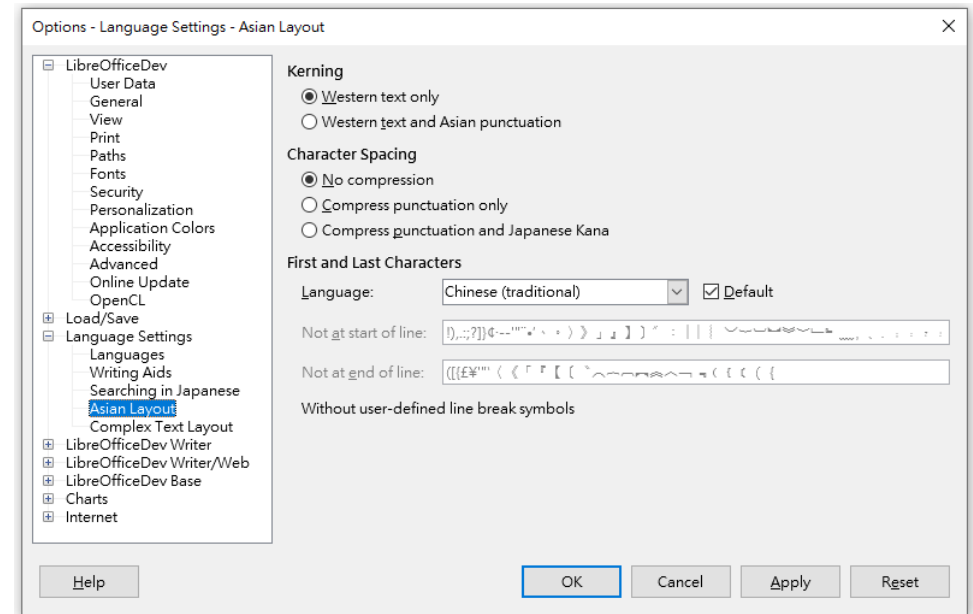
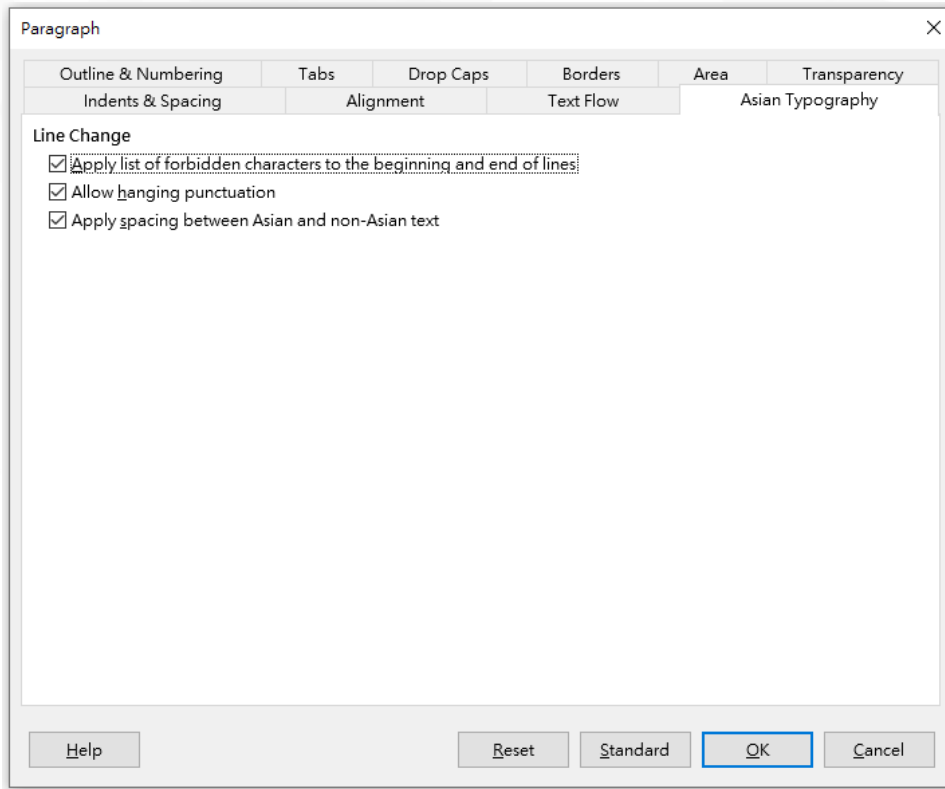




