

Multilingualism and multiscrptism in TEI publishing: DH2022

Yifan Wang (International Institute for Digital Humanities / Graduate school of the University of Tokyo), Kiyonori Nagasaki (International Institute for Digital Humanities), Ikki Ohmukai (The University of Tokyo), Masahiro Shimoda (Graduate school of the University of Tokyo)

Language-based markup

- Report, Vol. 33, pp. 1-11.
- [7] Baochang, G. 耿宝昌 (1993). *Ming Qing jianqing 明清瓷器鉴定* [Ming and Qing Porcelains Inspection]. Beijing: The Palace Museum.
- [8] Jun Z. 朱军 (2002). "Mingmo Qingchu huagu jianqing 明末清初青花花瓶鉴定 [A late early Qing dynasty blue and white goblet identified *Wenwu Shijie* 文物世界 4, pp. 38-42.
- [9] Jingjing, X. 徐菁菁 (2017). "Mingqing yuanyu jiqi tezheng 明清瓷瓶源流及特征 [Source Characteristics of Ming and Qing beaker vessels 艺术品. 11, pp. 66-73.
- [10] Medley, M. (1987). "The Ming-Qing Chinese Porcelain", *Artis. Asiatiques* 42, pp. 65-73.
- [11] Jenyus, S. (1955). "The Wares of the Transitional Period between the Ming and Ch'ing, 1620-16

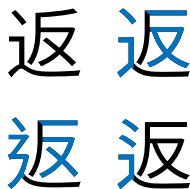
We have published the book of abstracts of DH2022 Tokyo using XSL-FO software suite provided by ADHO. The abstracts contain a large number of names and phrases in multiple Asian languages, which are not supported by the default setup. We thus decided to make a large expansion on its multilingual function.

```
</bibl style="text-align: left;">Jain
<hi style="text-align: left;">Koba
<hi rend="Japanese">文政期前後の
<hi rend="Japanese">浮世絵風景画
<hi rend="italic" style="font-fa
</bibl style="text-align: left;">Lee.
```

Author index is improved to accept tailored name sorting and normalization of spelling variants based on manual annotation in TEI files.

Yuan, Yiguo	605
van Zaanen, Menno	571, 606
Zafar, Huma	342
Zaghrouni, Wajdi	907
Zaidonak, Itay	633, 634
Zarei, Alireza	406
Zarrieß (Zarriess), Sina	473, 527
Zhan, Hanna Yaqing	104
Zhang, Hailin	225

Locale-specific typefaces



(source: https://commons.wikimedia.org/wiki/File:Source_Han_Sans_Version_Difference.svg)

We prepared different font files for CJK languages, namely: Simplified and Traditional Chinese, Japanese, and Korean. Unlike Western European languages, each region has varied standard for same characters, and there is no "neutral" shape that all local standards can agree with.

There are also passages of medieval Latin abbreviation, Sanskrit, and Thai found in the source files, that we needed to find specific fonts for. For the complex layout scripts, some fonts seemed to have minor incompatibilities with the processor (Apache FOP), and we needed to select a working font for our environment.

pecore penitentiaugen
signis extiorib penitetic
delectant ut pote usu
ciliciipfundis suspiriis
oŕoniß cre

of letter elements (right to left, e.g. *k+i*) the stroke order (left to right, e.g.), irregularly typeset dots, which indicate (*anusvāra*), and some lengthy consonant as (*स्त्व*, *rtsm-y-a*). Focusing on these in a were insufficiently handled by the preced

☹ CJK languages are accustomed to use different fonts for each locale, but most Cyrillic fonts implement regional variants through the fonts' OpenType features. That would make localization difficult in XSL-FO processing with Apache FOP, which currently does not support OpenType features.

Mixed-script typesetting

Intertextuality has been a staple of scholarly communities: like *Redaktionsgeschichte*校勘學 in China have laid foundations for debates works, editions and aut

Intertextuality has been a staple of scholarly communities: like *Redaktionsgeschichte*校勘學 in China have laid foundations for debates works, editions and aut

In the tight line setting of the book of abstracts, we needed some adjustment to the vertical position of CJK text, lest it disrupt the consistent line height of surrounding alphabet lines. The balancing can be only done on a case-by-case basis, depending on what fonts are combined.



Edge case: some Thai typefaces are designed smaller to allow the safe margin for stacking diacritics. Thai letters were enlarged lest be awkwardly undersized among ordinary Roman typefaces.

nuwat ทาริน ค

☹ Although our abstracts contain small portions of Hebrew, RTL scripts did not cause us problems as far as we know, mostly because the default Times New Roman font supports Hebrew out of the box. We did not test how markup-based font specification interacts with RTL intermixing.

☹ Ideally, the language of the text should be specified in @Lang in TEI, but we added them via custom @rend values following the existing style processing logic of TEI to PDF Book Creator.

Our modified version of TEI to PDF Book Creator program for DH2022 is available at: <https://github.com/747/tei-to-pdf-dh2022/>