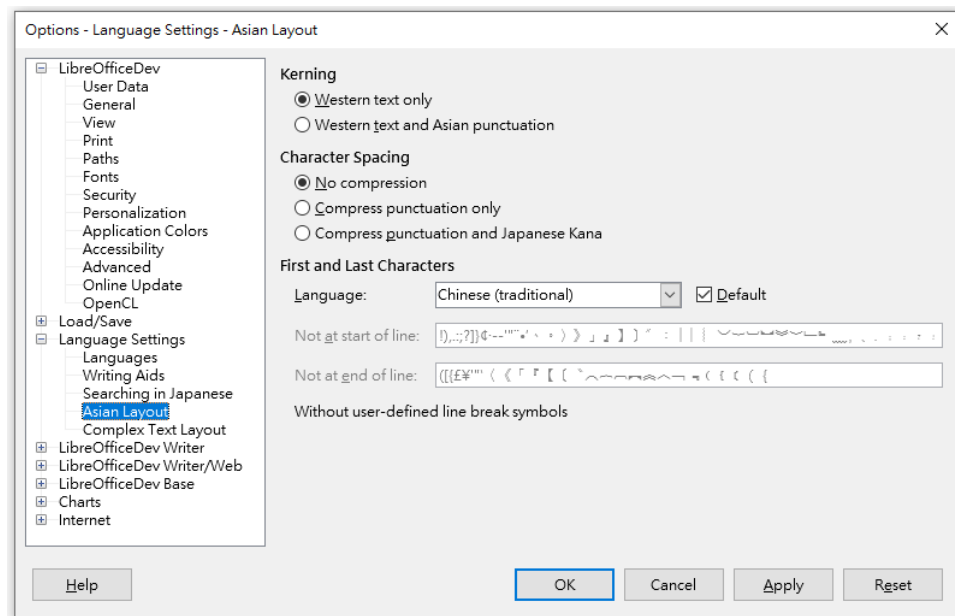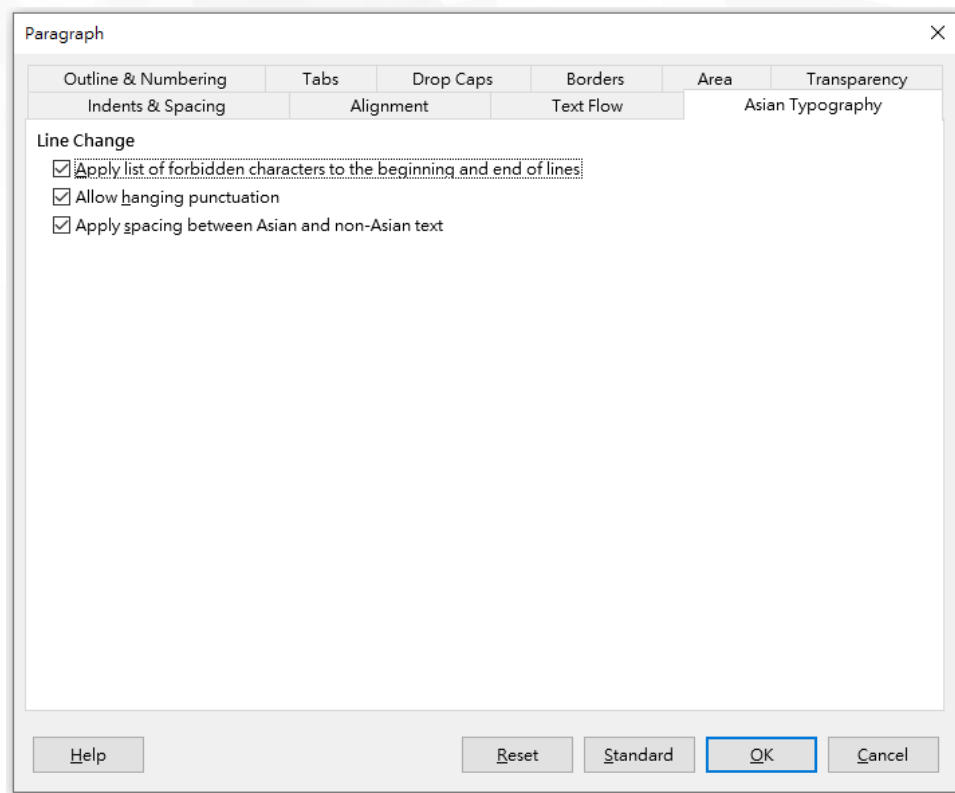