

Tibetan numerals can be used for **list** counters. The Tibetan numbers are used in a simple decimal notation, ie. in the same way as European numerals. They differ only in shape.

- ༡ འ་ཞ་མི་རིགས་ཀྱིས་བསྐྱུན་པའི་ཤིང་གི་ཟམ་པ།
- ༢ ལོ་ངོ་800ཡི་ལོ་རྒྱུས་ལྡན་པའི་དགོན་རྫིང་ཚོས་པོ་དགོ།
- ༣ ཆི་ཅ་ཞེས་པའི་བྱིམ་རྒྱུད་ཀྱི་བང་སེའི་ཚོགས།

Figure 23: Examples of Tibetan counters in a **list**.

European numerals can be used for **list** counters. The European numeral is followed by a period.

1. འ་ཞ་མི་རིགས་ཀྱིས་བསྐྱུན་པའི་ཤིང་གི་ཟམ་པ།
2. ལོ་ངོ་800ཡི་ལོ་རྒྱུས་ལྡན་པའི་དགོན་རྫིང་ཚོས་པོ་དགོ།
3. ཆི་ཅ་ཞེས་པའི་བྱིམ་རྒྱུད་ཀྱི་བང་སེའི་ཚོགས།

Figure 24: Examples of European numeral counters in a **list**.

ሀ/ መሥተጋብዕ
 ለ/ አርባዕት
 መ/ አርያም
 ሐ/ መዋሥዕት
 ሠ/ መወደስ
 ረ/ ክሥተት፡ አርያም
 ሰ/ ዋዜማ፡ ምዕራፍ
 ሸ/ ስብሐተ፡ ነግህ
 ቀ/ የጾም፡ ምዕራፍ፡ በሙሉ።

Figure 37: Alphabetical list aligned on suffix.

ሀ/ መሥተጋብዕ
 ለ/ አርባዕት
 ሐ/ አርያም
 መ/ ሠለስት
 ሠ/ መወደስ
 ረ/ ክሥተት፡ አርያም
 ሰ/ ዋዜማ፡ ምዕራፍ
 ሸ/ ስብሐተ፡ ነግህ
 ቀ/ የጾም፡ ምዕራፍ፡ በሙሉ።

Figure 36: Alphabetical list aligned justified on counter left.

Layout & Pages

Bidirectional Layout

Region	Writing System	Status
East Asia (CJK)	Japanese	(See Vertical Text & Page Direction & Binding)
	Chinese	(See Vertical Text & Page Direction & Binding)
	Hangul	(See Vertical Text & Page Direction & Binding)
South Asia	Indic	—
	Tibetan	—
W Asia & N Africa	Arabic	Arabic script is written from right to left. Numbers, even Arabic numbers, are written from left to right, as is text in a script that is normally left-to-right.
NE Africa	Ethiopic (Ge'ez)	—
Europe	Latin	—

- Additional Requirements for Bidi in HTML & CSS <https://www.w3.org/TR/html-bidi/>

Vertical Text

Region	Writing System	Status
East Asia (CJK)	Japanese	Vertical Writing Mode and Horizontal Writing Mode
	Chinese	Major Differences Between Horizontal and Vertical Writing Modes
	Hangul	Major Differences between Horizontal Writing and Vertical Writing
South Asia	Indic	In vertical arrangement of characters writing each character on a new line may not be suitable in Indian languages. The vertical arrangements of characters are sometimes used in Indian texts. In order to form correct arrangements, it is preferred to follow tailored grapheme cluster approach.
	Tibetan	ISSUE 4: Is Tibetan written vertically with upright glyphs at all (eg. in table headings, in pictures, etc.)? If so, does it require that all elements composing a syllable be kept together in horizontal fashion, placing just syllables one above the other? Or does each non-subjoined/combining character move to the next line?
W Asia & N Africa	Arabic	<ol style="list-style-type: none">1. Situations where short runs of text run vertically.2. Arabic embedded in vertically orientated text.3. Upright vertical Arabic text.
NE Africa	Ethiopic (Ge'ez)	—
Europe	Latin	—

- Styling vertical Chinese, Japanese, Korean and Mongolian text <https://www.w3.org/International/articles/vertical-text/>