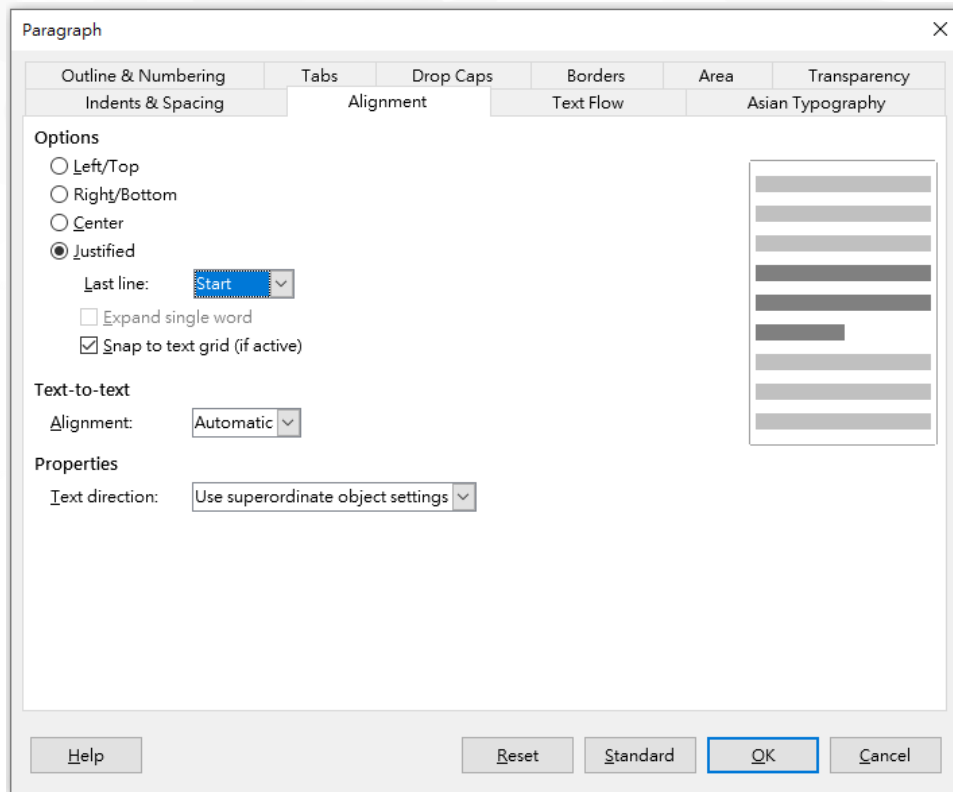
LibreOffice CJK のバグ、 修正、そして物語

Mark Hung (marklh9@gmail.com)

アジアの組版



段落の「両端揃え」 (Justification)