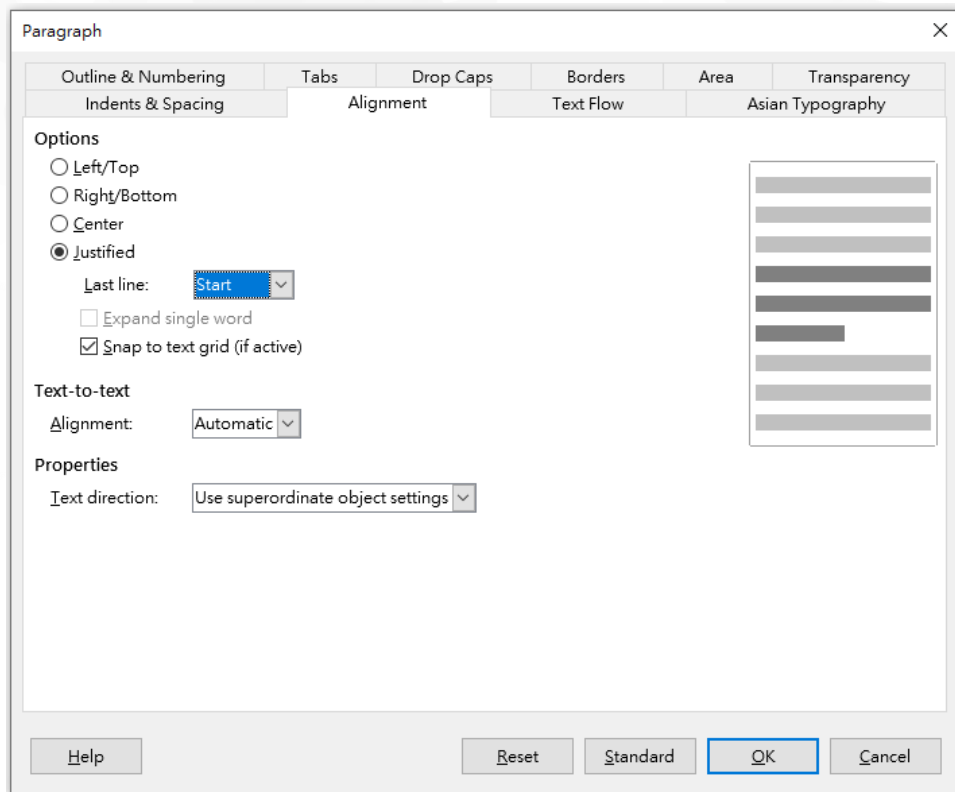# LibreOffice CJK Bugs, Fixes, and Stories.