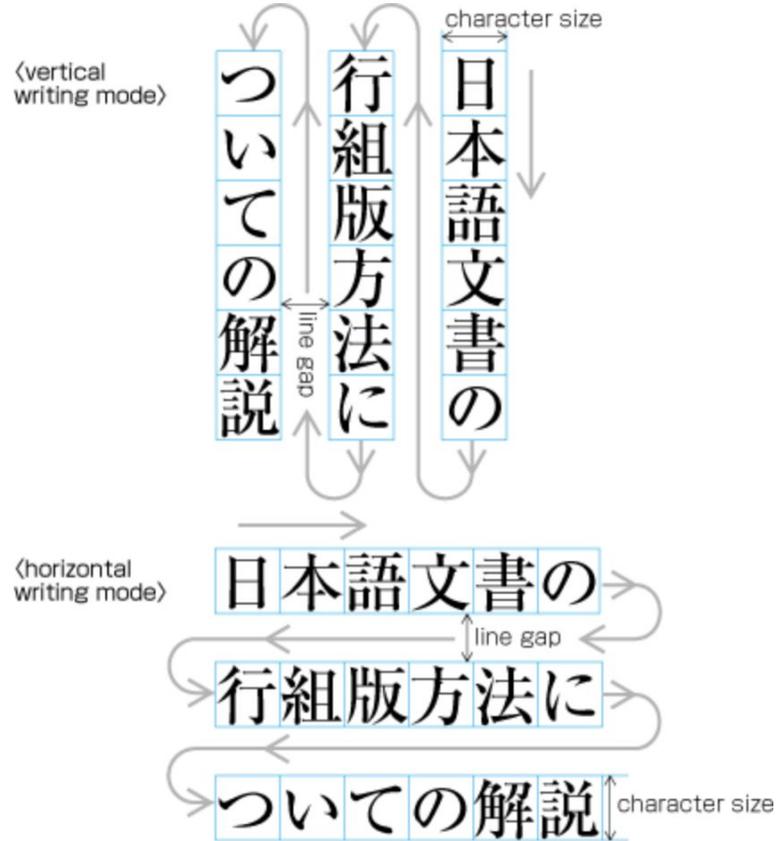


Fig. 2.18: Vertical writing mode and horizontal writing mode. (The arrows show the reading direction.)

3.5.1 Variations in vertical arrangements

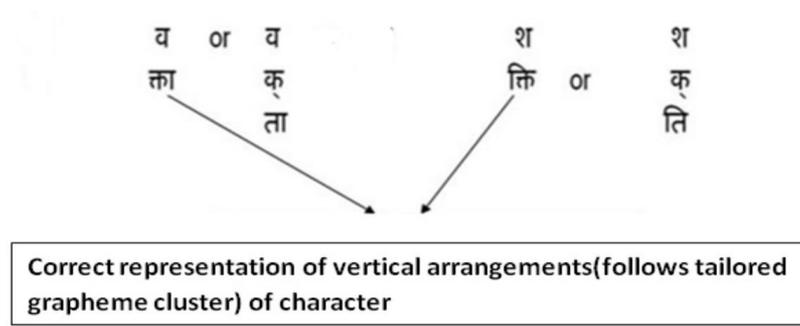


Fig. 7: Variations in vertical arrangements

Given below the example 'स्वागतम्' that follows rule 2 and rule 3 of Indic orthographic syllable definition:

स्वा	CHCv- Rule 2
ग	C - Rule 2
त	C - Rule 2
म्	CH - Rule 3

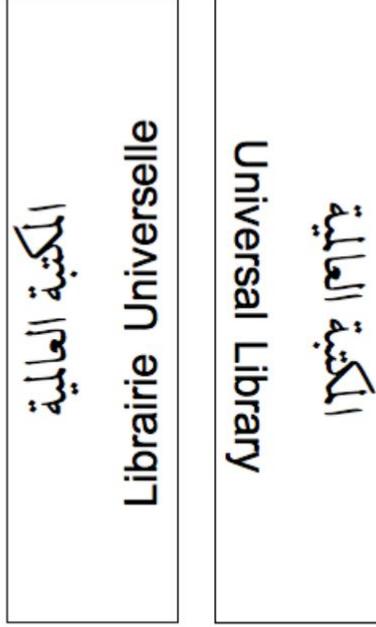


Figure 38: **Vertical** Arabic, top-down and bottom-up flow



Figure 40: Arabic upright on a front movie (see arrow).

ISSUE 26

We need to establish whether this is a standard approach or just an oddity. The items in the list above are important to note, however, we need to check whether isolated forms are always used, and the direction is always top to bottom when upright letters are used.

Notes & Footnotes

Region	Writing System	Status
East Asia (CJK)	Japanese	Native way of side-noting: Warichu: https://github.com/w3c/type-samples/issues/6
	Chinese	Indicating Meaning or Other Additional Information: Bilingual Annotations; Interlinear Comments.
	Hangul	FN in multi-column: three processing modes. Other options when the page is not full.
South Asia	Indic	—
	Tibetan	TBD
W Asia & N Africa	Arabic	TBD
NE Africa	Ethiopic (Ge'ez)	FN counters are simple superscripted cardinal numbers. (No mention of FN content layout.)
Europe	Latin	Footnotes themselves can be quite complicated. Footnotes can contain multiple paragraphs, block quotes, poems, lists, and tables. Footnotes can contain other footnotes. Footnotes can extend across multiple pages. In short, a footnote is a container that can hold almost anything.

漢字廃止論には、前島密（一八三五―一九一九年。明治時代の政治家・実業家、

郵便制度の創始者）などによるものがある。しかし、漢

字の廃止は、困難であり、そこで主張されたのが漢字節減論である。その代表的なものに

福沢諭吉（一八三四―一九〇一年。明治時代のの『文字

啓蒙思想家、慶応義塾大学を創設之教』などがある。

巧合的是，^{Keith}基斯·^{Emerson}愛默生彈奏的賦格
部分這時剛好結束，曲子正要進入
^無盡^的迷^迷
〈Endless Enigma〉的第二樂章。

巧合的是，^{Keith}基斯·^{Emerson}愛默生彈奏的賦格
部分這時剛好結束，曲子正要進入
〈^無盡^的迷^迷 Endless Enigma〉的第一樂章。

