

# Computers in the Translation Process: Getting the most out of IT Resources for Translations

Robert P. Batzinger

## 1. Introduction

In the past two decades, translators and computer-support staff have learned many lessons and techniques that have affected the impact of this technology on the process and the products of Bible translation. Word processors have helped to shorten the time and effort required to prepare and publish manuscripts. Text processing utilities have facilitated analysis and cross-checking of the translated text in ways that were not feasible using only manual methods. The same technology that allows local newspaper companies publish an edition half the size of a New Testaments each morning<sup>1)</sup> is now affordable by Bible translation project teams.

Despite all the advances in hardware and software, the actual results of any given translation project team depend mainly on the expectations, collaboration, diligence and creativity shared between the translator, the computer operator and the copyeditor. While support staff can attempt to teach new skills, effective application of technology seems to occur best when the relevance and value of the computer-based tools are appreciated by members of the team and are exploited for the benefit of the project. At the same time

---

1) The weekday Singapore Straits Times morning edition (excluding the classified section) averages 1.5 MBytes of text printed on 50 pages of newsprint, containing illustrations and photographs that account for approximately 30% of the page. This volume of material and graphics is similar to that found in Gospels and Acts of the Good News Colour Reference Bible (which represents half of the NT).

it has been noted that the enhancement in speed of production provided by the technology comes with the increased risk of publishing mistakes or lack-luster translations.

In short, the classic and excellent advice to budding authors in the day of pencil and paper is just as relevant today to translation teams in the modern Cyber Age:

“Things written with little effort are read with little delight”<sup>2)</sup>

## 2. Enhanced Access to the Biblical Material

Knowledge of Biblical languages has always been a major asset to translation teams. While the computer technology cannot replace the value of personal and professional study of these languages, it can enhance appreciation of the Biblical text by presenting information in a way that is easier to analyze and understand particularly in the following ways:

- **Search for specific words in Hebrew and Greek text bases:** In recent years, computer systems have begun to adopt Unicode, a standard character encoding that support major writing systems of the world. This has made it possible for software to search the Biblical text in the native scripts of Asia in much the same way one would search for English or Korean. Table 1 displays but a few of the 64 Biblical occurrences of the Greek verb βαπτίζω
- **Parsing of word forms into morphemes:** The Biblical languages affix morphemes to root words to modify the meaning or sense of a word. While students of Hebrew and Greek struggle to learn how to parse the Biblical text for themselves, Biblical scholars have parsed the entire corpus.<sup>3)</sup> The computer makes it possible for the electronic text bases of the Biblical text to be linked this parsing information via a single mouse click providing expert assistance with the interpretation and translation of the Biblical text. The parsing of various forms of the Greek work βαπτίζω shown

---

2) A quotation attributed to Mark Twain.

3) The two most popular parsing attempts have been published by Gramcord and Baker. The electronic databases are available in numerous electronic study Bible products.

in Table 1 comes from the translator's editor called Paratext.<sup>4)</sup>

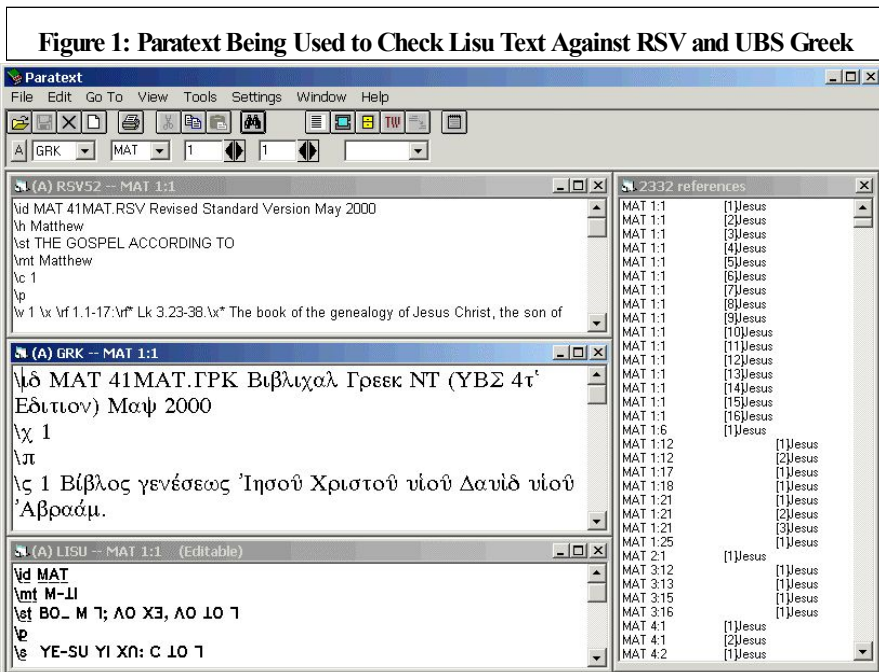
**Table 1: Resources Associated with Occurrences of the Greek Word βαπτίζω**