Mark Hung (marklh9@gmail.com)

LibreOffice

Asian Typography

Paragraph Justified Alignment

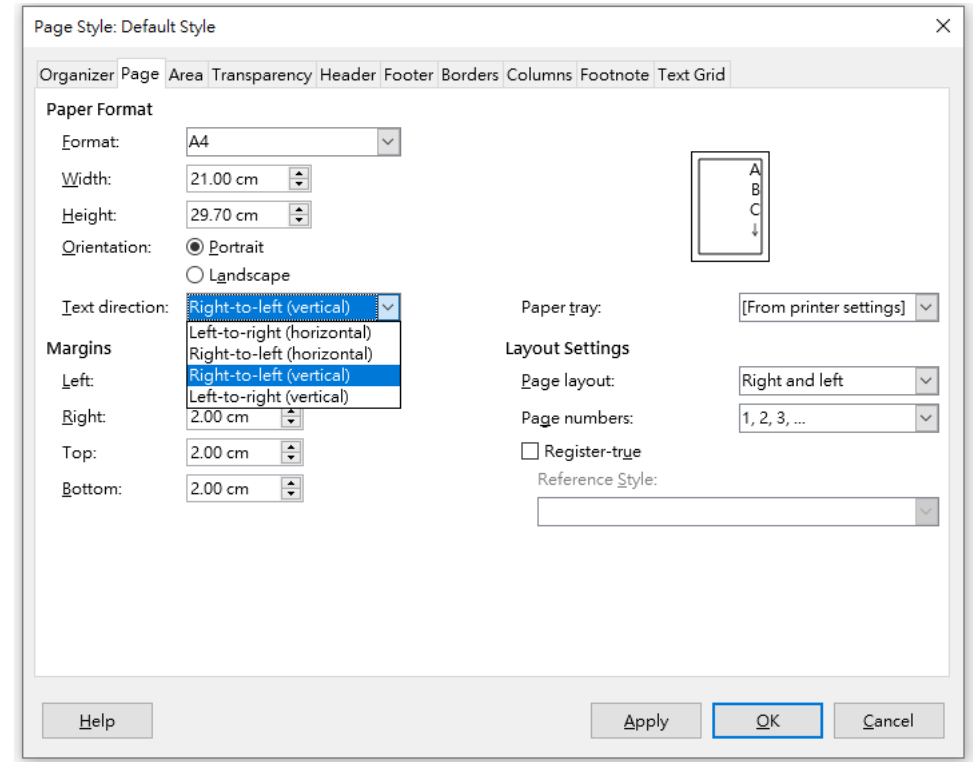"Justified" for Japanese and Chinese

"Justified" for Korean

LibreOffice

Text Grid & Vertical Writing

Asian Phonetic Guide

Other Language Tools

LibreOffice

And .....

- 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/
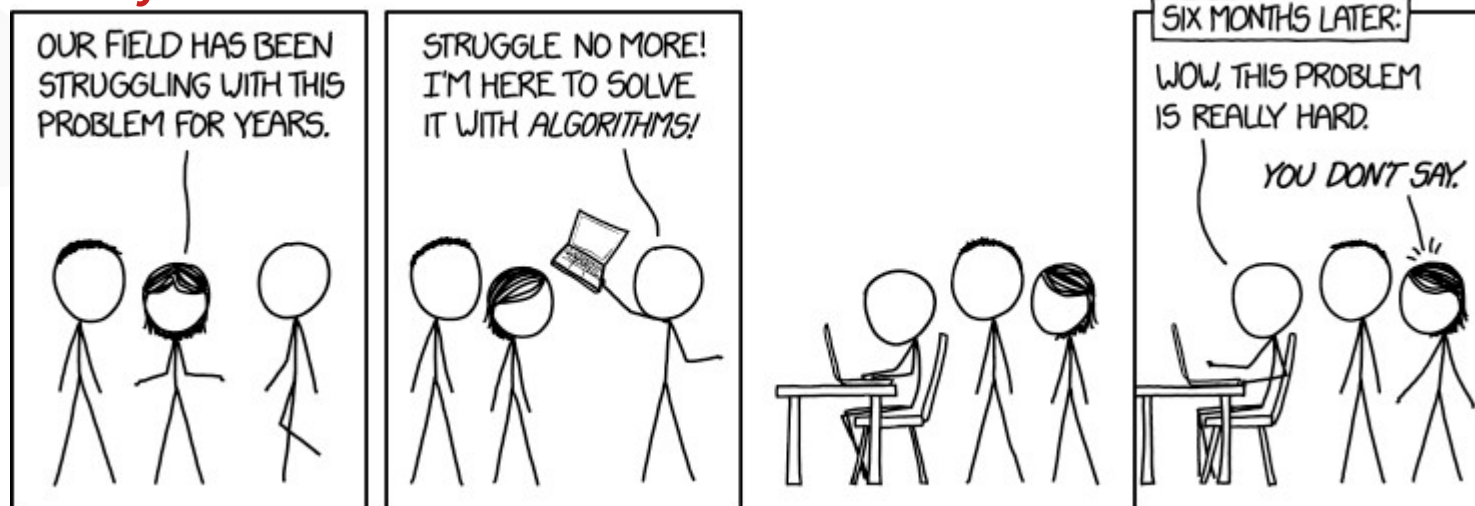
My First Patch

- I submitted my first patch to gerrit in Nov 2014.
- We switched to LibreOffice in next year.

```
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;
```

Link Here

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

Various Things I've Worked On

2014 (Nov)
- Numbered lists for traditional Chinese

2015
- Symbols in doc & docx.
- Character rotation issue.
- Hanging Punctuation
- Character Compression
- Copy&paste messed up.
- PPTX text color & bullets.
- Ruby: import & export.

2016
- NonBMP & IVS justification
- PPTX: custom shapes
- Table formatting in Writer

2017
- Text grids layout
- Copy & paste of tables (Writer to Impress)
- Impress editing & undo

2018
- Impress editing & undo
- Ruby: vertical-right.
- RTL high priority issues.
- IVS: backspace.
- Ruler in Impress
- EPUB ruby & vertical writing
- Slideshow & Animations

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」

「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.

