



LibreOffice
The Document Foundation



State of CJK issues of LibreOffice, 2018 edition Shinji ENOKI

shinji.enoki@gmail.co
m

TIRANA | 27 Sept.



2018
LIBOCON
TIRANA 2018



Agenda

- About the CJK features & issues
- CJK issue status
- Next challenge for the CJK issues



Who am I?

- Shinji Enoki (榎真治)
- A member of LibreOffice Japanese Team(2011-)
- A member of The Document Foundation (2014-)
- Activity: organizing events, grows community, sometimes QA
- Other community:
 - Chairperson of Japan UNIX Society
 - A member of Kyoto OpenData Community
 - A staff of KANSAI OPEN SOURCE



Do you know the word CJK bug?





What are CJK issues / bugs

- CJK is an abbreviation for **C**hinese-**J**apanese-**K**orean
- LibreOffice has many language-specific features and issues, CJK issue is one of them.
- CJK issues happen when used in CJK's environment
 - A function of CJK is broken
 - A bug occurs only with the CJK environment



Next Decade Manifesto

For the past ten years, the OpenOffice.org community has developed, supported and promoted the world's leading open-source office productivity suite. We have attracted the support of tens of thousands of individuals and corporate bodies during this period.

We now call on all our supporters to follow us into the next phase of our development, as we become an independent Foundation.

Our Values

WE COMMIT OURSELVES:

To eliminate the digital divide in society by giving everyone access to office productivity tools free of charge to enable them to participate as full citizens in the 21st century

To support the preservation of mother tongues by encouraging people to translate, document, support, and promote our office productivity tools in their mother tongue

To allow users of office productivity software to retain the intellectual property in the documents they create by use of open document formats and open standards

To an open and transparent peer-reviewed software development process where technical excellence is valued

WE REJECT:

The ownership of office productivity tools by monopoly suppliers which imposes a de-facto tax on global electronic free speech and penalises the economically disadvantaged

The creeping domination of computer desktops by a single language, forcing all people to learn a foreign language before they can express themselves electronically

The ownership of file formats by proprietary software companies - documents belong to their creators, not software vendors

A closed software development process where errors can lie hidden and poor quality is accepted



The Document Foundation: Next Decade Manifesto

“To support the preservation of mother tongue by encouraging people to translate, document, support, and promote our office productivity tools in their mother tongue”

- I think these “Our Values” are very good



Why do CJK people need to work CJK bugs?

- Many LibreOffice developers are European people
- It is hard to notice / understand CJK bug unless it is from people of CJK
- Then, It is very important for CJK users to report / reproduce CJK bugs



CJK Meta issues

- CJK bugs are tracked META issues
 - [Bug 83066](#)(CJK) - [META] issues

Bug 83066 (CJK) - [META] CJK (Chinese, Japanese, Korean, and Vietnamese) language issues

Status: NEW

Alias: CJK

Product: LibreOffice

Component: LibreOffice (show other bugs)

Version: unspecified (earliest affected)

Hardware: Other All

Importance: medium normal

Assignee: Not Assigned

URL:

Whiteboard:

Keywords:

Reported: 2014-08-25 14:50 UTC by Matthew Francis

Modified: 2018-08-21 16:27 UTC (History)

CC List: 15 users (show)

See Also: Font-Rendering Vertical-Text RTL-CTL

Crash report or crash signature:

Depends 35208 36709 39935 43671 50349 Ruby 50901 55954 63011 65266 66791 69809 71329 72042 75790 77803 81272 81484 82115 82116 82293 on: 82689 82690 82692 82930 83844 85795 88752 89422 89979 90874 93695 96458 96459 101679 101895 104378 104503 105501 Text-Grid 107195 107383 107466 107467 107484 107485 107487 108042 109030 112477 112745 113048 113184 113189 113191 CJK-Chinese-Traditional CJK-Chinese-Simplified CJK-Japanese CJK-Korean 114520 114746 114760 114761 114763 114764 115130 115321 115970 117857 118260 118476 119211 119219 40091 40335 44784 47719 49073 57139 58604 62750 64975 66126 67814 70097 76255 76291 77076 77490 77514 77971 77972 78789 79766 80272 80346 80473 80513 80604 80788 80812 81144 82176 82818 83060 83145 83178 83252 83559 83752 83755 84138 85954 87626 88169 88527 88855 89409 91288 92655 92782 94008 94066 95379 95656 95836 96091 96197 96457 96829 97405 97702 98165 98269 98527 98879 101729 102041 102074 102253 103528 103710 103718 103785 103850 104625 104726 104801 104828 104954 106097 106295 107056 107435 107462 109260 111717 111967 112066 112593 112807 112879 113481 114486 114736 115353 116577 118564