Ref	Concordance View		
Luk 3:12	ἦλθον δὲ καὶ τελῶναι	βαπτισθῆναι (verb, indicative, future, active, third person, singular) <i>Be baptized</i>	καὶ εἶπαν πρὸς αὐτόν, Διδάσκαλε, τί ποιήσωμεν;
Luk 3:16	ἀπεκρίνατο λέγων πᾶσιν ὁ Ἰωάννης, Ἐγὼ μὲν ὕδατι	βαπτίζω (verb, participle, aorist, passive, nominative, masculine, plural) <i>Baptize</i>	ὑμᾶς· ἔρχεται δὲ ὁ ἰσχυρότερός μου,
Luk 7:29	Καὶ πᾶς ὁ λαὸς ἀκούσας καὶ οἱ τελῶναι ἐδικαίωσαν τὸν θεόν	βαπτισθέντες (verb, participle, aorist, passive, nominative, masculine, plural) <i>Having been baptized</i>	τὸ βάπτισμα Ἰωάννου
Luk 7:30	οἱ δὲ Φαρισαῖοι καὶ οἱ νομικοὶ τὴν βουλὴν τοῦ θεοῦ ἠθέτησαν εἰς ἑαυτοὺς μὴ	βαπτισθέντες (verb, participle, aorist, passive, nominative, masculine, plural) <i>Have been baptized</i>	ὑπ' αὐτοῦ.
Luk 11:38	ὁ δὲ Φαρισαῖος ἰδὼν ἐθαύμασεν ὅτι οὐ πρῶτον	ἐβαπτίσθη (verb, indicative, aorist, passive, third person, singular) <i>Wash</i>	πρὸ τοῦ ἀρίστου.
Luk 12:50	βάπτισμα δὲ ἔχω	βαπτισθῆναι, (verb, infinitive, aorist,	καὶ πῶς συνέχομαι

4) Paratext is a translations tool originally written by Renier du Blois and further developed and made available to Bible translators by the United Bible Societies. It provides multiple windows with synchronized scrolling allowing the translator to edit the target text while comparing it against model Bible translations. For further information, please refer to the main support website for this product: <http://www.ubs-translations.org>

		passive) <i>To be baptized with</i>	ἕως ὄτου τελεσθῆ
--	--	--	---------------------

- **Cross-indexing to lexicons and other reference materials:** In the days before computers, it was typical to find the writing desk of translators cluttered with nearly a dozen Bibles commentaries and other Biblical reference text opened to the verse under consideration. As the translator reviewed this material a draft of the translation was penned into his workbooks. Modern computer software for translators mimics this work environment by linking text bases and other information together in a manner that can increase the efficiency and effectiveness of the translator.



Paratext shown in Figure 1 has become a very popular working environment for translators and copy-editors of Scripture. It not only supports synchronized scrolling of multiple windows of translations, it also has the ability to link to some commentaries and the UBS Translator Handbook series in the same fashion.

- **Concordant word views:** The various senses, nuances and shades of meaning ascribed to a word are most commonly derived by careful

comparison of the context for each occurrence of a word. For example, as shown in Table 1, the Greek word βαπτίζω can mean either to wash or to baptize depending on the context. The computer can render concordant views that span the entire corpus of the Biblical text base within seconds. Armed with this information and the ability to cross index against model and target text, a translator can easily verify that the proper nuance has been accurately preserved in the target translation.

### 3. Critical Analysis of Model and Target text

- **Comparison of similarities between translations:** When there are multiple attempts to translate the same material, it is possible to map the similarities between the approaches taken. The most common means to do this is to measure the degree of similarity between the different versions that exist by separating the words of both translations into 3 lines: those representing words unique to the first version, those common to both translation and finally those unique to the second translation. An example of a verse-wise comparison between RSV and GNT are given in Figure 2. This representation displays the degree of similarity. A numerical value of the similarity or difference, which can be calculated by the following formulae:

$$\text{Similarity} = \frac{2 \times (\text{Number of words in Common})}{(\text{Total Number of words in both versions})} = \frac{2 \times 3}{22} = 0.273$$

$$\text{Difference} = \frac{(\text{Num of unique V1 wrds}) + (\text{Num of unique V2 wrds})}{(\text{Total Number of words in both versions})} = \frac{8+8}{22} = 0.727$$

**Figure 2. Comparison of the Rendering of Phi 4:5 Between RSV and GNT**

RSV	Let all men know your forbearance.		at hand.
Common		The Lord is	
GNT	Show a gentle attitude towards everyone.		coming soon.

With modern personal computers, the comparison between full Bibles process can be done in a matter of hours. The results can be tabulated as in Table 2 or plotted using Cluster Mapping Techniques as in Figure 3.

