Adapting Digital Biblical Resources for Translation and Teaching:
Harvesting low-hanging fruit and growing next generation resources

Jonathan Robie
biblicalhumanities.org
jonathan.robie@biblicalhumanities.org

Abstract: When biblicalhumanities.org launched in 2013, we naively assumed that software developers would know how to take advantage of high quality, innovative data if it were published in open formats under free licenses. We understood the importance of building communities that know how to take advantage of this data, software and frameworks that make it easily accessible, and tailoring data to specific use cases, but we assumed that this would happen if we just made the data freely available. Freely licensed base texts, morphologies, contextual glosses, syntax treebanks, discourse analyses, lexicons, textual variants, images, conjectural emendations, and grammars are available now, and some are at least as good as commercial resources, but they are not yet widely used.

In the last year or so, Jupyter Notebooks and academic conferences have started to bring this data into academic study, but we have barely begun to tap its potential for translation, language learning, and scripture engagement for those who know the languages. To realize this potential, we need to understand the use cases and mindset of potential users and create software and other resources tailored to their needs. Over time, we need to gather feedback from these same users, walk with them as they learn and grow, and welcome them into communities that will create the next generation of resources.

Linguists, translators, and language teachers often face the same questions, but they generally face them at different levels of abstraction, in different contexts, and in different sequences and time frames. A linguist often needs to examine the entire corpus to identify all instances of a phenomenon. A language teacher needs ways to systematically teach the constructs of a language. A translator needs to understand what a particular word, phrase, clause, sentence, or passage means in order to render it into the target language.

In the last five years, a wide variety of freely licensed high-quality digital resources have been released for biblical languages, including base texts, morphologies, contextual glosses, syntax treebanks, discourse analyses, lexicons, textual variants, images, conjectural emendations, and grammars. Some of these are as good or better than the best commercial resources, others are not on the same level, and some important resources have no freely licensed equivalent. In addition, many important older resources have been digitized, including excellent technical commentaries. Some of this data was developed by Bible translation associations, including Greek syntax trees created by the Global Bible Initiative and discourse analysis done by SIL's Stephen Levinsohn. This data is beginning to be used in the research community, including multiple presentations at the Society for Biblical Literature and other academic conferences, and is being used to create materials for teaching biblical languages, but has not yet been adopted by the wider translation community.

These resources provide valuable information for translators and translation consultants who have some knowledge of biblical languages, but they need to be tailored to efficiently answer the kinds of questions they have as they work on a translation, when they have the questions, without forcing them to read through large amounts of abstract text in English or German, and provided in ways that work with the

tooling, resources, and workflows that they use. And when our resources do not efficiently answer their questions, we need to provide feedback mechanisms that capture their questions for the resource development community so that we can improve our resources and tooling.

Focused

Bible translation is focused, prayerful work. In the West, there is an increasing realization that our technologies and resources can distract us from the kind of attention we need to do focused work.[1] What an individual translator needs to focus on can vary, but a translator who is looking at a circumstantial participle does not need to know all of the details of the aspect debate, a concrete answer is often better than an answer that requires mastering new abstractions, and a translator who needs a community that can help answer questions does not need to be distracted by a stream of notifications.

The goals of a translation determine the focus of the translator. If a translator is translating from the original languages then it is important to efficiently describe vocabulary, syntax, and morphology, provide representative translations in languages the translator understands, discuss idioms and other constructs that affect translation, and identify pitfalls that can lead to mistranslations.

For instance, consider a translator who is trying to translate this phrase from Matthew 1:20 "for that which is conceived in her is from the Holy Spirit." If the translator is working from a single hub language translation, this is probably straightforward. If the translator is working from multiple hub language translations, all conveniently aligned to show how this phrase is translated, then some translation choices become evident, but the reasons for these choices may not be obvious without additional background.