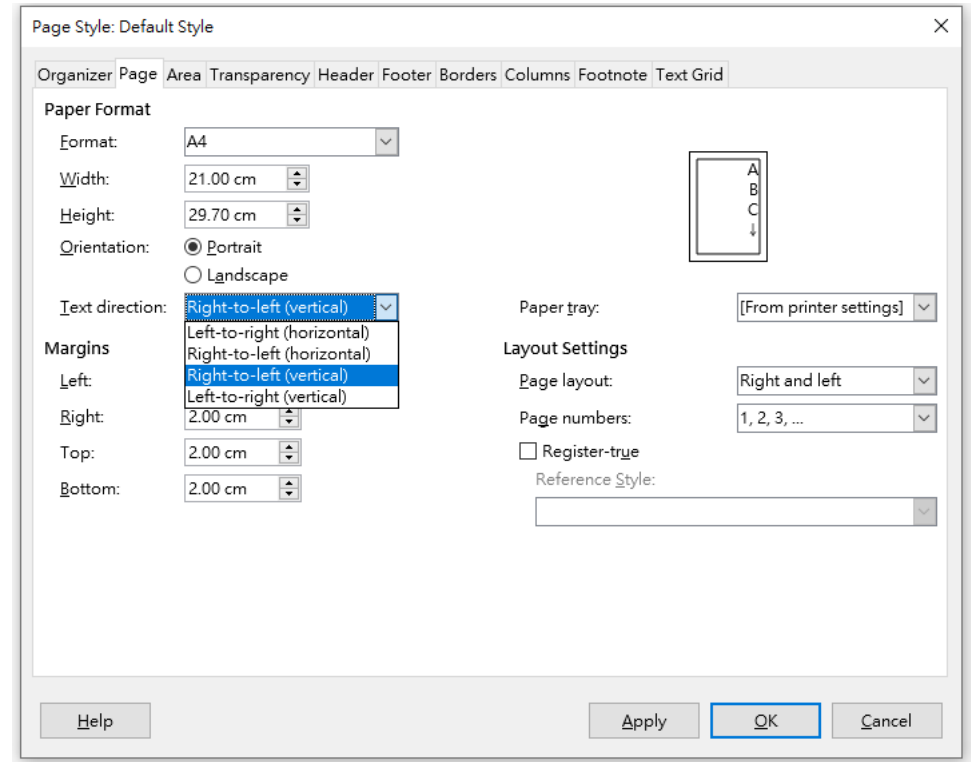
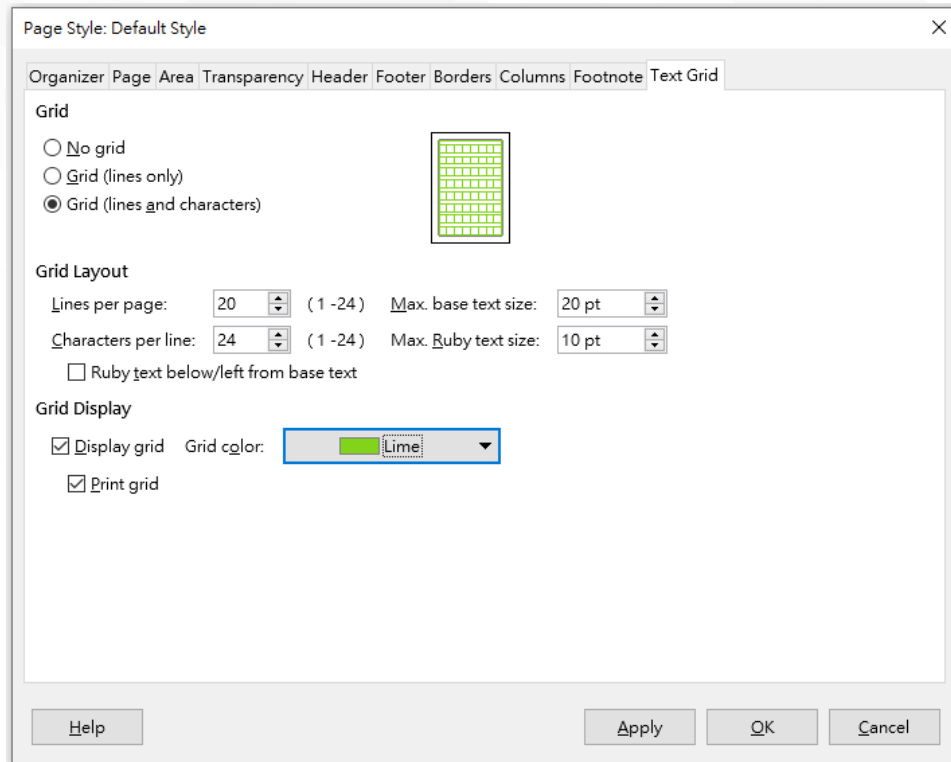
日本語と中国語の「両端揃え」

ほ ほ ほ word word
longlongword

韓国語の「両端揃え」

밀 밀 밀 word word
longlongword

行数と文字数、縦書き



ルビ (Asian Phonetic Guide)

