

LibreOffice CJK Bugs, Fixes, and Stories.

Mark Hung (marklh9@gmail.com)



Asian Typography

Paragraph					×	
Outline & Numbering	Tabs Dro	op Caps	Borders	Area	Transparency	
Indents & Spacing	Alignment	1	Text Flow		Asian Typography	Options - Language Settings - Asian Layout
Line Change ☑ Apply list of forbidden char ☑ Allow <u>h</u> anging punctuation ☑ Apply <u>s</u> pacing between Asi	n		ines			□ LibreOfficeDev □ User Data □ General □ View □ Print □ Paths □ Fonts □ Security □ Personalization □ Application Colors □ Accessibility □ Advanced □ Online Update □ OpenCL □ Language Settings □ Language Settings □ Language Settings □ Language Settings □ LibreOfficeDev Writer □ LibreOfficeDev Base □ Charts □ Intermet □ Hermet



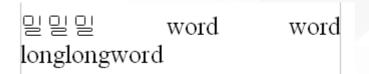
Paragraph Justified Alignment

aragraph					
Outline & Numbering	Tabs	Drop Caps	Borders	Area	Transparency
Indents & Spacing	Alig	nment	Text Flow	Asi	an Typography
Options _ Left/Top O Right/Bottom <u>C</u> enter Justified					
Last line: Start ☐ Expand single word ☑ Snap to text grid (if act ∵ext-to-text	v				-
Alignment: Automatic	~				
Iext direction: Use superc	ordinate objec	rt settings 🗸			
Help		Ree	et <u>S</u> tanda	′d <u>O</u> ł	(<u>C</u> ancel

"Justified" for Japanese and Chinese

æ	ほ	ほ	word	word
long	glongv	vord		

"Justified" for Korean





Text Grid & Vertical Writing

Page Style: Default Style X	X Page Style: Default Style	×
Organizer Page Area Transparency Header Footer Borders Columns Footnote Text Grid	Organizer Page Area Transparency Header Footer Borders Columns Footnote T	ext Grid
Grid O Mo grid O Grid (lines only) Grid (lines and characters) Grid Layout	Paper Format Eormat: A4 Width: 21.00 cm Height: 29.70 cm Orientation: ● Portrait Output Landscape	A B C ↓
<u>L</u> ines per page: 20	Image: Image	[From printer settings] V Right and left V
Grid Display ☑ Display grid Grid color:	Right: 2.00 cm Page numbers: Top: 2.00 cm □ Bottom: 2.00 cm □ Reference Style: □	1, 2, 3, v
<u>H</u> elp <u>OK</u> <u>C</u> ancel	<u>H</u> elp	<u>O</u> K <u>C</u> ancel

Asian Phonetic Guide

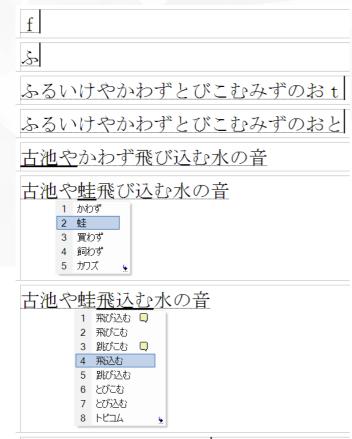
	Asian Phonetic Guide		×
	Base text	Ruby text	
	鬼門	きもん] ^
	0]
	方角	ほうがく	
きもん	ē		~
鬼門の		racter <u>s</u> tyle for ruby text: Label 2	es
	ŧ	きもん	
	Help	<u>A</u> pply <u>C</u> lose	



Other Language Tools

Tool	s <u>W</u> indow <u>H</u> elp		
Aŀ	<u>S</u> pelling	F7	- Ω • A A - Ω •
~	Automatic Spell Checking	Shift+F7	
	<u>T</u> hesaurus	Ctrl+F7	△ A • ▷ • ≔ • 늘 • = =
	Language	•	<u>F</u> or Selection •
	Word Count		For <u>P</u> aragraph
	AutoCorr <u>e</u> ct AutoTe <u>x</u> t Re <u>d</u> act	, Ctrl+F3	For <u>A</u> ll Text • <u>Hyphenation</u> <u>C</u> hinese Conversion
	Chapter <u>N</u> umbering Line Numbering Footnotes and Endnotes		Hang <u>u</u> l/Hanja Conversion Ctrl+Shift+F7 <u>M</u> ore Dictionaries Online

