

Scripture Forge Test on Upper Guinea Crioulo Revision Project

(June 29-July 4, 2024)

By: Jeff Shrum, CiT, Seed Company

Testing Scripture Forge Generated Drafts -Revision

While in Guinea-Bissau for a consulting trip for two other languages being worked on by *O Instituto de Tradução e Alfabetização* (ITA), I learned that the Crioulo project had already revised the New Testament and many Old Testament books. I became curious whether Scripture Forge could expedite their revision process. Our first step was to evaluate the quality of a Scripture Forge draft in Crioulo.

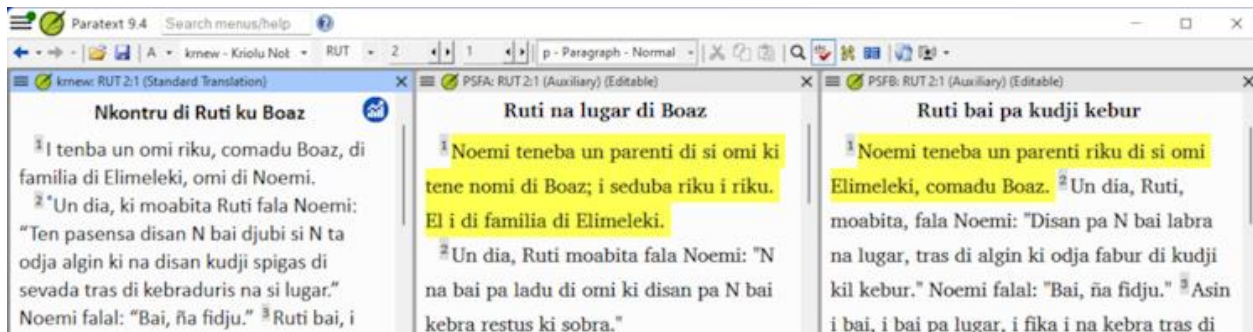
Comparing SF draft to existing consultant-checked books

To gauge the quality of Scripture Forge's output, we compared its drafts to books that had been consultant-checked. The team selected Ruth and Jonah for the comparison due to their manageable length and narrative style. Both books had been consultant-checked, and the team was well-acquainted with their content.

Working with Scripture Forge engineer Michael Martin, we produced two drafts for each book: one using the Nova Versão Internacional (**NVI-P**) as the model text and the other using the Almeida Revista Corrigida 1993 (**ARA93**). These drafts were imported into separate new Paratext projects for comparison against the team's existing revision project (**krnew**).

Evaluation Process

With all three projects open in Paratext, the team conducted a verse-by-verse comparison. In Chapter 1 of Ruth, they found that most verses were acceptable-to-excellent, though there were a few hallucinations and inconsistencies in names. Many verses generated by Scripture Forge were clear and natural, sometimes even better than the team's translation. Overall, the team felt they could easily correct the drafts if they had started with them. (See sample layout below)



As the evaluation progressed, the team was increasingly impressed with the quality of Scripture Forge's output. They began using the drafts as valuable resources to improve their translations of Ruth and Jonah. However, when introduced to a more dynamic Crioulo resource based on the Versão Fácil de Ler, the team found it too wordy and decided to stop using it.

Summary of Changes

After revising Ruth and Jonah with ideas from the Scripture Forge drafts, we used Paratext to track the number of changes made in each chapter:

Ruth -Verses Changed		
Chapter	Count	Percentage
1	10/22	45.5%
2	18/23	78.3%
3	9/18	50.0%
4	10/22	45.5%
Totals:	47/85	55.3%

Jonah -Verses Changed		
Chapter	Count	Percentage
1	11/17	64.7%
2	7/10	70.0%
3	5/10	50.0%
4	8/11	72.7%
Totals:	31/48	64.6%

The team felt that the revisions greatly improved their drafts and expressed interest in using this process for all their previously consultant-checked books. However, we informed them that such a change would require further negotiation.

Testing Scripture Forge Generated Drafts -As Draft Zero

Introduction

Our first test of Scripture Forge was very encouraging but took us in the unanticipated direction of using SF materials as *resources* and not as *drafts*. We decided that we needed a more quantitative test to judge how useful SF generated texts might be for a first draft or a draft zero. (The term draft zero is used by some if a draft is produced with a machine assisted process.) For this test we requested that drafts of Esther be generated one with the NVI-P as the model text and one with ARA93 as the model text just as before. Because of time we decided to only evaluate the NVI-P based draft.

Evaluation Process

The team scored each verse of Esther for clarity, naturalness, and accuracy, comparing them to their existing draft and the ARA93 and NVI-P Bibles. They were instructed not to edit the draft but

solely to evaluate the Scripture Forge-generated text. The evaluation took about one and a half days to evaluate the 167 verses in Esther, and the results are summarized below:

Issue	Count	Percent of total verses
Lack of Clarity	33	19.8%
Lack of Naturalness	16	09.6%
Lack of Accuracy	16	09.6%
Total with an issue	65	38.9%

**More detailed information can be found in the appendix.*

Conclusion

The team found that a Scripture Forge-generated draft would be very useful as a first draft. They were pleased with the results and would be happy to start from such a draft in the future.

Unknowns

Should it be decided that the Crioulo team will use Scripture Forge in the future there are some details to be worked out. It is possible to continue working with a Scripture Forge engineer, or to enable a Scripture Forge account to allow a user to generate drafts without help from Scripture Forge staff. This is probably the best course of action because it would free up Scripture Forge staff to work with other teams. Also, if a team member can generate the drafts, then it eliminates the problem of the time difference between USA and Guinea-Bissau.