LibreOffice

Docx Support for Ruby (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。

き もん　ほう がく　　ぎょうし
鬼門の方角を凝視する。
き もん　ほう がく　　ぎょうし
鬼門の方角を凝視する。
き もん　ほう がく　　ぎょうし
鬼門の方角を凝視する。
き もん　ほう がく　　ぎょうし
鬼門の方角を凝視する。
き もん　ほう がく　　ぎょうし
鬼門の方角を凝視する。
き もん　ほう がく　　ぎょうし
鬼門の方角を凝視する。

LibreOffice

Non-BMP Text Justification (2016)

CJK unified ideographs extension b
u+2a6b2
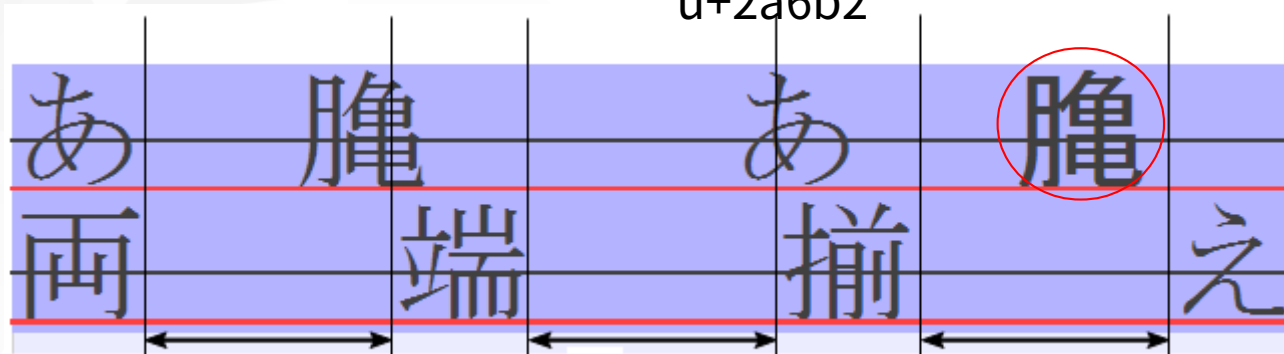
Bug43740 improper justification for hieroglyphics outside BMP.
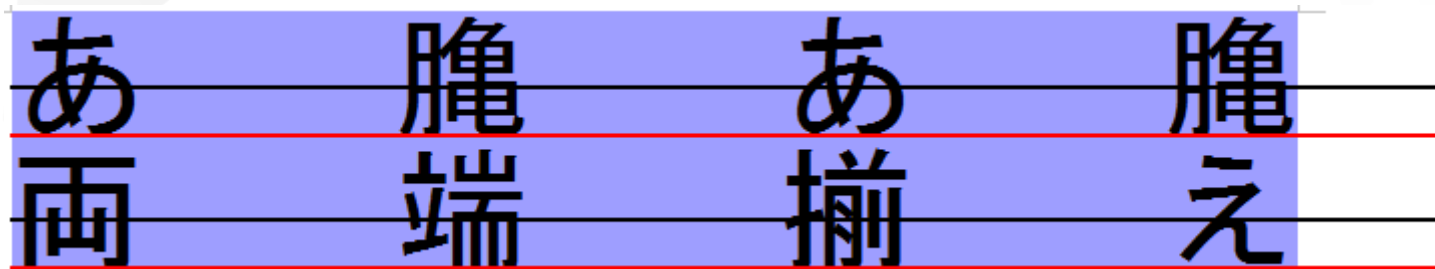
Bug43741 textlines extrusion in justified layout.

Borders & Underlines(2016)

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

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)

ബൃ ബൃബൃ ന്റ ന്റന്റ ന്റ ന്റ ഷൃ ഷൃ 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

LibreOffice

Text Grid (2017)

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

LibreOffice

Text Grid (2017)

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

Text Grid (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人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敬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：異盡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雨生學山！注車問沒可大加今有發己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人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：異盡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雨生學山！注車問沒可大加今有發己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.

Text Grid (2017)

