

Voice Marking Tools

[Click here to download the Voice Marking Tools^{\[1\]}.](#)

This is a collection of custom tools for Paratext. These tools all relate to either marking the words of Jesus for red-letter publications, or marking the boundaries of character voices as assigned in Glyssen.

For a quick introduction, watch this video:

Background: Markers and Glyssen

\wj Tags

If you want a print or electronic publication to display the words of Jesus in red, the publishing software (so far) requires that text is marked with `\wj ... \wj*` tags. However, USFM has some rather cumbersome demands of these tags:

- You must close and reopen the tags wherever the text is interrupted by a verse number, a paragraph marker, a footnote, a cross-reference, figure, a section header, a chapter break, or any such item that is external to the words of Jesus. (Violating some of these won't trigger marker warnings, but will trigger schema warnings nonetheless.)
- Any time character/word-level marking occurs within the words of Jesus, the markers need a plus sign (+) to indicate that this style is nested within the `\wj` style, not replacing it. For example, `\w ... \w*` must become `\+w ... \+w*`

Quote Milestones

USFM 3 introduced milestones^[2], a kind of self-closing marker that can be external to the hierarchical formatting requirements that `\wj` markers are stuck in. One kind of milestone, the quote milestone, is particularly designed for retaining in the text the boundaries of different voices that may be used in a multi-voice audio recording. Quote milestones are not necessary to mark the narrator's voice, as that is the default. Also, quote milestones are not used for content like section headings, as the section heading marker inherently indicates its voice. This means that when a section heading interrupts a voice, no milestones are necessary to close and reopen that voice.

The primary tool for dividing the text between voices is Glyssen. Glyssen knows the quotation rules for a Paratext project. Glyssen knows in which verses it can expect different characters to speak. And so Glyssen is able to help the translator sort out the remaining ambiguities much faster than he could do manually one by one. Having done this once, as the Paratext project continues to be edited, you can help Glyssen to keep track of those boundaries more accurately by inserting milestones back into the Paratext project.

Work done in Glyssen can be exported as a *glyssenscript* file that these tools read from in order to update the text in Paratext accordingly.

In the simplest case (which is all that Glyssen needs), the start quote milestone (`\qt-s*`) contains just the name of the character, like this:

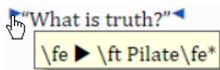
```
\qt-s |Pilate\q*"What is truth?"\qt-e\*
```

and the end quote milestone is empty. (`\qt-e*`)

Unfortunately, because Paratext currently displays the contents of all milestones inline, this can result in significant visual clutter. Until Paratext implements a less disruptive display option for milestones, one *temporary* alternative is to represent the milestone with a specially-formatted endnote, like this:

```
\fe ► \ft Pilate\fe*"What is truth?"\fe ◀ \fe*
```

which displays in Paratext with just blue triangles that can display the character ID when hovered-over like this:



Note that if you chose to employ this format, you'll probably need to filter these markers out of print/electronic publications with a changes rule that finds this regex. See the help file for details.

Inferring \wj from quote milestones on the fly at publishing time

PtxPrint and PublishingAssistant can automatically infer \wj markers from Jesus' quote milestones. Copy the contents of VM_PrintDraftChanges.txt (in the downloaded zip file) into a PrintDraftChanges.txt file in the Paratext project folder. (Publishing Assistant typesetters can do similarly into the job's Changes.txt file.)

For Scripture App Builder, which does not yet support such changes, you can temporarily add \wj markers into the project in Paratext, update the text in SAB, and then strip them out again.

Which should go innermost? Milestone or \wj?

USFM does not care whether the \qt-e* milestone or the \wj marker goes immediately adjacent to the words of Jesus, but if you're using the special endnote to represent a milestone, that cannot come inside the \wj ... \wj* run of text.