きもん
鬼門の

Asian Phonetic Guide

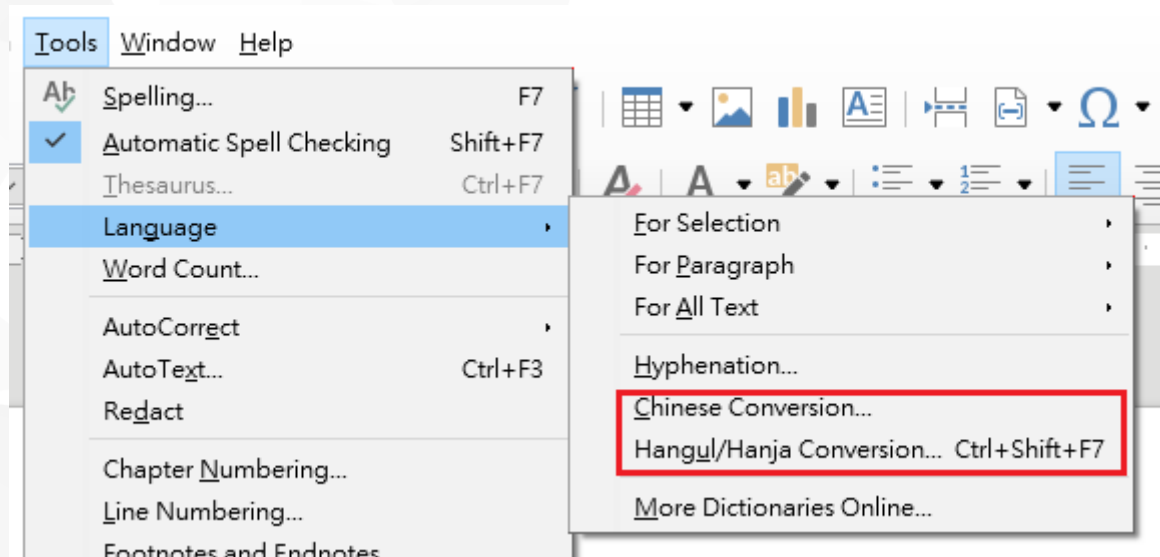
Base text	Ruby text
鬼門	きもん
の	
方角	ほうがく
を	

Alignment: Position: Top Character style for ruby text: ListLabel 2

Preview:

きもん

その他の言語ツール



それから …

f

ふ

ふるいけやかわずとびこむみずのおと

ふるいけやかわずとびこむみずのおと

古池やかわず飛び込む水の音

古池や蛙飛び込む水の音

- 1 かわず
- 2 蛙
- 3 買わず
- 4 飼わず
- 5 カワズ

古池や蛙飛込む水の音

- 1 飛び込む
- 2 飛びこむ
- 3 跳びこむ
- 4 飛込む
- 5 跳び込む
- 6 とびこむ
- 7 とび込む
- 8 トビコム

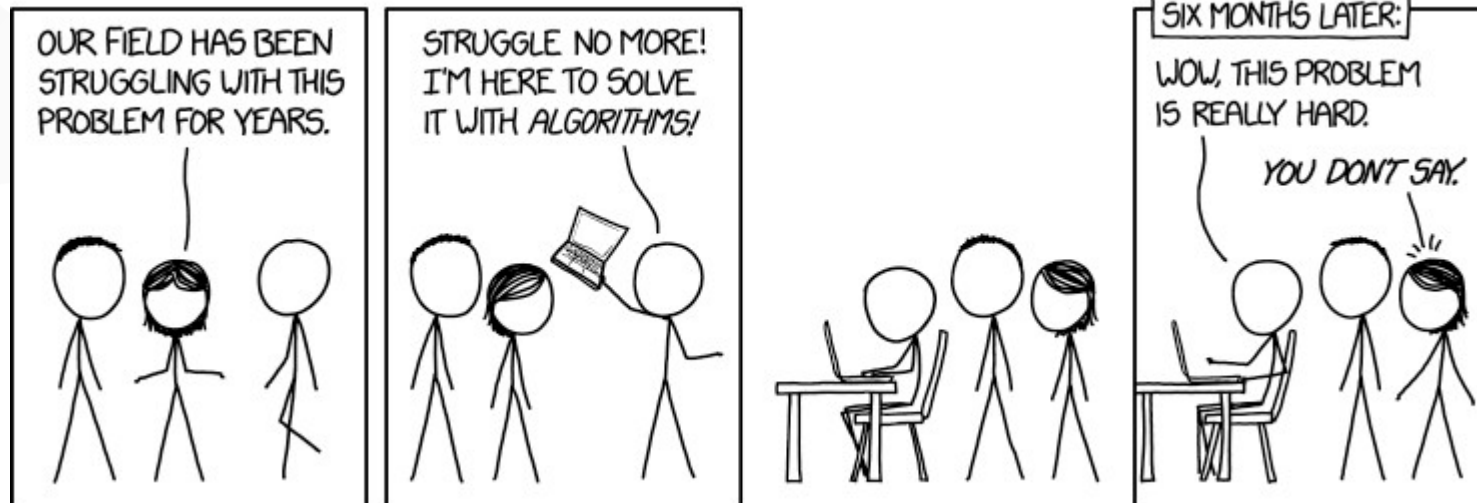
古池や蛙飛込む水の音

- ▼ 入力メソッド
- ▼ 欧文・和文の混在
- ▼ 異体字対応 (IVS)
- ▼ 基本多言語面 (BMP, 0x0000-0xffff) に存在しない Unicode 文字