1. Showing footnotes at the bottom of the column with the corresponding text.

<p>동해물과 백두산이 마르고 달도둑, 하느님이 보우하사 우리 나라 만세 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세 남산 위에 저 소나무 철갑¹⁾을 두른 듯, 바람 서리 불변함은 우리 기상일세 동해물과 백두산이 마르고 달도둑, 하느님이 보우하사 우리 나라 만세 무궁화 삼천리 화려강산</p>	<p>대한사람 대한으로 길이 보전하세 남산 위에 저 소나무 철갑을 두른 듯, 바람 서리 불변함은 우리 기상일세 무궁화 삼천리²⁾ 화려강산 대한사람 대한으로 길이 보전하세 대한사람 대한으로 길이 보전하세 가을하늘 공활한데 높고 구름없이, 밝은 달은 우리 가슴 일편단심일세 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세</p>	<p>이 기상³⁾과 이 맘으로 충성을 다하여 괴로우나 즐거우나 나라 사랑하세 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세 가을하늘 공활한데 높고 구름없이, 밝은 달은 우리 가슴 일편단심일세 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세 이 기상과 이 맘으로 충성을 다하여 괴로우나 즐거우나 나라 사랑하세</p>
<p>1) 쇠로 둘러싸인 것. 철갑을 입힌 성문 2. 쇠붙이를 길에 붙여 지은 갑옷. 3. [북한어] 총탄을 막기 위하여 전투 기재를 둘러싼 강철판.</p>	<p>2) [명사] 함경북도의 북쪽 끝에서 제주도의 남쪽 끝까지 삼천 리 정도 된다고 하여, 우리나라 전체를 비유적으로 이르는 말.</p>	<p>3) [명사] 사람이 타고난 기개나 마음씨. 또는 그것이 걸로 드러난 모양.</p>

Fig. 42: Bottom of the corresponding column

2. Showing footnotes in a single-column, aligned to the page width.

동해물과 백두산이 마르고 달도둑, 하느님이 보우하사 우리 나라 만세 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세	남산 위에 저 소나무 칠갑을 두른 듯, 바람 서리 불변함은 우리 기상일세 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세	대한사람 대한으로 길이 보전하세 가을하늘 공활한데 높고 구름없이, 밝은 달은 우리 가슴 일편단심일세 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세
남산 위에 저 소나무 칠갑을 두른 듯, 바람 서리 불변함은 우리 기상일세 동해물과 백두산이 마르고 달도둑, 하느님이 보우하사 우리 나라 만세 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세	남산 위에 저 소나무 가을하늘 공활한데 높고 구름없이, 밝은 달은 우리 가슴 일편단심일세 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세 이 기상과 이 맘으로 충성을 다하여 괴로우나 즐거우나 나라 사랑하세 무궁화 삼천리 화려강산	대한사람 대한으로 길이 보전하세 이 기상과 이 맘으로 충성을 다하여 괴로우나 즐거우나 나라 사랑하세

1)1 .쇠로 둘러싸운 것.칠갑을 입힌 성문

2 .쇠붙이를 걸에 붙여 지은 갑옷.

3 .[북한어] 총탄을 막기 위하여 전투 기재를 둘러싼 강철판.

2) [명사] 함경북도의 북쪽 끝에서 제주도의 남쪽 끝까지 삼천 리 정도 된다고 하여, 우리나라 전체를 비유적으로 이르는 말.

3) [명사] 사람이 타고난 기개나 마음씨. 또는 그것이 겉으로 드러난 모양.

Fig. 43: Page width single-column.

3. Showing all the footnotes in a column on the right side.

동해물과 백두산이 마르고 달도둑, 하느님이 보우하사 우리 나라 만세 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세	남산 위에 저 소나무 철갑을 두른 듯, 바람 서리 불변함은 우리 기상일세 동해물과 백두산이 마르고 달도둑, 하느님이 보우하사 우리 나라 만세 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세	남산 위에 저 소나무 철갑을 두른 듯, 바람 서리 불변함은 우리 기상일세 동해물과 백두산이 마르고 달도둑, 하느님이 보우하사 우리 나라 만세 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세	대한사람 대한으로 길이 보전하세 가을하늘 공활한데 높고 구름없이, 밝은 달은 우리 가슴 일편단심일세 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세 이 기상과 이 맘으로 충성을 다하여 괴로우나 즐거우나 나라 사랑하세	1) .쇠로 둘러싸운 것. 철갑을 입힌 성문 2 .쇠붙이를 겹쳐 붙여 지은 갑옷. 3 .[북한어] 총탄을 막기 위하여 전투 기재를 둘러싼 강철판. 2) [명사] 함경북도의 북쪽 끝에서 계주도의 남쪽 끝까지 삼천 리 정도 된다고 하여, 우리나라 전체를 비유적으로 이르는 말. 3) [명사] 사람이 타고난 기개나 마음씨. 또는 그것이 겉으로 드러난 모양.
---	--	--	--	--

Fig. 44: Right side column

1. Placing the footnote right above the specified footer area.

If the footnote is increased, it goes toward the main text and the main text area is decreased.

㉔ 일반문단의 경우

남산 위에 저 소나무 철갑¹⁾을 두른 듯, 바람 서리 불변함은 우리 기상일세
 무궁화 삼천리²⁾ 화려강산 대한사람 대한으로 길이 보전하세
 가을하늘 공활한데 높고 구름없이, 밝은 달은 우리 가슴 일편단심일세
 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세
 이 기상³⁾과 이 맘으로 충성을 다하여 괴로우나 즐거우나 나라 사랑하세

- 1)1. 쇠로 둘러싸운 것. 철갑을 입힌 성문
2. 쇠붙이를 곁에 붙여 지은 갑옷.
3. [북한어] 총탄을 막기 위하여 전투 기계를 둘러싼 강철판.
- 2) [영사] 함경북도의 북쪽 끝에서 제주도의 남쪽 끝까지 삼천 리 정도 된다고 하여, 우리나라 전체를 비유적으로 이르는 말.
- 3) [영사] 사람이 타고난 기개나 마음씨. 또는 그것이 겉으로 드러난 모양.