Tool Overview

The following tools comprise the voice-marking custom tools for Paratext:

- v. **Mark quote milestones (from Glyssen)**
 - o *Updates the scripture text with the milestones indicated by a glyssenscript file that the user provides.*
 - o *The user may select to have only the Jesus quote milestones in the text, or the milestones for all characters.*
 - o *The user may select either representation of milestones.*
- vi. **Convert quote milestone format**
 - o This simply converts one of the representations of quote milestones to the other.
- vii. **Mark words of Jesus with \wj (from Glyssen)**
 - o Updates the scripture text with \wj ... \wj* tags, based on a *glyssenscript* file that the user provides.
- viii. **Mark words of Jesus with \wj (from milestones)**
 - o Updates the scripture text with \wj ... \wj* tags, based on milestones already in the text.
- ix. **Remove all \wj marking**
 - o Removes all \wj marking, including un-nesting markers that had been nested within the \wj tags.
 - o Note: If you ever want to remove all USFM3 quote milestones, you can use RegExPal. See the help file for details.
- x. **Fix \wj marking**
 - o Wherever there's a pair of \wj ... \wj* markers, this fills in any required marker changes. For example, it closes and reopens \wj marking around verse numbers, section headings, chapter breaks, etc.; adds a plus sign (+) to nested style markers; etc.

More details on each tool are provided below.

Preparations in Glyssen

For the tools that read a glyssenscript file, this is how to prepare that file.

Tip: Using Glyssen can also help to identify quotation problems in Paratext that were not caught by the quotation checks, such as spoken verses that were not in quote marks. For this, use the “Verses with missing expected quotes” filter.

It’s strongly recommended that before using Glyssen, you ensure that the books you are going to use there pass the following basic checks: Chapter/verse numbers; Markers; Quotations.

If you have previously set up the project in Glyssen, go to the Glyssen project settings and click the Update button to get the latest version of the text from Paratext. Otherwise, set up your project in Glyssen by selecting the row with the language name and project code.

The main work that’s necessary in Glyssen is done from the “Identify speaking parts” tool. This has two views that you can switch between:

- “Rainbow view” (second button on the toolbar, “Match reference text for each colored row”)
- “Character view” (first button on the toolbar, “Select character. Who speaks this part?”)

As it doesn’t matter for our purposes whether or not you align the vernacular with a reference text (such as English), it will be simpler to use the Character view.

From the toolbar, you can select a filter to look at the problem areas. These filters will be particularly useful:

- **Unassigned quotes:** Every place that Glyssen has found quote marks that it can’t determine who was speaking.
- **Verses with missing expected quotes:** The translators may have forgotten to insert quote marks into the text. This will help you find such problems.
- **Needs review:** Issues that Glyssen is particularly uncertain about, such as ambiguity encountered when updating a project from Paratext.

For the selected verse segment in the left pane, select the correct character from the right pane, and click the “Assign Character” button. (If the correct character is not in the list, click “More Characters” and start typing the desired name to add more names to the list, then select it there. In that case, you may also need to select a Delivery method before the “Assign Character” button is enabled.)

Tip: You can double-click a character to make the assignment and automatically move to the next item. (If it doesn’t move to the next item, Glyssen probably also needs you to specify a Delivery option for that line).

If you only care about tagging the words of Jesus, those are the only assignments you need to make.

Once you’re satisfied that the important books for the words of Jesus have all of Jesus’ words identified, from the main Glyssen window, click the **View Recording Script** button, then click **Export** and **Export to HearThis...**

Click Browse and for a filename enter your Paratext project’s short name, such as XYZ. This should result in your file being exported to XYZ.glyssenscript in your **Documents/Glyssen** folder, where the Voice Marking Tools will look for it.

Installing these tools and using them in Paratext

- Copy the provided files into the CMS folder under your Paratext Projects folder.
E.g. C:\My Paratext 9 Projects\cms
- Restart Paratext.
- All of these tools can made widespread changes in your text. Before using any of these tools, be sure to mark a point in your project history, so that you can compare versions and roll back the changes if necessary.
- From the Project menu, select Custom Tools > Voice Marking, and then the specific tool.
- If you want to output to a new/different project, select it as the output.
- Select the books to apply this to.
- Verify that the options are correct.
- Click OK to run the tool.
- After running the tool, compare the before and after versions to ensure that you are satisfied with the changes.
- Each time you run one of these tools, it creates/replaces a log of its activity in the project folder in a file named **VoiceMarking.log**. If you get unexpected results, send this log file to Dan Em with a description of the problem.