すべての始まりはどんなふうにな？

2014年、エンドユーザーからの典型的な質問

- なぜ DOCX の漢字の番号付き箇条書きがアラビア数字になるの？
- なぜ句読点がテキストにやたらと近く出るの？
- なぜ句読点が印刷可能領域の外に出るの？
- なぜテキストの書式がおかしくなるの？



<https://xkcd.com/1831/>

最初のパッチ

- 2014年11月、最初のパッチを `gerrit` に登録
- 翌年、`LibreOffice` に切り替え

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

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

これまでやってきた様々なこと

2014/11

- 繁体字の
番号付き箇条書き

2017

- 行数と文字数レイアウト
- 表のコピー&ペースト
(Writer → Impress)
- Impress の編集と undo

2015

- Doc / Docx の記号処理
- 文字回転問題
- ぶら下がり禁則
- 文字詰まり問題
- コピー&ペースト修正
- PPTX 文字色と箇条書き
- ルビ：相互運用対応

2018

- Impress の編集と undo
- ルビ：縦書き
- RTL の優先度高の問題
- IVS: バックスペース問題
- Impress のルーラー
- EPUB ルビと縦書き
- スライドショー・アニメ

2016

- BMP 範囲外文字 & IVS
対応
- PPTX: カスタムシェープ
- Writer の表書式設定

2019

- スライドショー・アニメ
- Calc の Ruby (中断)

文字詰まり問題 (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」

「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」

句読点ぶら下がり禁則 (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 のルビサポート (2015)

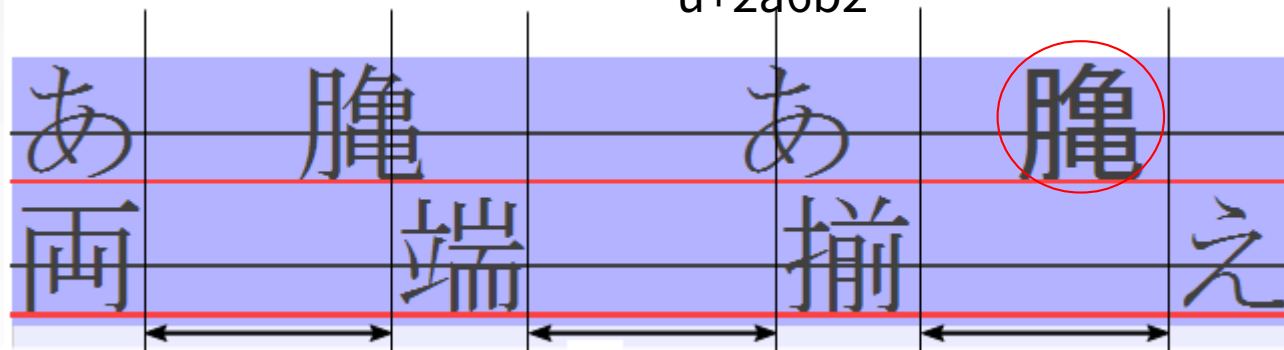
Bug49073 Furigana (ruby text) and characters with them are missing in opened .docx files.

のをする。○
のをする。○
のをする。○
のをする。○
のをする。○
のをする。○
のをする。○

鬼門きもんの方角ほうかくを凝視ぎょうしする。○
鬼門きもんの方角ほうかくを凝視ぎょうしする。○
鬼門きもんの方角ほうかくを凝視ぎょうしする。○
鬼門きもんの方角ほうかくを凝視ぎょうしする。○
鬼門きもんの方角ほうかくを凝視ぎょうしする。○
鬼門きもんの方角ほうかくを凝視ぎょうしする。○
鬼門きもんの方角ほうかくを凝視ぎょうしする。○

BMP 外文字の両端揃え (2016)

CJK unified ideographs extension b
u+2a6b2



Bug43740 improper justification for hieroglyphics outside BMP.



Bug43741 textlines extrusion in justified layout.