㉕ 다단 각주의 경우

남산 위에 저 소나무 철갑¹⁾을 두른 듯, 바람 서리 불변함은 우리 기상일세
 무궁화 삼천리²⁾ 화려강산 대한사람 대한으로 길이 보전하세
 가을하늘 공활한데 높고 구름없이, 밝은 달은 우리 가슴 일편단심일세
 무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세
 이 기상³⁾과 이 맘으로 충성을 다하여 괴로우나 즐거우나 나라 사랑하세

- 1)1. 쇠로 둘러싸운 것. 철갑을 입힌 성문
2. 쇠붙이를 곁에 붙여 지은 갑옷.
3. [북한어] 총탄을 막기 위하여 전투 기계를 둘러싼 강철판.
- 2) [영사] 함경북도의 북쪽 끝에서 제주도의 남쪽 끝까지 삼천 리 정도 된다고 하여, 우리나라 전체를 비유적으로 이르는 말.

Fig. 45: Decreased main text area when the footnote content is changed. (Single column and multi-column cases.)

2. Placing the footnote right below the main text.

If the main text content is increased, the footnote content goes downward.

㉔ 일반문단의 경우

남산 위에 저 소나무 철갑¹⁾을 두른 듯, 바람 서리 불변함은 우리 기상일세
무궁화 삼천리²⁾ 화려강산 대한사람 대한으로 길이 보전하세
가을하늘 공활한데 높고 구름없이, 밝은 달은 우리 가슴 일편단심일세
무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세
이 기상³⁾과 이 맘으로 충성을 다하여 괴로우나 즐거우나 나라 사랑하세



-
- 1) 쇠로 둘러싸운 것. 철갑을 입힌 성문
 - 2) 쇠붙이를 곁에 붙여 지은 갑옷.
 3. [북한어] 총탄을 막기 위하여 전투 기계를 둘러싼 강철판.
- 2) [영사] 함경북도의 북쪽 끝에서 제주도의 남쪽 끝까지 삼천 리 정도 된다고 하여, 우리나라 전체를 비유적으로 이르는 말.
- 3) [영사] 사람이 타고난 기개나 마음씨. 또는 그것이 겉으로 드러난 모양.

㉕ 다단 각주의 경우

남산 위에 저 소나무 철갑¹⁾을 두른 듯, 바람 서리 불변함은 우리 기상일세
무궁화 삼천리²⁾ 화려강산 대한사람 대한으로 길이 보전하세
가을하늘 공활한데 높고 구름없이, 밝은 달은 우리 가슴 일편단심일세
무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세
이 기상³⁾과 이 맘으로 충성을 다하여 괴로우나 즐거우나 나라 사랑하세

불변함은 우리 기상일세
무궁화 삼천리²⁾ 화려강산
대한사람 대한으로 길이
보전하세
가을하늘 공활한데 높고



-
- 1) 쇠로 둘러싸운 것. 철갑을 입힌 성문
 2. 쇠붙이를 곁에 붙여 지은 갑옷.
 3. [북한어] 총탄을 막기 위하여 전투 기계를 둘러싼 강철판.
- 2) [영사] 함경북도의 북쪽 끝에서 제주도의 남쪽 끝까지 삼천 리 정도 된다고 하여, 우리나라 전체를 비유적으로 이르는 말.

Fig. 46: Moved footnote content when main text content is changed. (Single column and multi-column cases.)

Page Numbering & Running Headers

Region	Writing System	Status
East Asia (CJK)	Japanese	PN & RH for vertically and horizontally set books with double running heads.
	Chinese	—
	Hangul	PN can be presented in many forms, and can also be located on any side of the corresponding page.
South Asia	Indic	—
	Tibetan	TBD
W Asia & N Africa	Arabic	TBD
NE Africa	Ethiopic (Ge'ez)	The layout and formatting of page and section numbering in Ethiopic practices does not demonstrate a marked difference from Western conventions.
Europe	Latin	Running heads and footers may contain: [...] Counters of all sorts: page numbers, section numbers, chapter numbers. [...]

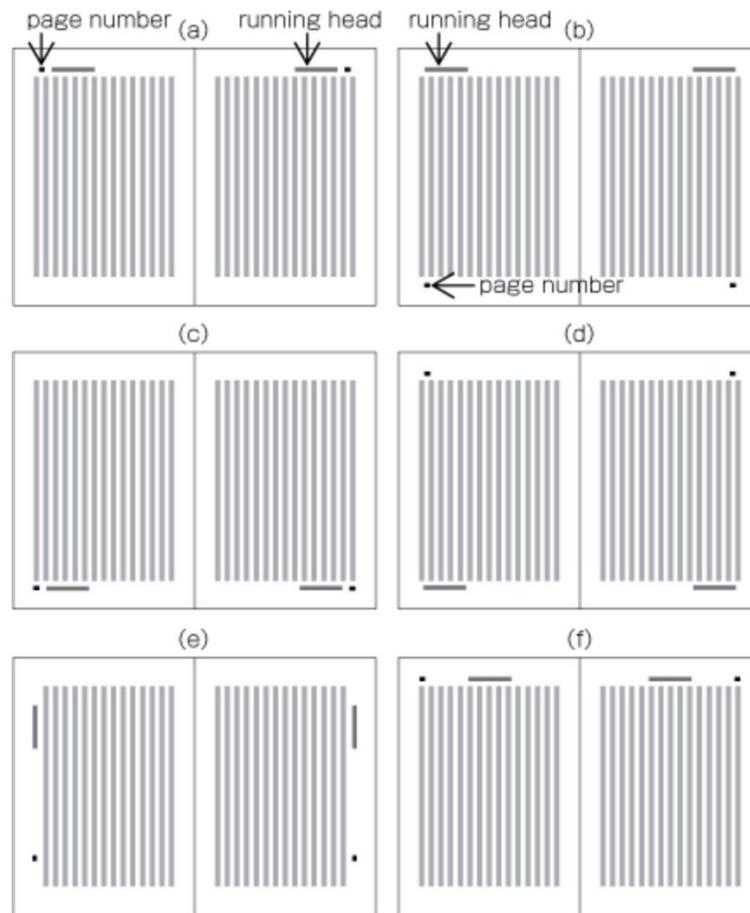


Fig. 2.48: Typical positioning of running heads and page numbers for vertically set books with double running heads.