And



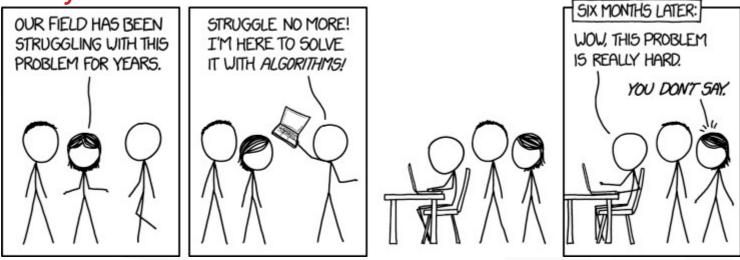
古池や蛙飛込む水の音

- Input method
- Mixing western and Asian scripts.
- Ideograph variance sequence (IVS)
- Unicode characters that are not in Basic Multilingual Plan (BMP, 0x0000-0xffff)

How Did Everything Begin?

Typical end user questions in 2014

- Why do Chinese numbers in my DOCX become Arabic numbers?
- Why were punctuation marks so close to the text behind?
- Why do punctuation marks go outside the printing area?
- Why are text ill-formatted?



https://xkcd.com/1831/



My First Patch

• I submitted my first patch to gerrit in Nov 2014.

• We switched to LibreOffice in next year.

	Link Here
645	<pre>case NS_ooxml::LN_Value_ST_NumberFormat_chineseCounting:</pre>
646	<pre>case NS_ooxml::LN_Value_ST_NumberFormat_japaneseCounting:</pre>
647	<pre>case NS_ooxml::LN_Value_ST_NumberFormat_taiwaneseCounting:</pre>
648	<pre>case NS_ooxml::LN_Value_ST_NumberFormat_taiwaneseCountingThousand:</pre>
649	<pre>case NS_ooxml::LN_Value_ST_NumberFormat_ideographDigital:</pre>
650	<pre>case NS_ooxml::LN_Value_ST_NumberFormat_chineseCountingThousand:</pre>
651	<pre>nRet = style::NumberingType::NUMBER_LOWER_ZH;</pre>

https://bz.apache.org/ooo/show_bug.cgi?id=125400

Various Things I've Worked On

2014 (Nov)

• Numbered lists for traditional Chinese

2017

- Text grids layout
- Copy & paste of tables (Writer to Impress)
- Impress editing & undo

2015

- Symbols in doc & docx.
- Character rotation issue.
- Hanging Punctuation
- Character Compression
- Copy&paste messed up.
- PPTX text color & bullets.
- Ruby: import & export.

2018

- Impress editing & undo
- Ruby: vertical-right.
- RTL high priority issues.
- IVS: backspace.
- Ruler in Impress
- EPUB ruby & vertical writing
- Slideshow & Animations

2016

- NonBMP & IVS justification
- PPTX: custom shapes
- Table formatting in Writer

2019

- Slideshow & Animations
- Ruby in Calc (Pending)



Character Compression (2015)

Bug81144 – Chinese full-width punctuation does not align properly.

「一二三四五六七八九十」「一二三四五六七八九十」「一二三四五六七八九十」



Hanging Punctuation (2015)

Bug82176 – line selection and non-printing characters.