Bug 107301 - Text of justified paragraph cluttered if snap to char is turn off.
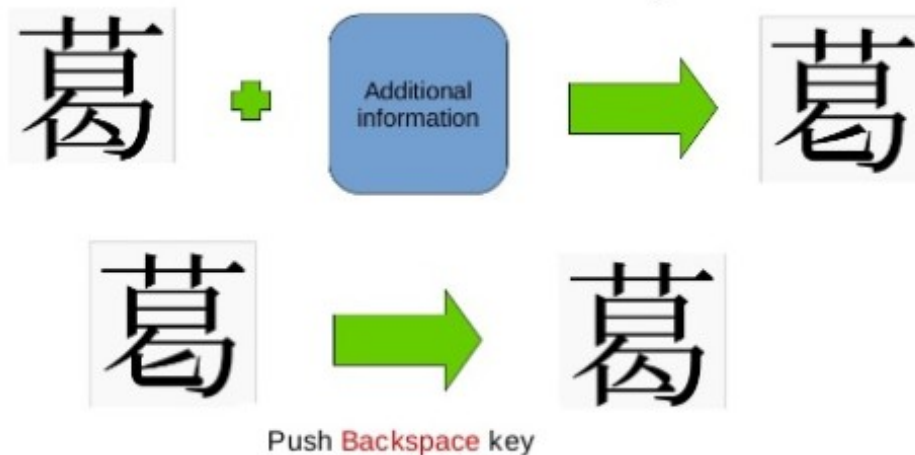
Text Grid (2017)
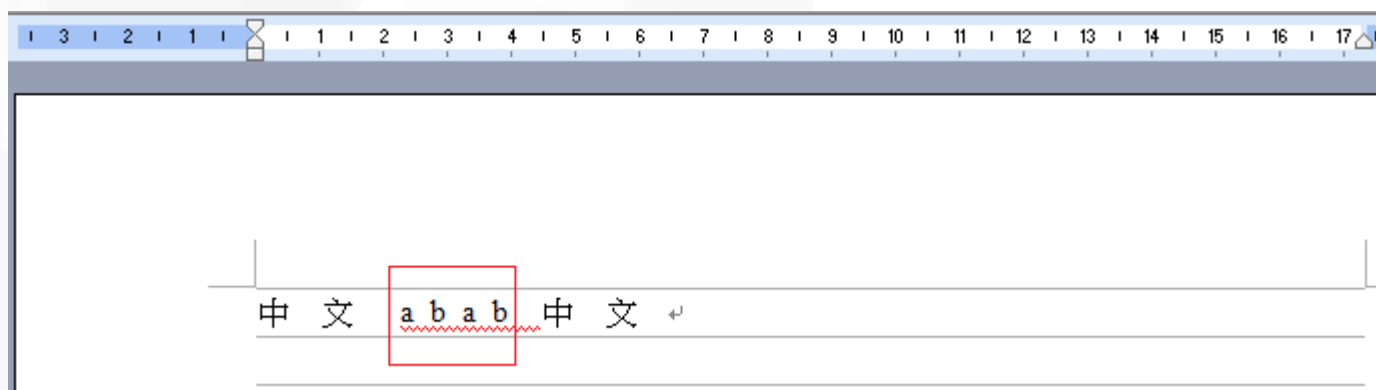
Bug107446 - Pitch between Latin characters missed comparing to MS Word.

Backspace on IVS (2018)

IVS problem is not fixed

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

THIS IS A REAL

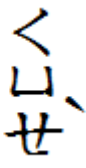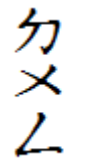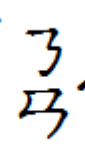葛 + Additional information → 葛

葛 → 葛

Push Backspace key

22

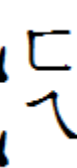State of CJK issues of LibreOffice, Shinji Enoki
Retreived from Slideshare.