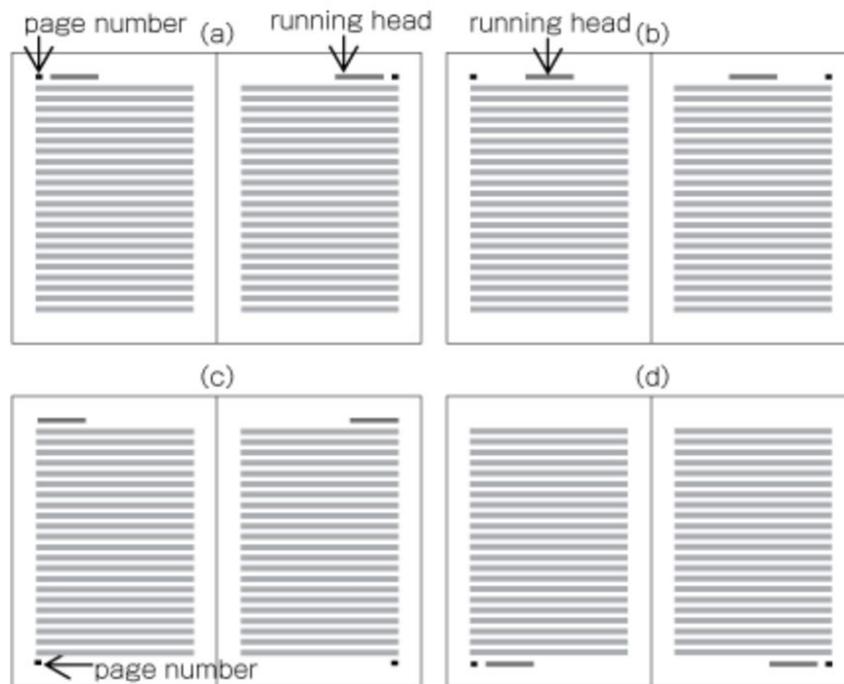


Fig. 2.49: Typical positioning of running heads and page numbers for horizontally set books with double running heads.

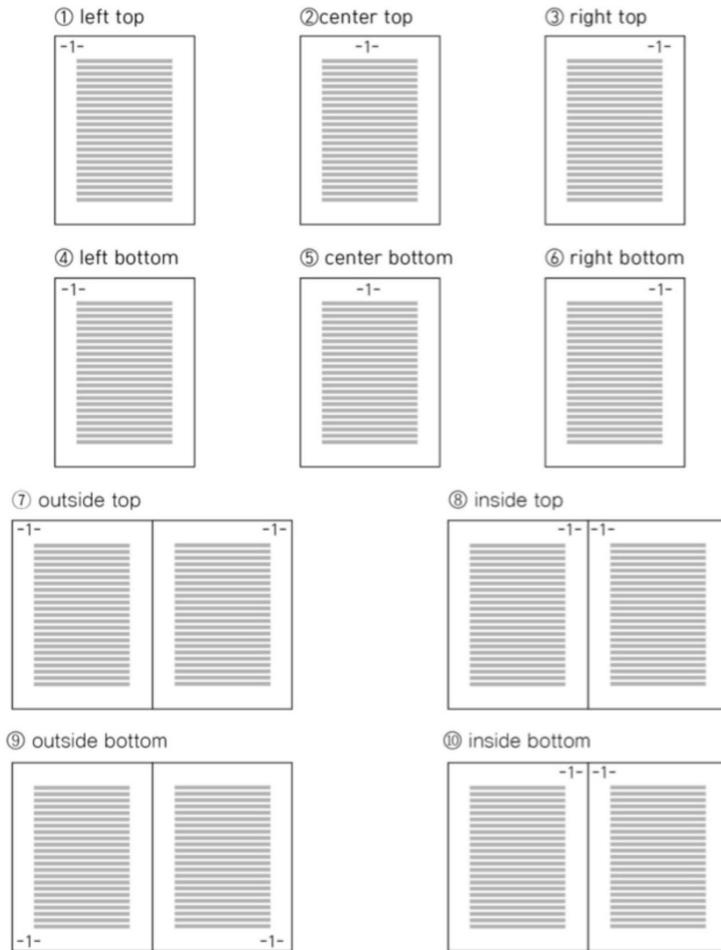


Fig. 49: Positions for page numbers

39

የግዴጩው ፡ ቀራሽ ፡ ።

፴፱

ጎዘን ፡ እንጂ ፡ ደስታ ፡ አለመሆኑን ፡ አሰብ ። ዕድለኞቹ ፡ ጀግ
ኖች ፡ ወታደሮች ፡ ጠላቶቻቸውን ፡ አስቀድመው ፡ እየተንተ

Figure 42: Page numbering in two systems from *Metsehafe Chewata Sigawi* by Zeneb Ethiopia.

Position and Size of Elements non-Paragraphs

Region	Writing System	Status
East Asia (CJK)	Japanese	Extensive details on Positioning of Illustrations & Processing of Tables.
	Chinese	—
	Hangul	Elements beside paragraphs (objects) are processed the same as characters, or as objects.
South Asia	Indic	—
	Tibetan	TBD
W Asia & N Africa	Arabic	TBD
NE Africa	Ethiopic (Ge'ez)	—
Europe	Latin	TK (!)

