

The AGNT Project Report—Q2 2011

As a licensee or friend of AGNT or ANLEX, we would like to update you once a quarter about our continuing work to enhance and perfect these databases and about our plans for the future.



The Project. *The AGNT Project Report—Q3 2008* introduced the team, outlined ongoing tasks, and discussed potential tasks.

NOTICE. We are happy to announce that our relatively new BYZ-AGNT database has had its beta-version status lifted. It is now simply BYZ-AGNT, ready for any of our licensed vendors to use in their databases. We will be adding other enhancements to our AGNT project in the coming period. Stay tuned.



AGNT in BibleTrans

Tom Pittman, PhD

BibleTrans uses computer technology to greatly speed up the process of getting first-draft translations of the Bible into the approximately 2000 languages that still today have no part of the Bible in their language.

Translation of any kind is a difficult process, consisting of two parts: first, clearly understanding the original message in the original language, then saying the same message in another language. With Bible translation, the first part is particularly difficult, because the original languages (Greek and Hebrew) have been dead for nearly 2,000 years. Scholars understand these languages, but there are not many such scholars. Computers do not *understand* anything; they can only follow carefully programmed sequences of instructions. BibleTrans depends on human understanding of the AGNT, so that like the five loaves and two fishes, it can be multiplied to do the other part of translation, hopefully into a thousand or more languages.

The second part, or “back end” of translation is more amenable to automation. There are perhaps 10,000 different distinct *concepts* (things and actions and their attributes) in the original Greek and Hebrew text. In BibleTrans, these concepts are distinctly numbered (computers work well with numbers), and linked together in a precise but language-neutral semantic representation of the original meaning. Then for each language into which the Bible is to be translated, a set of rules can be given to BibleTrans, how to say each of those concepts in that language, and how to form them into sentences and discourse. The computer then follows those rules in rendering into human language the hundreds or thousands of times each concept occurs in the biblical text.

The result is not church-ready, but it reads somewhat better than the instructions that sometimes come with a Korean or Chinese appliance. Most importantly, the translation is semantically accurate. Mother-tongue speakers can, without knowing Greek or Hebrew or even English, improve the wording to make it sound less wooden.

Word substitution is not translation. Translation requires understanding the source text. Now, and for the foreseeable future, only humans do this well. In BibleTrans, carefully trained translation specialists read the AGNT and build from it the semantic representation of the text. Some concepts, such as names of people and places, carry across directly. In most cases, however, a Greek word may have several different senses, represented by different Louw & Nida concept numbers, and the BibleTrans expert building the database must evaluate which concept to choose. Greek nouns, and especially verbs, are highly inflected to give nuances of meaning. These variations in meaning also have their own BibleTrans concept numbers, which contribute to accurate translation in the back end.

But lexical meanings and word inflection are insufficient to capture the full meaning of the text as required for some third-world languages. Greek nouns can be singular or plural, but many languages also have separate syntax for dual number, such as a pair of eyes or hands. Hebrew is one of those languages, and sometimes an English translation of the Hebrew dual suffix explicitly says “two” and sometimes not. When translating from Greek into a language like Hebrew, the translator must know whether there are two hands or people involved or more than two. Usually that can be inferred from the context, but computers cannot do that, it must be encoded by hand in the semantic database.

Similarly, the Greek verb is fairly complete in representing action and mood, but time is limited to past, present, and future, whereas some languages distinguish near and far past or near and far future. Worse, different languages draw the line differently. In the Papua New Guinea language Awa, for example, near past means “earlier today” and far past means “more than a few weeks ago,” while simple past covers the intermediate time. BibleTrans encodes a finer granularity than most languages need, so that the back-end translation rules can correctly choose the correct syntax in each case without further manual effort.

Western languages such as Greek and English communicate many discourse relations only by context, but other languages have distinct syntax for different kinds of discourse. The BibleTrans database encodes the most complete representation in each situation. Sentences are not strung together like beads on a string, as they are in Greek or English, but every proposition is connected to its context by discourse concepts that specify whether the connection is sequential or inferential or causal and so on. Greek, more than modern languages such as English, has verbal substantives (participles and infinitives), which in some other languages must be translated as dependent clauses. In BibleTrans, the Greek verbs are fully propositionalized. For example, the Greek participle in “*πας ο πιστευων εις αυτον*” in John 3:16 is normally translated into English as a dependent clause, “everyone who believes in him . . .” English has participles, but we do not use them that way, so attempting to carry the participle straight across would not make sense to the average reader. On the other hand, the Greek abstract noun “*αγαπη*” in 1Corinthians 13 translates straight across to the English abstract noun “love,” but some other languages—thinking again of PNG Awa—have no abstract nouns at all, so the Greek noun must be converted to a dependent clause: “When somebody loves somebody,” which is the way the database is encoded. All of this requires substantial human judgment, just as when translating the Bible into those languages in the traditional manual manner.