Now suppose we introduce a reverse interlinear to provide access to Greek. Here is the same phrase as displayed in the Lexham Greek-English Interlinear in Logos:

τὸ	γὰρ	ἐν	αὐτῇ	γεννηθὲν	ἐκ	πνεύματός	ἐστιν	ἁγίου	. [21]
to	gar	en	autē	gennēthen	ek	pneumatos	estin	hagiou	
DNSN	CAZ	P	RP3DSF	VAPP-SNN	P	NGSN	VPAI3S	JGSN	
what₄	for₃	in₆	her₇	has been conceived₅	from₉	Spirit₁₁	is₈	the Holy₁₀	

This is much more complex than comparing English translations, and it is also more confusing than helpful – it provides information about each word, but little information about the relationships among words and phrases. In this phrase, the Greek word order is quite different from that of the English translations, and nothing in the interlinear makes it obvious that "Holy Spirit" is one constituent and "what has been conceived in her" is also one constituent. This display contains many pieces of information, but provides little help in bringing this information together to understand the structure of the sentence.

---

[1] Newport, Cal. *Deep work: Rules for focused success in a distracted world*. Hachette UK, 2016.

Here is the same phrase as displayed in the Cascadia Syntax Trees in Logos:
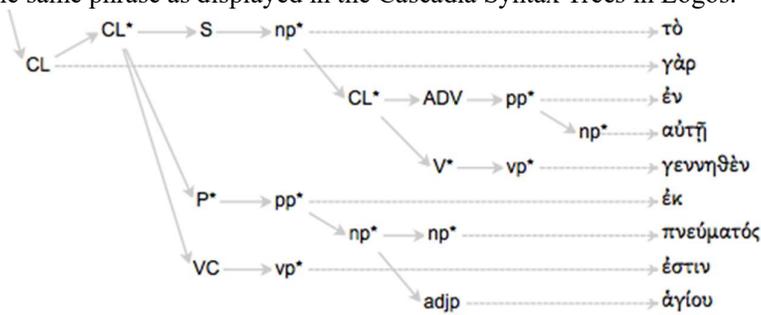
The Cascadia display does identify the constituents, but only if the translator has learned to read the crossing lines in the display, and it also contains nodes that are of more interest to a parser than to a human being trying to read this text.

There are, of course, many other ways that the syntax of this phrase can be visualized, each has advantages and disadvantages.2 Here is one way of displaying the structure of this sentence that makes the relationships among words and word groups clear with little additional overhead:

τὸ γὰρ ἐν αὐτῇ γεννηθὲν ἐκ Πνεύματός ἐστιν Ἁγίου·
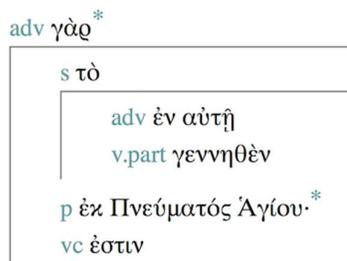
Contextualized glosses can be used for an interlinear display of the same Greek phrase:
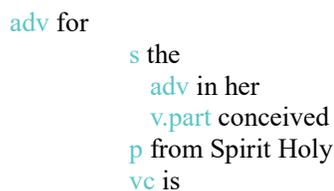
A good translation should maintain the relationship between the subject and predicate in the above display. What is from the Holy Spirit? The subject answers that question: "the in her conceived", which might be rendered as "that which is conceived in her", "what is conceived in her", or in a freer translation, perhaps "the child within her". Where is this child from? The predicate answers that question: "from the Holy Spirit".