Blocks: Languages

Show dependency tree / graph



Check [META] CJK issues

	2017/10	2018/9
Depends on:(All)	135	168
Open	59	73
Close	76	95

- Open: Including 6 meta bugs “Text Grid “, “ruby”, “CJK-Japanese” and othe (not meta: 67)
- Some cases are not added to CJK META issues due to CJK related buga



- Mark Hung is fixing some CJK issues
- Thanks Mark!





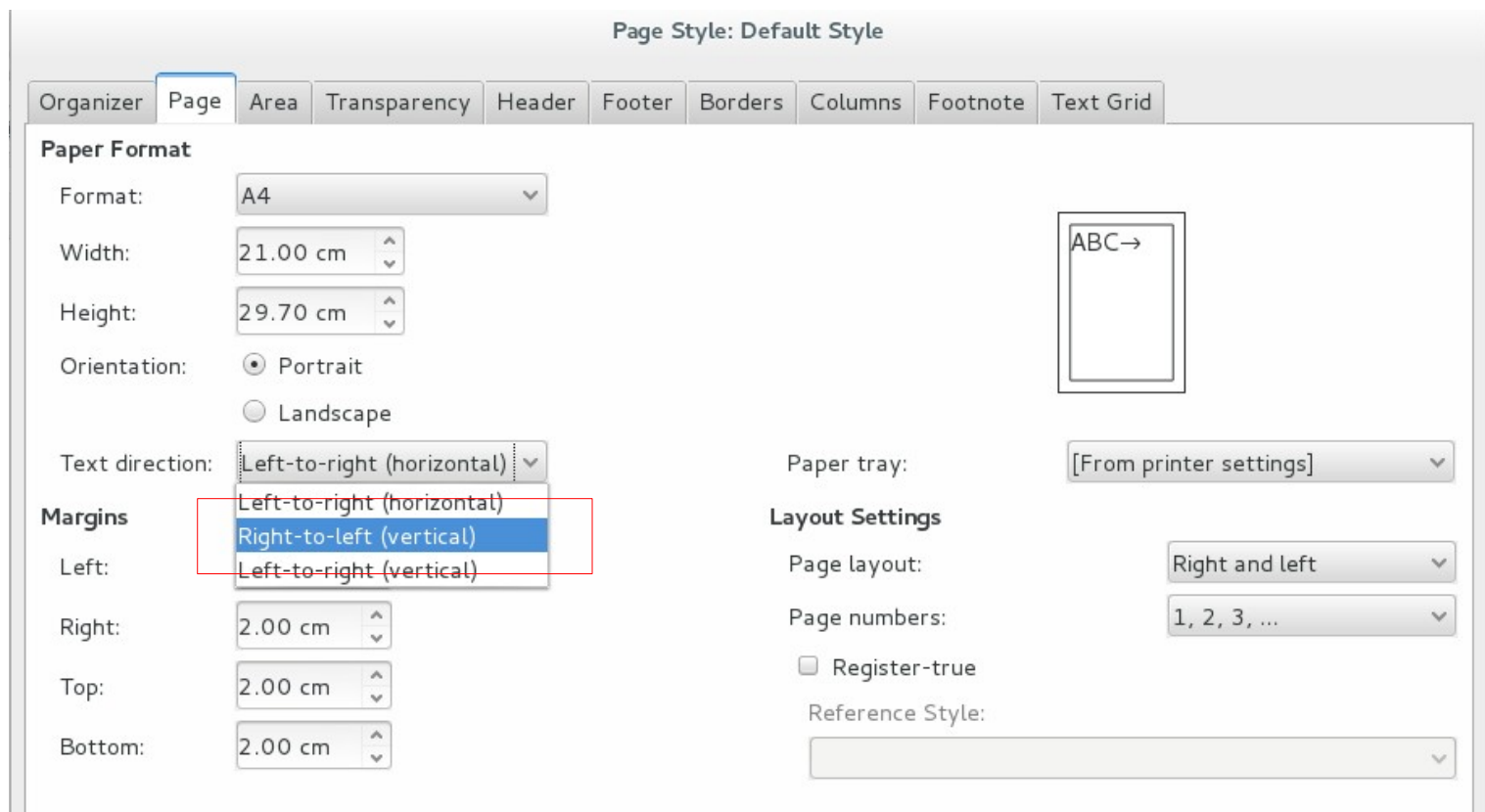
CJK features

- In common with CJK
 1. using Kanji (Chinese characters)
 2. vertical writing
 3. phonetic guides (ruby)
- Chinese characters are also different in Simplified Chinese (Mainland China), Traditional Chinese (Taiwan, Hong Kong) and Kanji (Japanese)
- Even with common functions, usage methods are not always the same
 - Vertical writing