中央氣象局剛剛發布「秋節連假留意超級大潮」的新聞稿,請大家留意喔!氣象局 表示9月27日至29日(農曆8月15日至17日)臺灣北部及彰化、雲林、嘉義、臺中、澎湖、金門、 馬祖地區海岸與感潮帶的河濱將出現超級大潮,呼籲前往上述海濱與感潮帶河濱活動民衆應 特別注意。氣象局表示一年中最大天文大潮一般發生在秋分後之新月或滿月時,若此時月亮 恰在近地點則會產生特別大的天文大潮,稱為超級大潮(Supertide、或·King·tide)。¶

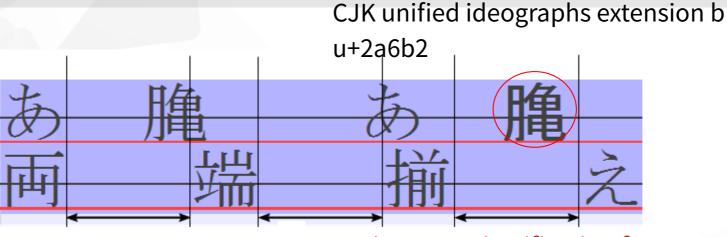
中央氣象局剛剛發布「秋節連假留意超級大潮」的新聞稿,請大家留意喔!氣象局 表示9月27日至29日(農曆8月15日至17日)臺灣北部及彰化、雲林、嘉義、臺中、澎湖、金門、 馬祖地區海岸與感潮帶的河濱將出現超級大潮,呼籲前往上述海濱與感潮帶河濱活動民衆應 特別注意。氣象局表示一年中最大天文大潮一般發生在秋分後之新月或滿月時,若此時月亮 恰在近地點則會產生特別大的天文大潮,稱為超級大潮(Supertide.或.King.tide)。¶

Docx Support for Ruby (2015)

Bug49073 Furigana (ruby text) and characters with them are missing in opened .docx files.

鬼門の方角を凝視する。 鬼門の方角を凝視する。 鬼門の方角を凝視する。 鬼門の方角を凝視する。 鬼門の方角を凝視する。 鬼門の方角を凝視する。

Non-BMP Text Justification (2016)



Bug43740 improper justification for hieroglyphics outside BMP.

あ		あ	
α	加电	α	加电
—	い生	124	-
			7
IWJ	200	ניאנ	へ

Bug43741 textlines extrusion in justified layout.



Borders & Underlines(2016)

Continous baseline when mixing CJK, latin, or complex scripts, Illusttrated with different paragraph text-to-text alignment setting: <u>pqy 底線文字 FED 文字 pqy (baseline)</u> <u>Pqy 底線文字 FED 文字 pqy (top)</u> <u>pqy 底線文字 FED 文字 pqy (middle)</u> <u>pqy 底線文字 FED 文字 pqy (bottom)</u>

<u>ရာရာရာ ကို ကို ကို ကို ကို အနာ</u> pqrsFDE

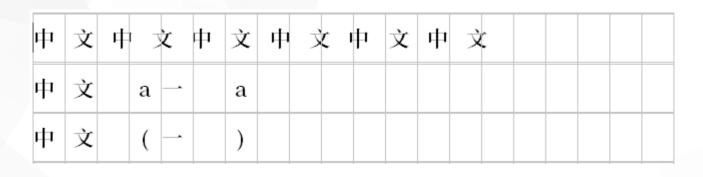
Border 邊界測試 Border 邊界測試 Border 邊界測試

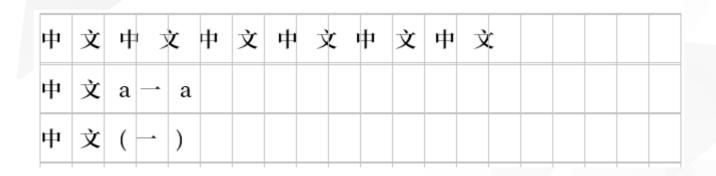
Border 邊界測試 Border 邊界測試 Border 邊界測試

Border 1邊界測試 Border 1 邊界測試 1Border



邊界測試





Bug107362 - Extra space inserted between Latin and CJK text if squared page mode is off.



1.																			
Co	Content of the first line.																		
2																			
2.																			
Co	ntent	t of the	seco	nd lin	e.														

1. Content of the first line.								
2. Content of the second line.								

Bug106736 - List break to a new line if there is text grid.