---

2 The blog post "Nine Kinds of Ancient Greek Treebanks" discusses the relative advantages of nine different approaches. See http://jonathanrobie.biblicalhumanities.org/blog/2017/12/20/treebanks-for-ancient-greek/

This syntactic display is much less busy than the interlinear display, but provides much more information about the basic relationships among words and word groups in the phrase. It could be used by a translator with some knowledge of Greek to understand a difficult phrase, or by a translation consultant who wants to explain this structure to a translator with little or no knowledge of Greek. If a phrase can be interpreted in more than one way, a display of each corresponding structure can illustrate the different interpretations much more efficiently than pages of text in theological English or German, and with much less distraction. The translator does not need to understand the differences between cognitive linguistics, dependency grammars, and constituency grammars to grasp this structure, and there is no need to read the literature on hyperbaton to understand the basic meaning behind this particular word order well enough to translate accurately.3

Now let's return to the topic of focus. A translator who is working from a single hub language translation will not face the questions that this representation answers. For this passage, a translator working from several hub language translations may not face these questions either, though other passages might well raise questions that this representation answers. But if a translator or a translation consultant is working from biblical languages or even referring to biblical languages, these kinds of questions will come up, and we need to find efficient ways to provide the answers without overly disrupting the prayerful focus needed for good translation work. An interlinear provides a great deal of information about individual words, but very little information about the structure of the sentence as a whole, and this is particularly important for anyone who wants to understand the sentence. The representations we choose may display only a fraction of the information available in our linguistic representations. To keep translators focused on the data they actually need, we often may want to omit data that we would otherwise need for research or for creating optimal displays, but that are less relevant to translators.

The goal is to answer questions as they come up as simply as we can, but fully enough to understand and translate the sentence accurately. We want to provide information without disrupting the focus needed to translate. To do this well, we need to understand what information translators need at their fingertips as they do their work.

Concrete

"Every layer of abstraction loses 90% of the audience".4 That means we should avoid introducing abstractions that are not needed to convey what needs to be said, especially when dealing with an audience that is not used to academic abstractions.

Academic Westerners tend to imagine that translation requires advanced education and mastery of complex abstractions, and we may be alarmed to see translators with the equivalent of a high school or even junior high school education.5 In practice, mother tongue mastery of the target language is the most important qualification. Anyone who works with refugees regularly sees high school and junior high school students who are quite adept at translating for their parents. These children probably do not know the abstract terms used to describe languages in universities, but they understand the relevant

---

3 This passage is an example of hyperbaton in which a phrase is interrupted by a word that is not part of the phrase. Experts disagree on the significance of this kind of hyperbaton, and understanding the various possible explanations is not necessary for translating the phrase.

4 I originally learned this as "Bosworth's Law" when Microsoft's Adam Bosworth used it in a presentation. I do not know if it is original with him.

5 This is not meant to imply that all mother tongue translators have little formal education. Some have a great deal of formal education. We should avoid stereotypes while providing for the needs of a wide variety of translators.

languages and learn how to translate as needed. We see the same phenomenon for biblical languages. There are people who have mastered the vocabulary of linguistics, but are not particularly adept at reading and understanding biblical languages. There are other people who have become adept at reading biblical languages, but have difficulty reading the metalanguage used to describe them. Some people are comfortable in both the language and the metalanguage. This means that materials created for translators, translation consultants, and linguists may be significantly different even when conveying the same material – and even people who have similar roles may prefer different kinds of materials due to education levels, cultural factors, or personal preference. We may need to produce different kinds of materials from the same base data.

Maturing

Both translations and people mature over time. The first translation into a new language may be done from a single hub language translation and a set of translator's notes with no knowledge of biblical languages, and the emphasis may be on producing a translation relatively quickly where no translation currently exists. This translation may be inferior to the translations available to English speakers, and it may be the only translation available for the language. Ideally, this translation will be revised and improved over time, and perhaps more than one translation will eventually be produced (though this is sadly rare for most languages).

Similarly, a translator may mature as a translator, begin to understand why translations differ, learn biblical languages, and eventually become a qualified translation consultant. We need to raise up Christians throughout the world so that Christians from all nations work together as equals. For instance, we might introduce biblical languages using intensive living language courses6, or perhaps online learning7 or programmed materials8 could help make biblical languages more accessible. We could explore establishing centers of excellence in biblical languages in various locations that are strategic for bible translation. For translators new to working with biblical languages, we could produce translation resources and study materials that provide scaffolding for less experienced readers and online forums for translators to discuss their questions and learn from each other.9 We should envision growth paths that translators and communities can follow if they choose to, removing barriers when we can.