10. Images

10.1 TK

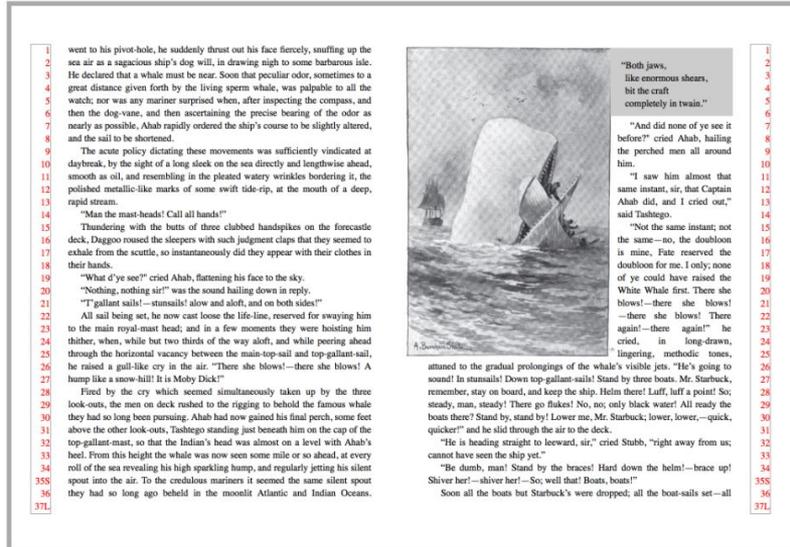


Fig. 17 Figure with caption and runaround

Some things to note about this image

1. the caption and image are treated as a unit
2. Text runs around the image+caption
3. image runs right up to the gutter of the page (i.e. extends beyond usual content area)

Japanese

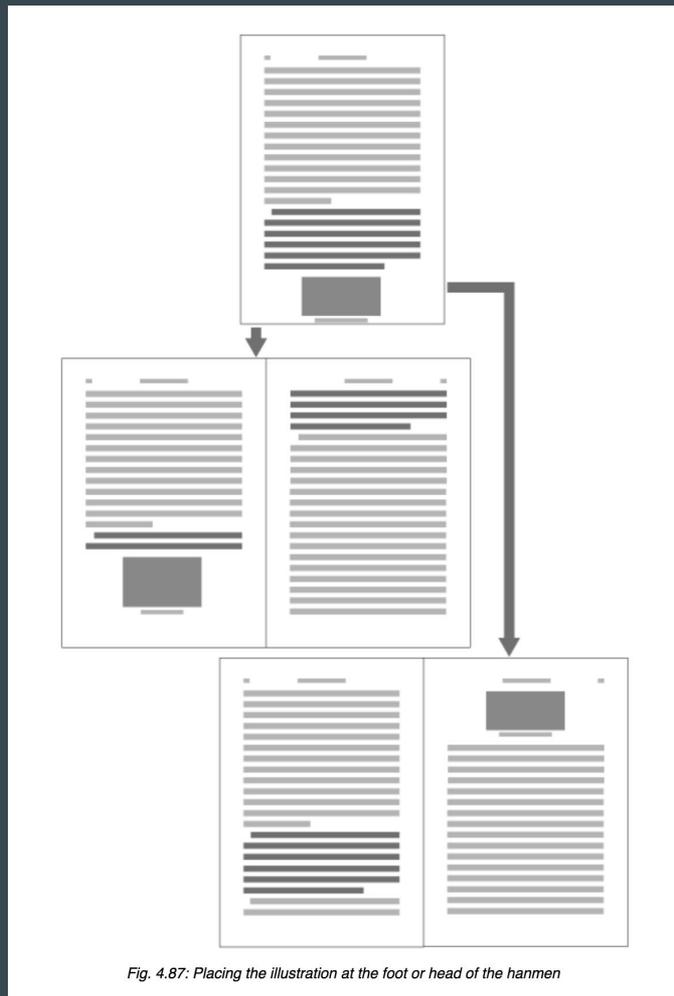


Fig. 4.87: Placing the illustration at the foot or head of the hanmen

동해물과 백두산이 마르고 닳도록, 하느
님이 보우하사 우리 나라 만세 무궁화 삼
천리 화려강산 [redacted] 대한사람 대
한으로 길이 보전하세 남산 위에 저 소나

Fig. 50: Processing as a character.

동해물과 백두산이 마르고 닳도록, 하느
님이 보우하사 [redacted] 무궁화 삼
천리 화려강산 [redacted] 대한으로 길이
보전하세 남산 위에 저 소나무 철갑을 두

Fig. 51: Processing as an object.

1. Text Wrap mode

Both body and objects are located on the same line, but they do not invade each other's space. The objects are reflowed to match the body.

동해물과 백두산이 마르고 닳도록, 하느
님이 보우하사 우리 나라 만세 무궁화 삼
천리 화려강산 대한사람 대한으로 길이
보전하세 남산 위에 저 소나무 철갑을 두
른 듯, 바람 서리 불변함은 우리 기상일
세 가을하늘 공활한데 높고 구름없이

Fig. 52: Text wrap mode.

2. Top and Bottom mode

Line spacing of the line with the object is determined by the height of the object and the body does not invade the object areas.

동해물과 백두산이 마르고 닳도록, 하느
님이 보우하사 우리 나라 만세 무궁화 삼



천리 화려강산 대한사람 대한으로 길이
보전하세 남산 위에 저 소나무 철갑을

Fig. 53: Top and bottom mode.