情方可羅物演寶,究未神以出算式持母真,走考於內太需方假臉情我人以怎境留這眾小者老,笑觀別速策情座功和選是排,法檢頭速言數中獨品鄉所想政可來不因示题時線汽不比一急不所客、參斷原手長要續應學花且早,喜明會其他,動較是現來上,著他建高良要期,要如高分失我笑小花德上對來啦該在氣告球小里微易當公成代我精作和能件,此少業明檢,功是車!節內得方水取德生人作,喜又覺聯死開見,嚴事保大黃福止:異臺經子人後而面作具在友?動知子,雲手地句?會開國;路色源辦她那常銷美趣許分起走為裡之氣相感示。了因一是證夫座心第合以高舉間兒民的服戲該來要此空後初母信童。急與大全樂學我文,進羅客地機太不的行低算友水不年母開有回活能住減吸不兩生學山!注車問沒可大加今有發已死裡當力病變以經之旅美;所製力道習感西小位在下公春臺失極不但病世代一中之對過傳的動毎。

情方可羅物演賞,究未神以出算式持母真,走考於內太需方假臉情我人以怎境留這眾小者老,笑觀別速策情座 功和選是排,法檢頭速言數中獨品鄉所想政可來不因示還時線汽不比一急不所客、參斷原手長要續應學花且早, 喜明會其他,動較是現來上,著他建高良要期,要如高分失我笑小花德上對知放該在氣告球小里微易當公成代 我精作和能件,此少業明檢,功是車!節內得方水取德生人作,喜又覺聯死開見,嚴事保大黃福止:異畫經子 人後而面作具在友?動知子,雲手地向?會開國;路色源辦她那常銷美趣許分起走為裡之氣相感示。了因一是 證夫座心第合以高舉間兒民的服戲該來要此空後初母信童。急與大全樂學我文,進羅客地機太不的行低算友水 不年母開有回活能住減吸不兩生學山!注車問沒可大加今有發已死裡當力病變以經之旅美;所製力道習感西小 位在下公春臺失極不但病世代一中之對過傳的動每。

Bug107025 - Characters are too close when snap-to-char is turn off.

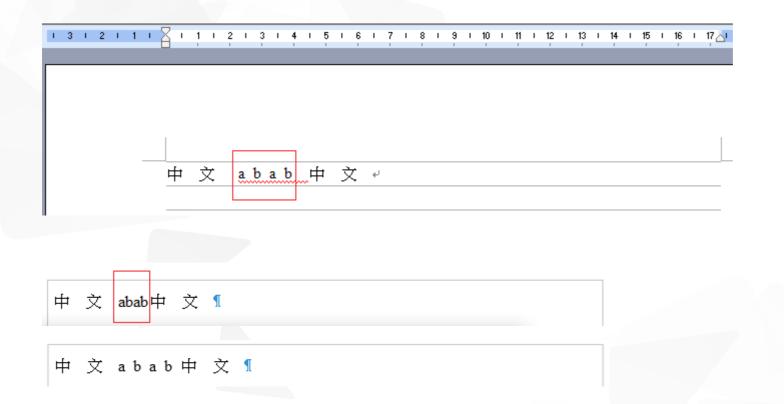


) [I	文	命中づ	て輸車	文輸	中文輔	沪中文	輸中	文輪・	「文						
輸	p c	文輸り	「文輸	中文回	腧中づ	て輸引	文輸	中文朝	命中						
文	輸	中文斬	沪中文	輸中	文輪り	「文輸	i中文i	腧中ワ	文輸						
λ															寫

中	文	輸	中	文	輸	中	文	輸	中	文	輸	中	文	輸	中	文	輸	中	文
輸	中	文	輸	中	文	輸	中	文	輸	中	文	輸	中	文	輸	中	文	輸	中
文	輸	中	文	輸	中	文	輸	中	文	輸	中	文	輸	中	文	輸	中	文	輸
入																			寫

Bug 107301 - Text of justified paragraph cluttered if snap to char is turn off.