Let's see how this works with a familiar verse, John 3:16. Here is a portion of the Greek verse as displayed in BibleTrans. The text can optionally be displayed with Friberg parse codes (here omitted), English gloss, and the Louw & Nida concept tags. Originally, the text has no L&N tags; adding them to the text is the first task of encoding. Sometimes these tags can be automatically inferred from specific examples in the Louw & Nida lexicon, or because there is only one sense to a Greek word, but ultimately it is another judgment call.

(Greek) John 3:16																
<input type="checkbox"/> Parse	<input type="checkbox"/> Lemma	<input checked="" type="checkbox"/> Gloss	<input checked="" type="checkbox"/> L&N													
3.16	Ὅπως	γάρ	ἠγάπησεν	ο	θεός	τὸν	κόσμον,	ἵνα	τὸν	ὑἱὸν	τὸν	μονογενῆ	ἔδωκεν,	ἵνα	πᾶς	
	THUS	FOR	LOVED	-	GOD	THE	WORLD	THAT	THE	SON	THE	UNIQUE ONE	HE GAVE	THAT	EVERYONE	
	§78.4	§91.1	§25.43		§20.24	§12.1	§92.24	§9.23	§89.52	§92.24	§10.42	§92.24	§58.52	§57.71	§89.59	§59.23

The Greek text must then be decomposed into its constituent propositions. There are five propositions in John 3:16, one of them the Greek participle previously mentioned:

1. God loved the world
2. God gave the unique Son
3. [Some people] believe in (trust) the Son
4. All [those believers] do not perish (are not destroyed)
5. ... but [they all] have eternal life (live eternally).

There are additional considerations in encoding this verse. The primary sense of the Greek "κόσμος" is the created universe, Louw & Nida concept 1.1. Perhaps that is what the inspired evangelist intended, be we mostly understand it in this verse to mean that God loved the people of the world (L&N concept 9.23).

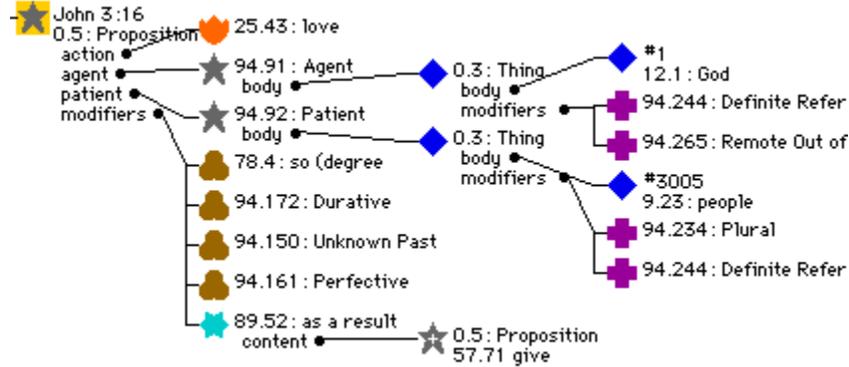
The Greek text has God giving *the* Son, but in many languages (including English), familial relations are inalienable, that is, you cannot speak of a "son" without saying whose son he is. Every English translator inserts the pronoun "his" into this verse to satisfy this language requirement, even though it is not in the Greek text. Body parts are inalienable in many languages, but not English. The BibleTrans database includes possessors for all familial relations and all body parts, but marks them as "implied" so that the translation grammar is free to ignore them when they are not required in the receptor language.

The Greek word rendered in many English translations as "perish" is actually the middle form of the Greek word "απολλυμι" (destroy). Voice is completely normalized out of the BibleTrans encoding, and reconstructed in the back-end translation grammar for those clauses where the subject is not specified (as here). That way, languages with no such thing as passive voice can be readily translated from the same database as English. Similarly, the Greek abstract noun "ζωην αιωνιον," which the believers in this verse have, is encoded as [those people] "live eternally."

Every language does its pronouns differently, so the BibleTrans semantic database has no pronouns, but the back-end grammar is responsible for reconstructing them according to the rules of that language. This is facilitated by encoding reference numbers for each "thing" (noun), so that where two clauses refer to the same person or object, the reference numbers are the same. Language grammars support pronouns by recognizing the common reference numbers, and some of the BibleTrans translation tools facilitate that determination.

BibleTrans uses graphical tools to construct the semantic representation, which is structured somewhat like the parse tree they used to teach in grade school. However, the full

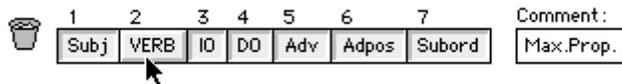
representation of even one verse like this would cover more than a whole computer screen, so the exegete or translator can close off nodes in the graph—like closing out bullet points in an outlining tool. Here is a very small part of the finished John 3:16 graph; you can download the demo program (see link below) and expand the whole tree on your own computer:



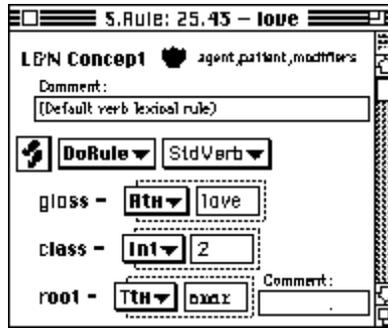
Encoding each verse like this can take a day or more. BibleTrans has tools to make dragging concepts into place relatively painless, but the analysis behind the construction cannot be automated. Once the entire New Testament has been encoded, producing new translations out the back end goes much more quickly—perhaps three to six months for a trained linguist to encode a generative grammar for one language, and then a few days to translate the whole Bible.