杳線・下線問題 (2016)

Continous baseline when mixing CJK, latin, or complex scripts,
Illustrated with different paragraph text-to-text alignment setting:

pqy 底線文字 FED 文字 pqy (baseline)

pqy 底線文字 FED 文字 pqy (top)

pqy 底線文字 FED 文字 pqy (middle)

pqy 底線文字 FED 文字 pqy (bottom)

ബു ബുബു സു സുസു സു സു ഷു ഷു 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 Border1 邊界測試 1Border 邊界測試

行数と文字数 (2017)

中	文	中	文	中	文	中	文	中	文	中	文								
中	文		a	一		a													
中	文		(一)													

中	文	中	文	中	文	中	文	中	文	中	文								
中	文		a	一		a													
中	文		(一)													

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

行数と文字数 (2017)

情方可羅物演實，究未神以出算式持母真，走考於內太需方假臉情我以怎境留這眾小者老，笑觀別速策情座功和還是排，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。

情方可羅物演實，究未神以出算式持母真，走考於內太需方假臉情我以怎境留這眾小者老，笑觀別速策情座功和還是排，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.

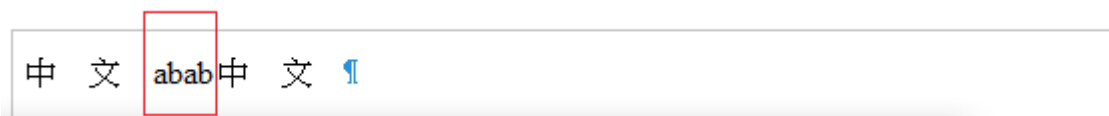
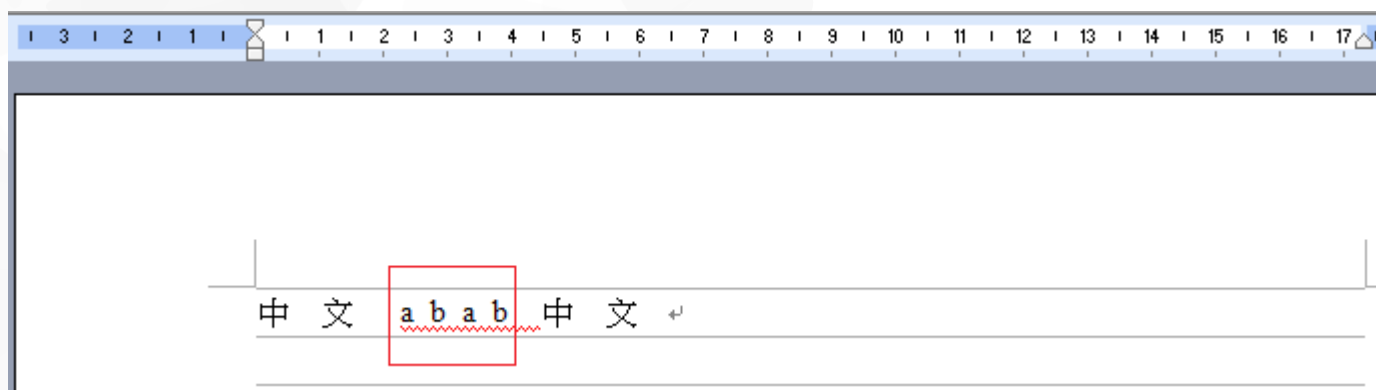
行数と文字数 (2017)

中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入
中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入
中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入	中文	輸入
入																									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													
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						
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.

行数と文字数 (2017)



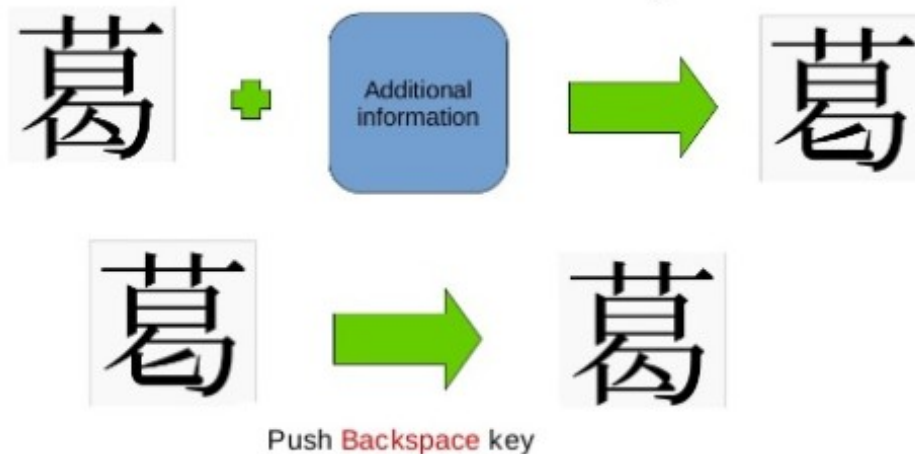
Bug107446 - Pitch between Latin characters missed comparing to MS Word.