Vertical Ruby (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

LibreOffice

Vertical Ruby (2018)
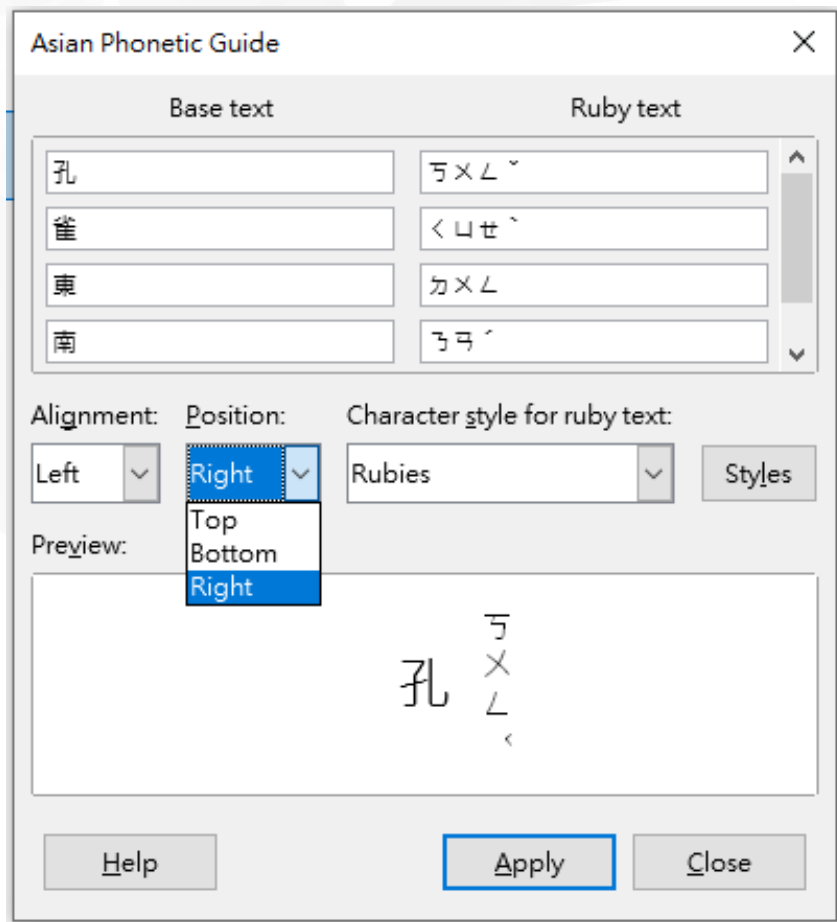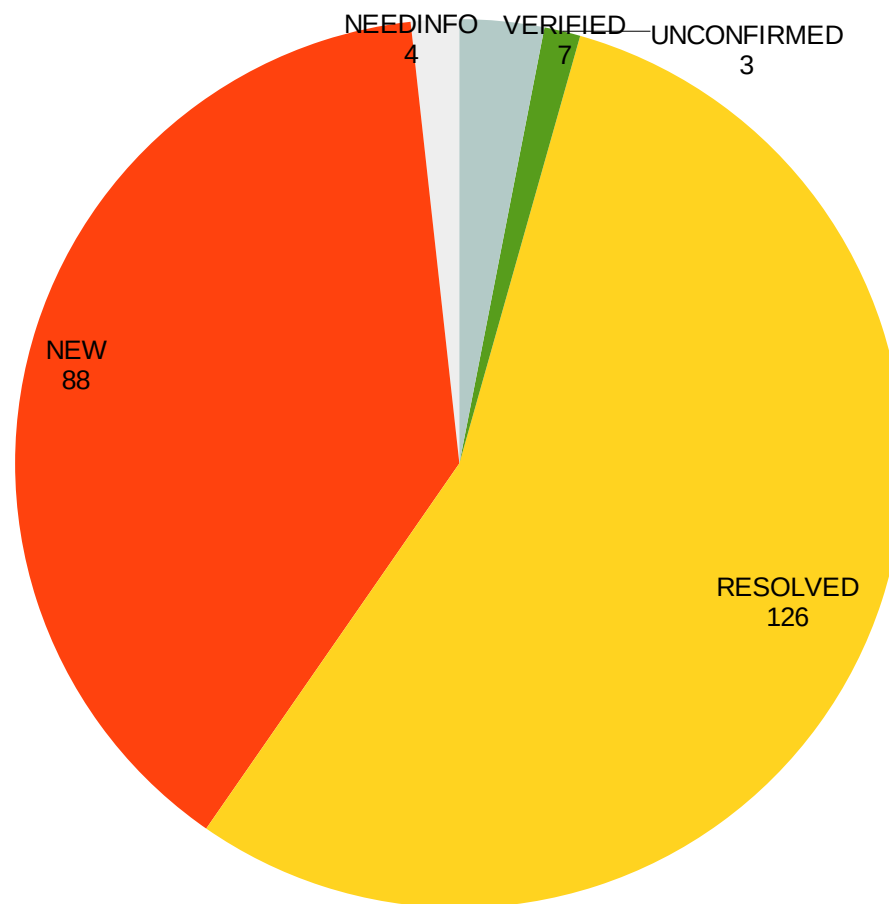
- Text layout is ready.
- Ruby dialog updated.
- Docx / rtf / odt support.
- Tone marks positioning
  - Works with "Bopomofo GPOS Regular" font by But Ko (6.2)
  - Broken with Source Hans 2.0.1

What's Next?

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

NEEDINFO 4 · VERIFIED 7 · UNCONFIRMED 3 · NEW 88 · RESOLVED 126

What's Next?

- 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.

LibreOffice

What's Next?

- 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

What's Next?

- 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

What's Next?

- 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)

LibreOffice

What's Next?

- Welcome to join us.