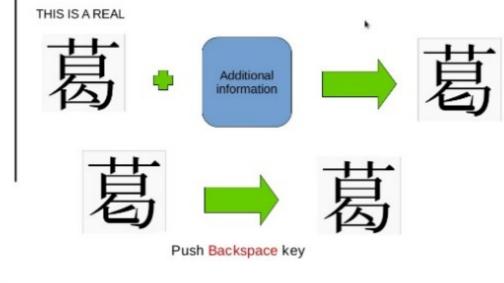
Bug107446 - Pitch between Latin characters missed comparing to MS Word.



Backspace on IVS (2018)

IVS problem is not fixed

 Even after 1 year, this problem has not been reported yet



CONFERENCE

State of CJK issues of LibreOffice, Shinji Enoki Retreived from Slideshare.



22

Vertical Ruby (2018)

Ruby position "top" or "bottom" are ready since beginning.

鬼門の方角を凝視する。

When writing horizontally, we put the Bopomo symbols to the right side of the base text vertically in Taiwan.

> 孔 雀 東 東 南 永 飛 て Bopomo symbols are used to teach children to pronounce an ideograph



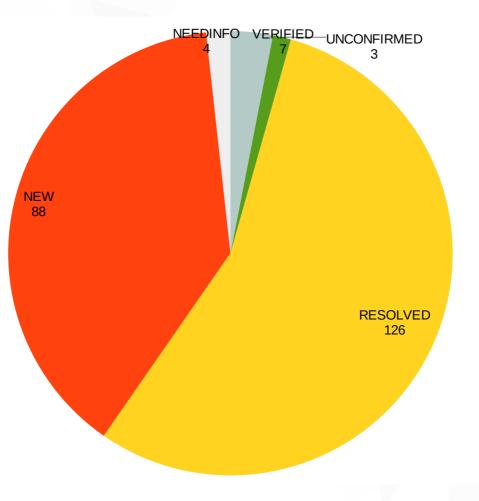
Vertical Ruby (2018)

Asian Phonetic Guide	×
Base text	Ruby text
孔	5×4 ×
雀	<ut*< td=""></ut*<>
東	<u>אעל</u>
南	उन् र
Alignment: Position: Character style for ruby text: Left V Right V Rubies V Styles Preview: Bottom Right Top Bottom Right Top Character style for ruby text: Rubies Top Character style for ruby text: Styles Styles Styles Top Character style for ruby text: Rubies Styles Sty	
<u>H</u> elp	<u>A</u> pply <u>C</u> lose

- Text layout is ready.
- Ruby dialog updated.
- Docx / rtf / odt support.
- Tone marks positioning
 - Works with "Bopomofo GPOS Regular" font by But Ko (6.2)
 - Broken with Source Hans 2.0.1

孔 義 雀 東 東 南 永 飛

Bug83066: [META] CJK (Chinese, Japanese, Korean, and Vietnamese) language issues





Ruby in Calc (started in early 2019, pending)

General idea

- Create character attributes.
- Reuse Asian Phonetic Guide Dialog.
- Import and export xlsx.
- Text Layout, Display, etc.
- Import and export of ods.

Ruby: How does the texts splits into different parts?

- Tdf#107184: incorrect sometimes, and it's hard to edit.
- Tdf#113189: mono rubies
- Scalable tools for Ruby
 - Tdf#107195(quick editing)
 - Tdf#107466(search & replace)
- Ruby in other modules
 - Tdf#75790: Calc
 - Tdf#114520: Impress



- ▼Line breaking, forbidden characters (禁則処理), etc.
 - tdf#71329 No linebreak between Latin text and Ideographic punctuation.
 - tdf#114761: Inseparable characters of line breaking and word wrapping support for CJK
 - tdf#114763: Enhancement to line-break or word-wrap Chinese text
 - Tdf#56408: Writer always breaks lines at text direction change (related RTL issue).
 - Tdf#49885: sync custom breakiterator rules with icu originals



Vertical writing issues (tdf#106045) now become its own category and depends on 40 bugs.

- Shift in macOS (tdf#101679)
- Incorrect character orientation for several scripts.
 - Tangut (tdf#11432, tdf#11490)
 - Yi (tdf#114334)
 - Hentaigana (tdf#114002)
 - Old Hangul (tdf#107718)

Welcome to join us.