Vertical writing

- Exsample: Witer page stype
 - Menu [Format] – [Page...]
 - [Page] tab



Vertical writing(2)

The screenshot shows a word processor interface with a menu bar (View, Insert, Format, Styles, Table, Form, Tools, Window, Help) and a toolbar. The font is set to IPAmj明朝 and the size is 10.5. The main text area contains vertical Japanese text. The status bar at the bottom indicates 807 words, 814 characters, Default Style, Japanese, and 160% zoom.

俊寛様の話ですか？ 俊寛様の話くらい、
たし、——自身の事さえ、でもない嘘が伝
いをなすってしまうし、わたしはそのを言
あの島の女と、夫婦のらいをなすった上、
した。前の琵琶法師の語った事が、もない
やはりい加減の出たらめなのです。
一体琵琶法師などと云うものは、どれか
わたしはあのの小屋に、俊寛様が子供た
つい涙さえ落しました。たとい嘘とは云
があるだけ、わたしでも今の内ありのま
こうあなたはおっしゃるのですか？ な
の事を御話しましょう。しかしわたしは
ない真実と云う事だけです。ではどうか

BOCON
ANA 2018

807 words, 814 characters | Default Style | Japanese | 160%



Features specific to Japanese

- Line breaking rules : ” 禁則処理” Kinsokushori
- [Paragraph style]-[Japanese appearance] tab





line adjustment by hanging punctuation

136

本製本は、綴じた中身を仕上げ、裁ちしてから表紙でくるみ、仕上げた本である。中身より表紙の方が大きくなる。中身よりはみ出した表紙の部分を子リとよんでいる。上製本または仮製本ともいう。

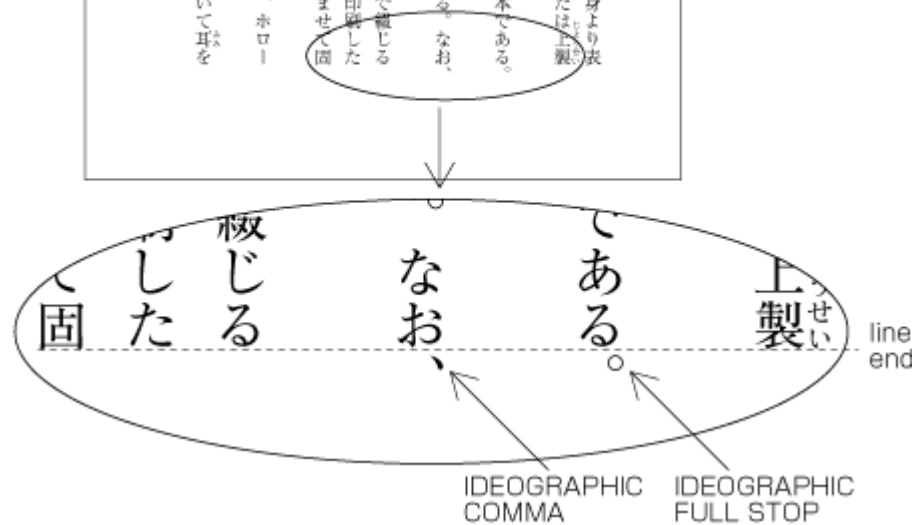
仮製本は綴じた中身に簡単な紙表紙を付け、中身とともに仕上げ裁ちして仕上げた本である。雑誌はほとんどが仮製本である。並製本または並製ともいう。

本製本と仮製本の区別は、表紙に子リがあるかないかにより判断するのが普通である。なお、見返しは本製本では必ず付けるが、仮製本では付ける場合と付けない場合がある。

本を綴じる場合、糸を用いて綴じる糸綴じが一般的であるが、今日は、接着剤だけで綴じる無糸綴じの一種であるあじろ綴じ（網代綴じ）で仕立てる例も多い。あじろ綴じは、印刷した紙を折り畳む際に、本の背にあたる部分に切り込みをいれ、そこから接着剤を浸み込ませて固める綴じ方である。

本製本の背の仕立て方には、フレキシブルバック（柔軟背）、タイトバック（硬背）、ホローバック（腔背）があるが、ホローバックが一般的である。

製本における背の様式としては、丸背と角背（角山）がある。角背のまま山をたたいて耳を出したものを角山と一般によんでいる。





External Character (異体字) & Kanji variants

Kanji variant is a Kanji character with the same meaning but by different shape.