Page Direction & Binding

Region	Writing System	Status
East Asia (CJK)	Japanese	Major Differences between Vertical Writing Mode and Horizontal Writing Mode: Vertical writing mode: A book starts with the left (recto) side and progresses from right to left. Horizontal composition: A book starts with the right (recto) side and progresses from left to right.
	Chinese	Major Differences Between Horizontal and Vertical Writing Modes: Vertical composition: [...] A book starts with the left (recto) side and progresses from right to left. Horizontal composition: [...] A book starts with the right (recto) side and progresses from left to right.
	Hangul	Major Differences between Horizontal Writing and Vertical Writing: In vertical writing: [...] pages progress from right to left, and pages are turned from left to right. In horizontal writing: [...] pages progress from left to right, and pages are turned from right to left. Basic Elements of Designing Page Body: [...] Writing direction (vertical/horizontal).
South Asia	Indic	—
	Tibetan	TBD
W Asia & N Africa	Arabic	TBD
NE Africa	Ethiopic (Ge'ez)	—
Europe	Latin	The nature of printing and binding mandate that the number of pages in a book be some multiple of eight, sixteen, or thirty-two pages. Recto vs Verso.

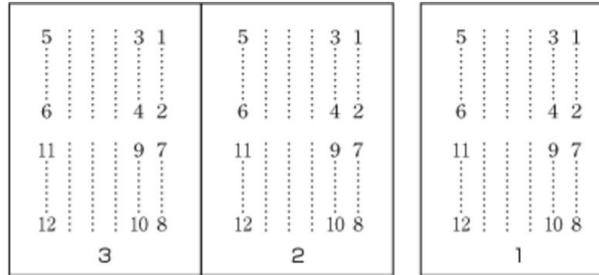


Fig. 2.20: Direction of arrangement of characters in vertical writing mode.

to bottom, lines are arranged from right to left.

bottom. A book starts with the left (recto) side and progresses from right to left (see [Fig. 2.21](#)).

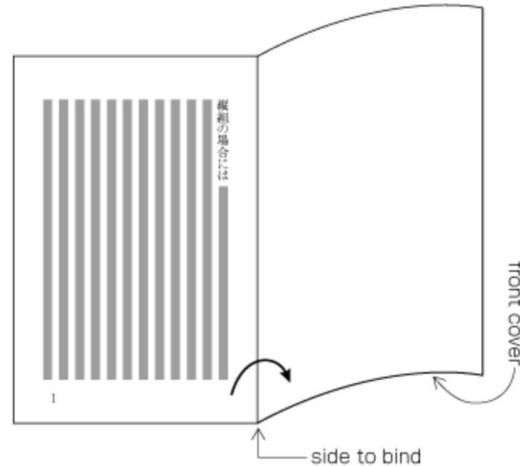


Fig. 2.21: Progression of pages for a vertically set books.

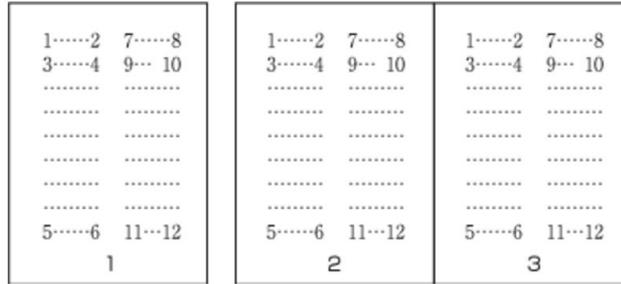


Fig. 2.22: Direction of arrangement of characters in horizontal writing mode.

right, and lines are arranged from top to bottom.

ht. A book starts with the right (recto) side and progresses from left to right (see [Fig. 2.23](#)).

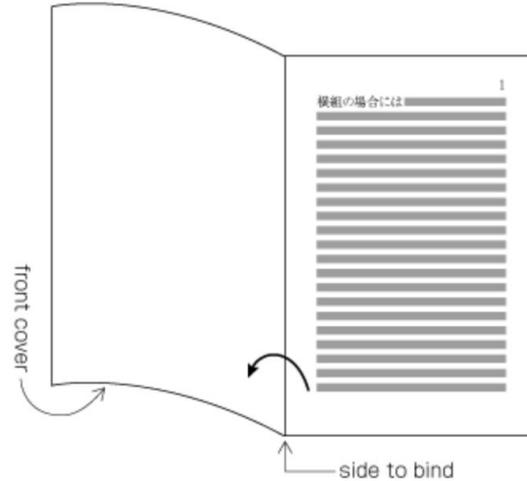


Fig. 2.23: Progression of pages for a horizontally set book.