Tool Details

Mark quote milestones (from Glyssen)

Adds milestone tags to the text to mark the start/end points assigned to various characters in Glyssen. This ensures that when the text round-trips back to Glyssen, the decisions previously made will be accurately reflected in the updated text. Alternatively, if Glyssen has only been used to mark the words of Jesus, or you'd rather not track the other voices in Paratext, you can opt to only have milestones for the words of Jesus.

If the Paratext project has been edited since the creation of the the Glyssen project, it's recommended that you first update the Glyssen project. (Glyssen project settings > Update.) Otherwise, this tool does fairly well at handling changes to the text.

Convert quote milestone format

Converts the format in which quote milestones are represented in this project, in either direction between two formats: 1. True USFM 3 `\qt-s*` and `\qt-e*` quote milestones. (Currently, Paratext displays these entirely inline in the text, so they are visually cluttered.) 2. Special endnotes like this: `\fe ► \ft Jesus\fe*“Go!”\fe ◀ \fe*` Choose the direction of conversion in the Options.

Mark words of Jesus with `\wj` (from Glyssen)

This Paratext custom tool adds `\wj` tags to the USFM text to mark the words of Jesus, based on the contents of a glyssenscript file exported from Glyssen.

Any existing `\wj` tags will be removed.

Word/character formatting found within the words of Jesus, such as `\w` or `\nd`, will be nested as `\+w` or `\+nd`.

In verses where this tool cannot be very sure of the tagging position, the List window will display these verses for you to manually verify the tagging, adjusting it if necessary.

Options

Filename of HearThis glyssenscript file exported from Glyssen

If you exported the file according the instructions above, leave this blank. Otherwise, specify the filename, including the full path if the file is not located in your Documents\Glyssen folder.

Minimum similarity ratio between Glyssen and Paratext

How closely must the text from Glyssen match the text from Paratext in order to apply the `\wj` tags? They may be slightly different if you've made minor edits in Paratext since setting up Glyssen. The similarity ratio is calculated based on the differences between the texts. A value of 1 means identical. 0.97 means very close. Do not specify this number as a percentage.

Recommended value: **0.97**

List all changes?

If you want a list of which characters were changed in which verses, set this to "Yes". If you prefer to leave the list as sparse as possible, to quickly notice warnings, set this to "No". In any case, the VoiceMarking.log file will contain the list of changes made.

Mark words of Jesus with `\wj` (from milestones)

Adds `\wj` tags to the text to mark the words of Jesus, based on milestone markers already present in your text. This means that you can normally edit the text without the clutter of `\wj` markers, just milestones, and then temporarily apply the `\wj` markers when publishing via software that can't interpret the words of Jesus milestones to display red letters. (The milestone markers could be either the standard USFM 3 `\qt-s`* and `\qt-e`* quote milestones, or special endnotes like this: `\fe ► \ft Jesus\fe* "Go!" \fe ◀ \fe*`) Any existing `\wj` tags will first be removed. Word/character formatting found within the words of Jesus, such as `\w` or `\nd`, will be nested as `\+w` or `\+nd`.

Remove all `\wj` marking

Clears all `\wj ... \wj*` markers from the selected books. Also removes the plus sign (+) from any character style markers (e.g. `\+w ... \+w*`) that had been nested within `\wj ... \wj*` tags.

Fix `\wj` marking

If your text already contains some `\wj ... \wj*` markers, this tool will close and reopen these markers wherever the words of Jesus are interrupted by verse numbers, paragraph markers, chapter breaks, section headings, footnotes, cross-references, figures, etc. It will also add a plus sign (+) to any character-level markers that become nested in the `\wj` text. (e.g. `\+w ... \+w*`)

If the words of Jesus are not already wrapped in `\wj ... \wj*` markers, rather than adding them manually in each location and then running this tool, it will be faster to use Glyssen to identify the words of Jesus (because it knows where to expect them, and only needs a little disambiguating) and then use the "Mark words of Jesus with `\wj` (from Glyssen)" tool instead of this one.

Versions:

2023-06-08: Updated to handle "Jesus (child)" in LUK 2:49.

^[1] <https://drive.google.com/drive/u/1/folders/1BVTFJeqORi83wPiBKONjWQ6EKWnEEUFm>

^[2] <http://ubscap.github.io/usfm/milestones/index.html>