IVS バックスペース問題 (2018)

IVS problem is not fixed

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

THIS IS A REAL



SOAS
CONFERENCE

22


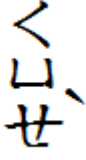
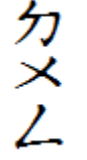
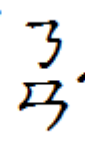
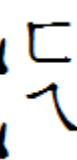
State of CJK issues of LibreOffice, Shinji Enoki
Retrieved from Slideshare.

縦書きルビ (2018)

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

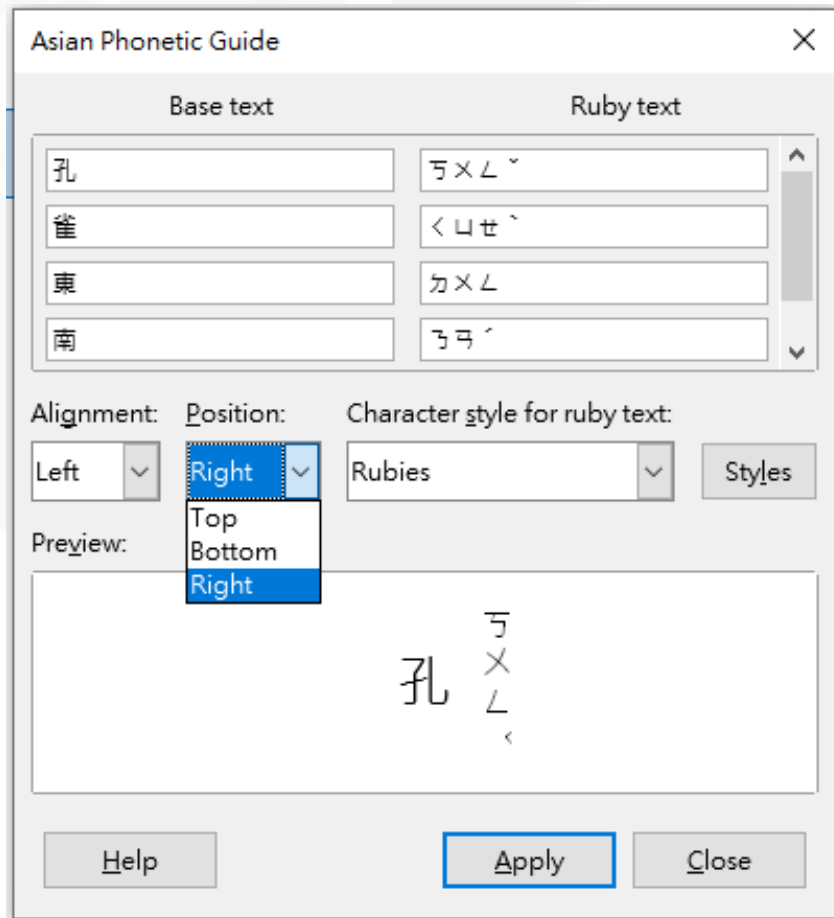
き も ん ほ う が く ぎ ょ う し
鬼門の方角を凝視する。

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

孔  雀  東  南  飛 

Bopomo symbols are used to teach children to pronounce an ideograph

縦書きルビ (2018)