The resources we provide over time must also mature over time, as we learn what is most effective for our target audience. To do that well, we need to capture the questions they have and learn about the strengths and shortcomings of our materials as they perceive them.

Excellent

Our translations and the resources we provide to translators should both strive for excellence. Even if the first translation in a given language cannot yet be described as excellent, we should plan for excellence over time. This is equally true of the resources we develop for translators.

---

[6] Paul Nitz's Youtube channel shows how he teaches Greek in Malawi using a living language approach. Randall Buth and Christophe Ricco each teach biblical languages as living languages and produce useful materials for teachers. Micheal Palmer and Jonathan Robie have been adapting these techniques for a text-driven approach that also draws from SIOP.

[7] Such as BibleMesh Greek and Hebrew courses or Nicolai Winther-Nielsen's Moodle courses for Hebrew.

[8] I have recently become aware of John Werner's "Greek: a programmed primer", which may be useful for this purpose.

[9] These same materials may be useful for biblical language learners all over the world. Seminary education is not accessible to all, and Western seminaries and universities are generally reducing their offerings in biblical languages.

Open data resources are extremely important for providing access and removing barriers to research, but they are only useful if they are also good, which generally requires testing, curation, error reporting, and engagement with users over time.

At this point, the data for some freely licensed resources are at least as good as commercially available resources, including these:

- MorphGNT morphology10
- GBI11, Lowfat12, PROIEL13, Shebanq14, and Text Fabric15 syntax trees
- Stephen Levinsohn's *Greek New Testament Discourse Features16*
- Alan Bunning's textual variants17

These resources can be leveraged to provide useful new resources for translators – resources that are currently unavailable to most translators.

But many important existing resources are not openly licensed and the development of some important new resources are funded in ways that do not permit open licenses. For instance, the Greek texts used for translation and the best hub language translations in most languages are not freely licensed, neither are the best translator's notes and lexicons. Translators must have access to the best resources available, and the Digital Biblical Library18 provides authorized and licensed access to these resources for translators worldwide. The goal of open data is not to restrict access to the best information, even if it is not freely licensed. In the secular software world, open source has established itself by providing truly excellent solutions that can compete with commercial offerings, not by restricting access to commercial software, and the two are frequently used together in the same environments. The same is likely to be true for open source and open data in the translation world.

---

[10] https://github.com/morphgnt/sblgnt

[11] https://github.com/biblicalhumanities/greek-new-testament/tree/master/syntax-trees, in the sblgnt and nestle1904 subdirectories.

[12] https://github.com/biblicalhumanities/greek-new-testament/tree/master/syntax-trees, in the sblgnt-lowfat and nestle1904-lowfat subdirectories.

[13] https://github.com/proiel

[14] https://shebanq.ancient-data.org/

[15] https://github.com/Dans-labs/text-fabric/wiki

[16] https://github.com/biblicalhumanities/levinsohn

[17] http://greekcntr.org

[18] https://thedigitalbiblelibrary.org/about/

Shared

The software and data that we share is an important part of our community. In the translation community, ParaTExt is not only the vehicle for sharing resources, it also shapes the way people work, manage their projects, measure consistency and quality using built-in tools, communicate, and publish their results. It provides a common frame of reference at the heart of a community.

The open resources discussed earlier are at the heart of another community that provides freely licensed data on GitHub, building on each other's work. Until recently, people in this small community either used queries or programs without GUIs or built their own, and it was difficult to become part of this community without significant knowledge of both programming and biblical languages. Over the last year or so, the Eep Taalstra Center began using Jupyter Notebooks[19] to provide this kind of data in a loosely coupled GUI environment that can be used with less programming background, though it still requires programming or queries, and a handful of researchers have presented results of this approach for Hebrew linguistics at several academic conferences. A Jupyter Notebook is a document that can contain live code or queries and display results, and a module has also been written to support queries for syntax and morphology with biblical Greek.[20] This kind of environment is particularly well suited for data integration using various resources that may have been developed separately. The open data community uses a variety of tools, but we know each other, participate on each other's Slack channels, use each other's data, learn from each other, and see each other at conferences. This community is well-placed to create new resources and curate data over time, and we have tools that are useful for prototyping displays that might be used in software, but the tools we use are probably beyond the average non-programmer.

