April 17, 2023

To: Lev Israel, Chief Data Officer, Sefaria

From: David E. S. Stein, on behalf of the Jewish Publication Society

Re: Revised note on the formatting of the RJPS translation — USX files

This version incorporates several hundred content and formatting corrections, while fixing dozens of bugs in embedded cross-references. As in the original dataset from July, RJPS resembles the USX version of the updated portion of NJPS dated 1/27/21. Given that some of the encodings are useful only to the print medium (including two new codes at chapter openings), I trust that you will simply collapse those distinctions.

As with CJPS and NJPS, not all footnotes have a lemma. Some notes apply to what *follows* the footnote’s callout (e.g., the whole verse) rather than what *precedes*.

Unlike both CJPS and NJPS, the footnote callers are coded as absolute sequential letters of the alphabet. They restart with each chapter. In the two chapters (in Ezekiel) that have more than 26 notes, the sequence begins again with “a.”

We have embedded **links to all cross-references** to other passages within the *Tanakh*. That being said, the following are known bugs, with suggested fixes.

* To generate links to book-chapter-verse locations, ignore the link-href attributes of char tags. Instead use the loc attributes of ref tags.
* 74 loc attributes of ref tags have no verse, i.e. they are just book-chapter locations, e.g., <ref loc="NUM 32">. So, you’ll probably need to add “:1” (colon-one) to those, either literally or conceptually.
* 6 loc attributes of ref tags are “ab-qualified,” e.g., <ref loc="1KI 6:6a"> and <ref loc="1CH 20:1b">.  So, you’ll probably need to strip the “a” or “b” off of those.
* 362 loc attributes of ref tags are ranges, e.g., <ref loc="2KI 13:22-25">. So, you’ll probably need to drop the end point of those ranges.

Sefaria may also wish to embed links to rabbinic literature, as cited in the footnotes to 2Sam 21:8; 2Kgs 23:13; Eccles 10:4; Isa 1:28; 3:17.

Cross-references that are merely directional (“see previous note”) are used only if the target note occurs within the same verse. No link is embedded in such a case.

Last and probably least, among the assigned paragraph markers (styles) is “b”, which is used to represent a half-line space. You will occasionally find two in a row, which represents a full-line space. (This oddity seemed worth mentioning even though Sefaria’s formatting seems to ignore all line spaces.)

If you have questions, let me know; I’ll do what I can to help.

*Attachments*: Inventory of USX Markers; Style Sheet and Word List; Glossary, Abbrev, Translit.