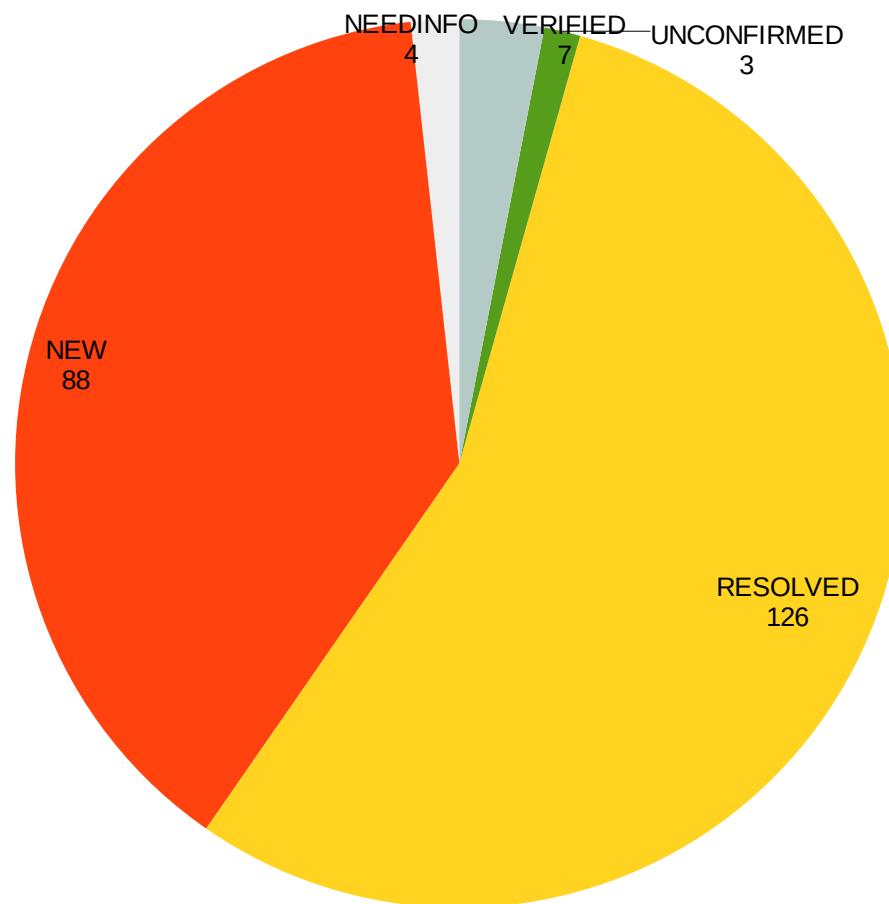


- ▶ Text layout is ready.
- ▶ Ruby dialog updated.
- ▶ Docx / rtf / odt support.
- ▶ Tone marks positioning
 - ▶ Works with "Bopomofo GPOS Regular" font by But Ko (6.2)
 - ▶ Broken with Source Hans 2.0.1

孔_{ㄅㄨㄣˊ} 雀_{ㄑㄩㄝˇ} 東_{ㄉㄨㄥˊ} 南_{ㄋㄢˊ} 飛_{ㄈㄟˊ}

次は何をやるうか？

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



次は何をやるうか？

- ▼ Calc のルビ (2019 初頭に開始、中断中)
 - ▼ 基本的なアイディア
 - ▼ 文字の属性を作成
 - ▼ Writer のルビダイアログを流用
 - ▼ xlsx へのインポート・エクスポート
 - ▼ テキストレイアウト、表示など
 - ▼ ods へのインポート・エクスポート

次は何をやるのか？

- ▼ ルビ：テキスト分割はどういうふうに行われるか？
 - ▼ Tdf#107184: incorrect sometimes, and it's hard to edit.
 - ▼ Tdf#113189: mono rubies
- ▼ ルビの使いやすいつール
 - ▼ Tdf#107195(簡易編集)
 - ▼ Tdf#107466(検索・置換)
- ▼ 他のモジュールへのルビ
 - ▼ Tdf#75790: Calc
 - ▼ Tdf#114520: Impress

次は何をやるうか？

- ▼ 行分割、禁則処理など
 - ▼ 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

次は何をやるうか？

- ▼ 縦書き問題 (tdf#106045) は独立したカテゴリとなつて、40 個のバグが登録されている
 - ▼ macOS でのズレ (tdf#101679)
 - ▼ いくつかの文字で文字の不正な回転
 - ▼ 西夏語 (Tangut) (tdf#11432, tdf#11490)
 - ▼ 彝文字 (Yi) (tdf#114334)
 - ▼ 変体仮名 (tdf#114002)
 - ▼ 旧ハングル (tdf#107718)

次は何をやるうか？

- ▼ 仲間入りを歓迎します