崎 崎 埼
碕 寄

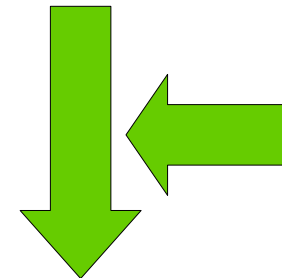


Ideographic **V**ariation **S**equence/Selector

In Japan, using a mechanism called IVS to make it easier to deal with a Kanji variant.

IVS is combine the characters,
as one of the character.

Unicode character



Variation
Sequence



UCSコードは、細かな字形の差異は区別せず、複数の字形を一つのコードに統合している。

	UCSコード		VS
⋮	⋮		
	845A		E0100
	845B		E0101
⋮	⋮		E0102
		⋮	⋮
	8FBA		
	8FBB		E0100
⋮	⋮		E0101
		⋮	⋮

UCSコードと字形を指定するVS(字形選択子)を並べて記述することで、字形を指定することができる。
 ※VSの値は実際の規格と異なります。



IVS using case bug#113481

- Deleting the character behind two same IVS characters with Backspace will result in another character
- Demo



IVS using case bug#113481 (2)

- I explained this bug at LibreOffice conference 2017
- Mark Hung created a patch
- 6.1 : FIXED



Bug #35301: Allow vertical ruby text in the Asian

Hanzi
(Characters)

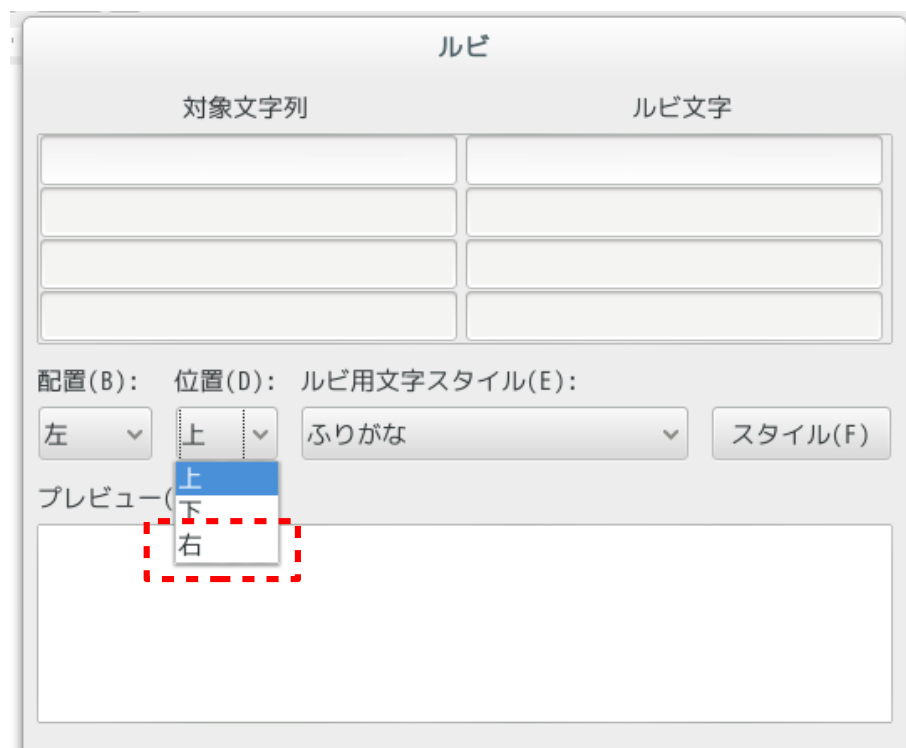
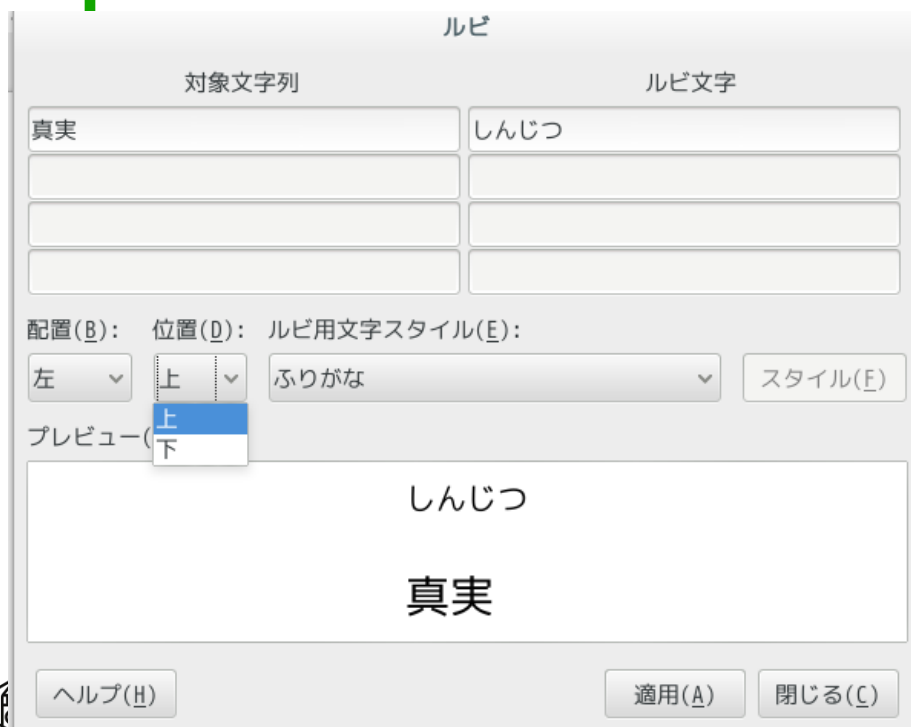
符	フ メ
號	フ ム