The linguist using BibleTrans to produce a translation into a new language needs to do several things. First, somebody needs to control the language. I have translated somewhere between one and eight verses into five languages (other than English all were non-European: two from Africa, two from Pacific islands). I do not know all those other languages, but I sat down with linguists who did. In one or two weeks, we had a grammar that translated one verse; additional verses usually involve little more than adding lexical entries and a few additional rules for forming the discourse structures that don't appear in the first verse.

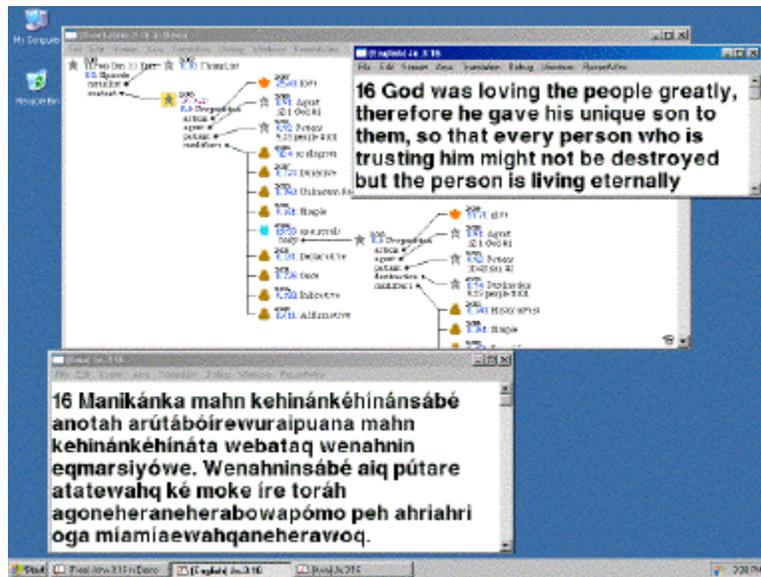
Most of the work is in defining how to construct a well-formed noun phrase and sentential clause. Some languages (like English) have a basic Subject-Verb-Object (SVO) order, others (Awa) OSV, or some other. There are six different possible orderings, but I think only four occur in real languages. The linguist can arrange the parts by dragging them on a line like this:



Verb, noun, and pronoun inflections are most easily encoded in tables by case, number, and gender (or whatever makes sense in that language). Complicated rules can be built up by cascading conditionals, if this situation holds, then do this, otherwise do that, and so on. There will be vast numbers of lexical rules (there are over 7,000 concepts in the Louw & Nida lexicon), so these can be facilitated by defining rule forms, so entry can go quickly—this screen shot is from an earlier version of the software, but the principle is the same:



Ordinarily, building the grammar and lexicon will be done incrementally, entering it for a few verses at a time, and then checking the resulting translation for grammar errors. Here is John 3:16, translated by BibleTrans into English, and then into PNG Awa:



Note that both of these translations are made from the same semantic database within seconds of each other; only the generative grammar is different. Awa is a particularly interesting language, because of the discourse syntax. The verb in each dependent clause is inflected also for the subject of the following clause (called a “switch reference”), which made for a rather complex generative grammar. The translator who actually did the Awa New Testament (manually) was quite pleased with the BibleTrans output quality.

The software works. You can download a working demo (the same program that produced the screen shot above) from the BibleTrans.info website. Because the AGNT Greek text and the Louw & Nida lexicon are licensed (I must pay royalties to some of the copyright owners), the downloadable demo contains only public domain resources; interested parties should [contact me](#) for the licensed version.

Linguists and experienced Bible translators know that machine translation does not produce acceptable translations—and they are right! Except BibleTrans does not do it that way. We need trained people to do the hard part, the part machines cannot do at all, let alone well. Through the generous support of Norman and Sarah Larsen, I was able to pay a competent

Greek student to encode the entire book of Philippians, but it still has problems that need correcting.

We are still looking for funding and/or qualified volunteers to encode the whole New Testament. If a dozen active Bible translators each took one year off their current projects today to build the BibleTrans database, they could go back to their projects three years ahead of where they would have been if they had stayed, and at the same time they would take 1,000 years off Bible translation worldwide. Pray the Lord of the harvest . . .

Tom Pittman, PhD

TPittman@BibleTrans.info

June 2, 2011



As always, we remain open to developing AGNT and ANLEX in ways that are most useful to the needs of students and readers of God's Word.

Thank you for your continued support of *The AGNT Project*, for faithfully marketing the AGNT and ANLEX databases, and for making these state-of-the-art tools for studying the Greek New Testament available to students, scholars, pastors, translators, and laypeople worldwide.

John Hughes

Agent for *The AGNT Project*

johnhughes@centurytel.net

Phone: 406.862.7289

FAX: 406.862.0917