The team benefited from having 2 or more SF resources. I am unclear if this can be accomplished by a user (someone not on SF staff) in the current user interface. I will need to experiment with using multiple model texts with the same base text to understand better the workflow for generating multiple draft of the same text.

Future workflow

Should it be decided to make Scripture Forge part of the Crioulo team's workflow, then resources for one book would be generated as needed for a book being revised. Once that book completes the consultant check stage, that book would be added to the POV2014¹ project and POV2014 would be analyzed by SF again to generate the set of resources for the next book to be revised. Working on one book at a time is the method recommended by the SF team, because it will refine the model each time and should produce drafts of increasing quality.

Conclusion

Using model texts (NVI-P and ARA93) with similar translation styles to the Crioulo project, Scripture Forge created valuable resources that helped translators improve their Bible translation. Resources with different styles were not helpful and slowed the team down. Generating multiple resources increased the chances of finding ideas to improve the text.

¹ See Methodology Section of this report for more details of the Paratext project POV2014.

We recommend using Scripture Forge to generate "resources" for all Crioulo books that have not yet been consultant-checked. This approach should greatly improve the quality of these books with minimal impact on the translation schedule. The team also requested Scripture Forge-generated resources for books that have been consultant-checked, though this would require further negotiation with all stakeholders due to its potential impact on the schedule.

Our second test, using Esther, showed that Scripture Forge can produce an acceptable first draft (draft zero) if conditions are right. With 39% of the verses having issues, most were small and easily fixed. If a draft is 60-70% or higher in quality, it is worth using as a first draft and editing. Few first drafts done traditionally exceed this level of quality, while a Scripture Forge draft bypasses the tedium and errors of manual typing.

Methodology

Several new projects were created. One standard translation project called POV2014 that was registered as a revision of the **pov** project/text. The purpose of the POV2014 project was to contain the books from the revision project **krnew** that were deemed of sufficient quality to be ingested by *Scripture Forge*. (This step also had the benefit of minimizing any disruption to the team's normal work.) The New Testament, Pentateuch, Ruth and Jonah were copied to POV2014.

Next four auxiliary projects of POV2014 were created into which four *Scripture Forge* generated drafts were placed. This was done initially to facilitate the comparison of the various drafts generated by *Scripture Forge*. The names of the four projects and their basic composition are as follows:

Short name	Model text	Base text	
PSFA	NVI-P	POV2014	
PSFB	ARA93	POV2014	
PSFC	RC69	POV2014	
VFLKA	VFdL	POV2014-NVI-P	
VFLKB	VFdL	POV2014-ARA93)	

I requested the drafts from the SF engineer, Michael Martin and he would send them to me the next afternoon as zipped files. I would import them into the appropriate projects. PSFA, PSFB, PSFC, VFLKA, and VFLKB were all configured as axillary projects to POV2014. This simplified how the projects were registered.

Acknowledgement

I would like to thank the people of ITA and especially the Crioulo Revision team for allowing us to perform this test. Without their help and participation this text would not have been possible. I would also like to than Michael Martin from the Scripture Forge team for his very timely help despite the time difference and that we were working on a US holiday.

Appendix

History/Background

Upper Guinea Creole or Crioulo (Kiriolo) as it is called locally is a creole language whose lexicon is largely derived from Portuguese, with grammar and additional lexicon from Balanta and other languages of the Casamance region. The first edition of the Crioulo Bible was published by the Evangelical Church of Guinea-Bissau in 1998; © Evangelical Church of Guinea-Bissau; © WEC International. The Evangelical Church of Guinea-Bissau has a translation and Literacy department called *O Instituto de Tradução e Alfabetização* (ITA).

Scripture Forge generate draft of Esther in Crioulo using NVI-P as model text

Esther Chapter 1:

Issue	Count	Percent of total verses
Clarity	3/22	13.6%
Naturalness	3/22	13.6%
Accuracy	1/22	04.5%
Total with issue	7/22	31.8%

Esther Chapter 2:

Issue	Count	Percent of total verses
Clarity	7/23	30.4%
Naturalness	0	00.0%
Accuracy	4/23	17.4%
Total with issue	11/23	47.8%

Esther Chapter 3:

Issue	Count	Percent of total verses
Clarity	2/15	13.3%
Naturalness	3/15	20.0%
Accuracy	0	00.0%
Total with issue	5/15	33.3%

Esther Chapter 4:

Issue	Count	Percent of total verses
Clarity	3/17	17.6%
Naturalness	2/17	11.8%
Accuracy	1/17	05.9%
Total with issue	6/17	35.3%

Esther Chapter 5:

Issue	Count	Percent of total verses
Clarity	1/14	07.1%
Naturalness	3/14	21.4%
Accuracy	4/14	28.6%
Total with issue	8/14	57.1%

Esther Chapter 6:

Issue	Count	Percent of total verses
Clarity	2/14	14.3%
Naturalness	0/14	00.0%
Accuracy	3/14	21.4%
Total with issue	5/14	35.7%

Esther Chapter 7:

Issue	Count	Percent of total verses
Clarity	1/10	10.0%
Naturalness	0/10	00.0%
Accuracy	2/10	20.0%
Total with issue	3/10	33.3%

Esther Chapter 8:

Issue	Count	Percent of total verses
Clarity	7/17	41.2%
Naturalness	1/17	05.9%
Accuracy	0/17	00.0%
Total with issue	8/17	47.1%

Esther Chapter 9:

Issue	Count	Percent of total verses
Clarity	5/32	15.6%
Naturalness	2/32	06.3%
Accuracy	1/32	03.1%
Total with issue	8/32	25.0%

Esther Chapter 10:

Issue	Count	Percent of total verses
Clarity	2/3	66.7%
Naturalness	0/3	00.0%
Accuracy	0/3	00.0%
Total with issue	2/3	66.7%