Bopomofo
(Phonetic Symbols)



Bug#35301 (2)

- LibreOffice 6.1 can become on the right side with ruby in horizontal writing
- Mark Hung created a patch





Bug 115258 - FILESAVE: text with Asian phonetic guide (ruby) disappear after save to docx and reopened

- Text with Asian phonetic guide (ruby) disappears after saving as “docx” and then reopening
- Mark Hung created a patch
- 6.1 : FIXED



Multilingual issues that matter even at CJK

- Multilingual problem that becomes a problem with CJK
- In some time zone Date is shifted: (Already fixed)
- Bug 59850 - EDITING: typed date decreased one day
- Bug 63230 - Calc: Date decreased by 1 depending on Timezone



Bug 118849 - Impress hang when input multibyte character.

- On Japanese Input mode. Select multiple lines, Input multibyte character, hang.
- Only Linux
- 6.0.5 reproduced, 6.1 not reproduc
- demo



Meta issue for each CJK language

- [META] Japanese language-specific CJK issues
- https://bugs.documentfoundation.org/show_bug.cgi?id=113195
- [META] Traditional Chinese (zh_TW, zh_HK)
 - https://bugs.documentfoundation.org/show_bug.cgi?id=113193
- [META] Simplified Chinese (zh_CN)
 - https://bugs.documentfoundation.org/show_bug.cgi?id=113194
- But, it is not used now



Checking is difficult

- It is difficult to understand the meaning
- Test case, sample file is not enough
- I do not understand when I read the comment to the end



Next task as I think

- 1st: Re-make or arrange the summary and test cases in Bugzilla
 - Against high risk and high probability of regression
- 2nd: CJK Meta issue and list maintenance
- 3rd: Finding critical bugs that users don't immediately notice



Telegram: LibreOffice CJK

- If you want contribute to report / resolving CJK bug, please join this channel
- https://web.telegram.org/#/im?p=s1119055114_15674911573967734518



Conclusion

- Sharing the Unique functions and issues about CJK with a lot of members
- I'd like to work to solve issues with our Japanese members
- wanting to progress organization of Bugzilla such as CJK issue

Thank you !



All text and image content in this document is licensed under the Creative Commons Attribution-Share Alike 4.0 License (unless otherwise specified). "LibreOffice" and "The Document Foundation" are registered trademarks. Their respective logos and icons are subject to international copyright laws. The use of these thereof is subject to trademark policy.