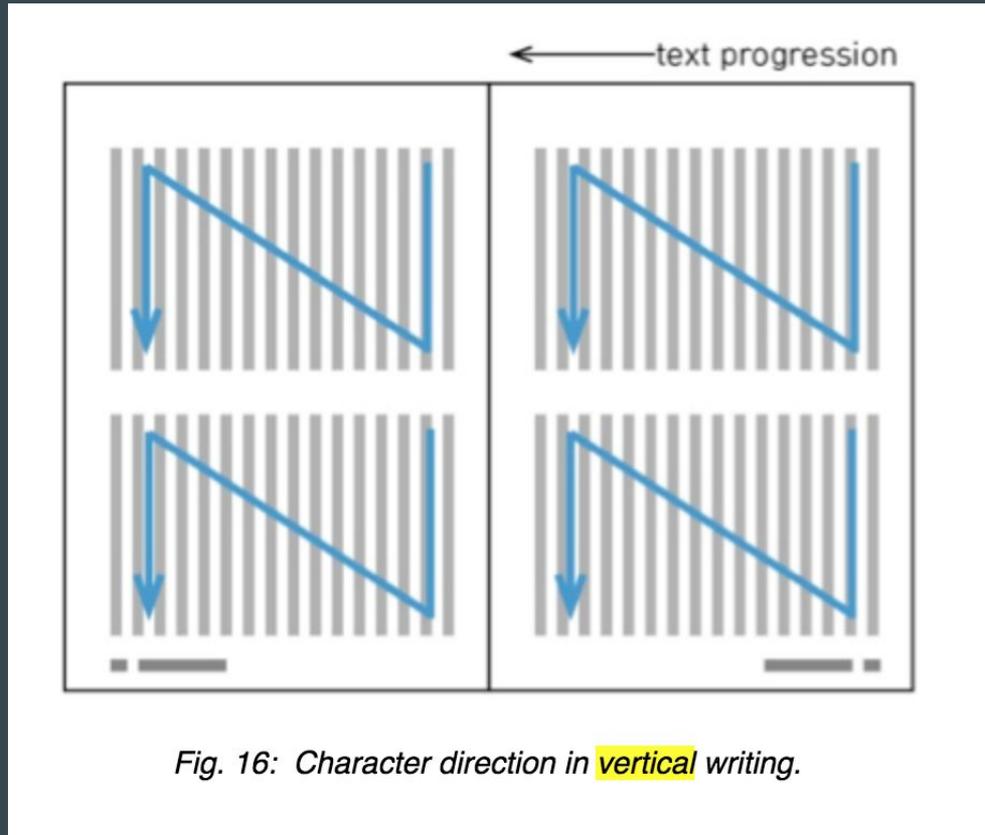


Fig. 16: Character direction in **vertical** writing.

text progression →

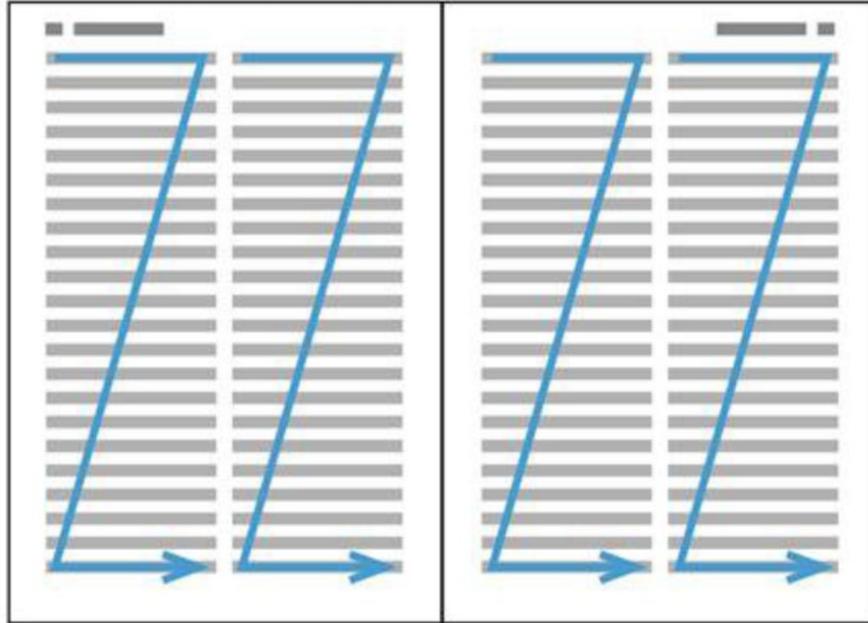


Fig. 17: Character direction in horizontal writing.

Conclusions & Future Work

Conclusions

- How languages/writing systems/cultures measure themselves?
 - TFs start with the features most important to them.
 - Features not different from Latin/Western are usually left out in the documents.
 - It is common to assume anything not mentioned explicitly is not different.
 - But, sometimes “Not Applicable” needs more research before reaching the green stage.
- Comparing requirement documents helps with finding new questions & answers.
 - Specially in the age of web, it's important to highlight what does NOT apply to a language/script/culture.
 - E.g. **bold** and *italic* are not applicable to Chinese characters. Other methods are used for emphasis.

Future Work

- Existing requirement documents
 - Hangul / Korean
 - The Working Draft has outstanding issues that need to be addresses before reaching WG Note.
 - Ethiopic
 - Active, but moving forward slowly. The Task Force can use help of more editors to reach the next stage sooner.
- New requirement documents, tests, and implementations
 - A new approach for the layout work, going forward: ie. to start with a gap analysis, then use that to drive areas of work on the requirements document, tests, followed by communication of requirements to spec/browser developers.
 - <https://w3c.github.io/typography/gap-analysis/language-matrix.html>
 - <https://w3c.github.io/typography/gap-analysis/index?data=ethi>

See also...

- Type Samples repository
 - <https://w3c.github.io/type-samples/>
- Next Steps for Japanese Script Support
 - Richard Ishida, Japan (March 2017)
 - <https://www.w3.org/International/talks/1703-jleq/>
- Strings and bidi
 - <https://www.w3.org/International/articles/strings-and-bidi/>
- Notes on JSON strings and text direction
 - <https://w3c.github.io/i18n-discuss/notes/json-bidi>
- Ruby Markup and Styling
 - <https://www.w3.org/International/articles/ruby/markup>
 - <https://w3c.github.io/i18n-drafts/articles/ruby/styling.en.html>

質問？

שאלה?

سؤال؟

題？

문제？

پرسش؟

Questions?