Using data from the open data community in translation tools will benefit both communities, but we should also consider sharing with a third community: non-programmers who want to work in the original languages but are not translators. This is a larger community than the first two, and a community that the other two generally know and want to share with. We need a GUI-based tool that can provide the open resources we develop for translators to this wider community. One possibility would be to have a free version of Scripture Online Tools that provides only free resources. Another possibility would be to create a website. Either approach would benefit the majority world by giving free access to mature resources for understanding biblical languages, and it would benefit open data developers by providing a vehicle for sharing our work more easily with serious students of the Bible. And our communities overlap - the same person may be translator, a translation consultant, a pastor, a language teacher, or a scripture engagement expert, or several people may each play one or two of these roles in their community. A translator may ask questions in an online forum or suggest corrections in data created by a researcher, or create datasets that can augment existing datasets. Communication among these overlapping communities is much easier if they have access to the same materials – at least those that are openly licensed.

---

[19] https://jupyter.org/

[20] See http://jonathanrobie.biblicalhumanities.org/assets/greeksyntax-tutorial.html, a Jupyter Notebook tutorial which is itself a Jupyter Notebook.

Unfortunately, there are some obstacles that need to be considered when sharing resources. For instance:

- The NA28 Greek text is not openly licensed, the SBLGNT is not freely enough licensed to be used in some settings, and freely licensed texts are either old enough to be unfashionable (e.g. Nestle 1904, Tischendorf, Antoniades 1904), use a text that most translation societies do not prefer (e.g. the Robertson-Pierpoint Byzantine text), or are so new that we do not know how they will be received (e.g. the Bunning Heuristic Prototype). Tyndale House has indicated a desire to make the Tyndale House Greek New Testament open for non-commercial use, but it is not clear whether this will be done in a way that does not require each project to negotiate separately. Currently, open resources use a variety of Greek texts. We need a strategy for efficiently porting resources to multiple texts, and it would be helpful if most open resource providers used the same texts.
- We need standards for reference systems, referring to lexemes, and describing morphology. The variety of systems used to describe the same things makes queries that span data sources unnecessarily complex.
- A few additional open resources would significantly improve the usefulness of the existing ones. For Greek, a modern lexicon would be particularly helpful. We also need to expand our support for the Septuagint.
- Creating open resources for biblical studies is not currently a major priority for academic institutions or translation associations, and it requires labor. We need to find ways to fund this work.

We need to find appropriate venues for deciding how to best work together to avoid duplicate effort and make our resources as interoperable as possible. This will benefit all three communities mentioned in this section.

Conclusion

As seminaries and universities cut back on their offerings of biblical and classical languages, passionate communities are working hard to develop new resources and make them freely available. These resources are valuable for translators and for serious Bible students in the majority world, and they can help us gain a deeper understanding of biblical languages. We need to find ways to make our resources more interoperable and reduce duplication of effort, tailor our resources to support efficient and informed translation, and create tools that can make these resources available to a wider audience of serious Bible students in the majority world and elsewhere.

The initial steps may be as simple as finding ways to visualize these resources and make them available in translation software, but the long term impact is much greater if we see this as a community building exercise that strengthens the translation community, the open data community, and the global biblical studies community. In everything we do, we need to set clear goals, strive for excellence, and plan for everything we do to mature over time as we learn from each other in the communities we establish.

Over time, we hope this will make it easier for people throughout the world to learn biblical languages and study the original biblical texts using high-quality, freely licensed resources, and to establish an open global community of translators and scholars who can learn from each other and build on each other's work.