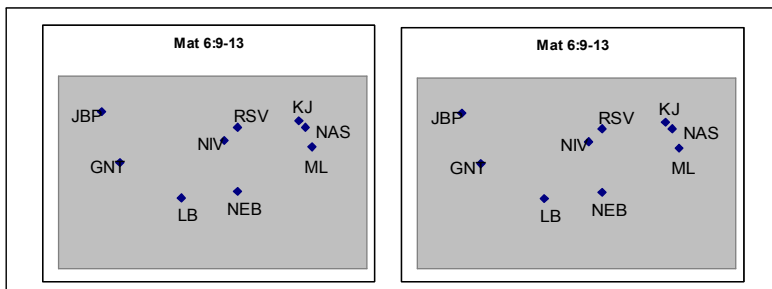


Figure 3. Similarities between English Translations

With modern personal computers, the comparison between full Bibles process can be done in a matter of hours. The results can be tabulated as in Table 2 or plotted using Cluster Mapping Techniques as in Figure 3.

**Table 2. Variation Between English Translations**

		% Variation Between Versions of Psa. 1:1								
		GNT	KJ	LB	ML	NAS	NEB	NIV	RSV	
GNT	Good News Translation		55	48	53	51	49	51	53	
KJ	King James Bible	55		60	32	22	56	18	13	
LB	Living Bible	48	60		40	60	62	54	61	
ML	Modern Language Bible	53	32	40		20	45	27	20	
NAS	New American Standard Bible	51	22	60	20		41	12	11	
NEB	New English Bible	49	56	62	45	41		43	42	
NIV	New International Version	51	18	54	27	12	43		9	
RSV	Revised Standard Bible	53	13	61	20	11	42	9		
		% Variation Between Versions of Mat 6:9-13								

		GNB	JBP	KJ	LB	ML	NAS	NEB	NIV	RSV
GNT	Good News Bible		33	42	29	40	42	24	30	36
JBP	J.B. Phillips Translation	32		45	37	44	44	35	35	36
KJ	King James Bible	41	46		38	10	4	36	21	16
LB	Living Bible	29	38	38		35	37	30	27	31
LML	Modern Language Bible	39	44	10	37		10	30	22	20
NAS	New American Standard Bible	40	45	4	37	10		34	21	16
NEB	New English Bible	23	35	35	31	29	34		24	27
NIV	New International Version	29	35	21	29	22	21	24		9
RSV	Revised Standard Bible	37	37	16	32	20	16	27	9	

This analysis is useful for ascertaining the history, derivation, and approach used for each translation that exists. Dynamic translations will be dispersed across the map while formal translations will tend to cluster around the Bible manuscript from which they are derived. It is also useful for demonstrating whether popular passages of Scriptures are being handled in the same manner as the rest of the translation.

- **Cluster mapping of semantics:** Translation can be described as a quest to transfer meaning from one language to another through a process of appropriate choices. It takes tremendous skill to take the ancient Hebrew words of King David and render them into words that would communicate the message of Psalms effectively to an eight-year-old Korean girl. In the past, scholars would attempt this task by guessing or prescribing which grammar would be used, and which words would populate the semantic domains of the target audience. However, recent studies of the medical profession have clearly shown that educated scholars grossly overestimate how well they communicate with the general public<sup>5)</sup>. My informal observations of youth listening to public readings of the Bible also indicate that translations used in worship do

---

5) According to the University of Michigan Health Services Patient Education Repository, the 1992 Adult Literacy Survey showed that about half of US adults read at or below the eighth grade reading level. Most of the materials submitted to the Patient Education Repository had to be rewritten at or below the eighth grade reading level. URL: <http://www.med.umich.edu/pteducation/read.html>

not speak as clearly as one might hope.

However, with modern personal computers, it is possible to create and analyze corpus of literature and collected sayings of any given target group in order to actually measure and describe the specific features of the actual form of the language used. Comparing this information against that of the Bible text is useful for identifying the parts of the Bible, which will be difficult to translate. This task has been made easier through the use of standard cataloguing systems developed for thesaurus in order to identify the semantic domains of each word in the lexicon.

As one compares the data, important clues for a translation project emerge. Words in one language often lack the precision of words used in another. For example, it is well known that the single English word *love* by itself does not describe the exact nature of the affection described by either Greek word: ἀγάπη, or φιλία. In addition, the number of word choices within the semantic domains of the Bible will vary greatly from one language to another. Sindhi speakers would want to know which of the 12 different kinds of earrings were used to make the golden calf. The Inuit of northern Canada would like to know which of 15 different types of snow represents the color of our forgiven sins. In short, where the grammatical and lexicon mapping between the source and the target languages are significantly different, one can expect difficulties in translation that will require special care and testing.

In addition, this approach can also be used to compare various attempts at translation to determine which approach would communicate better with a specific target audience. The translation that best mimics the vocabulary, grammar and style of the language of a target audience will have far less difficulty in conveying the Good News.

- **Checking for Readability and Understanding:** Developing dynamic translations that are easy to read and understand is the main goal of UBS Translations. Despite the importance of this, many translation teams do not actually measure these characteristics with member of the target audience. Without proper prepublication testing, a new translation represents a major risk to the reputation of a Bible publisher.

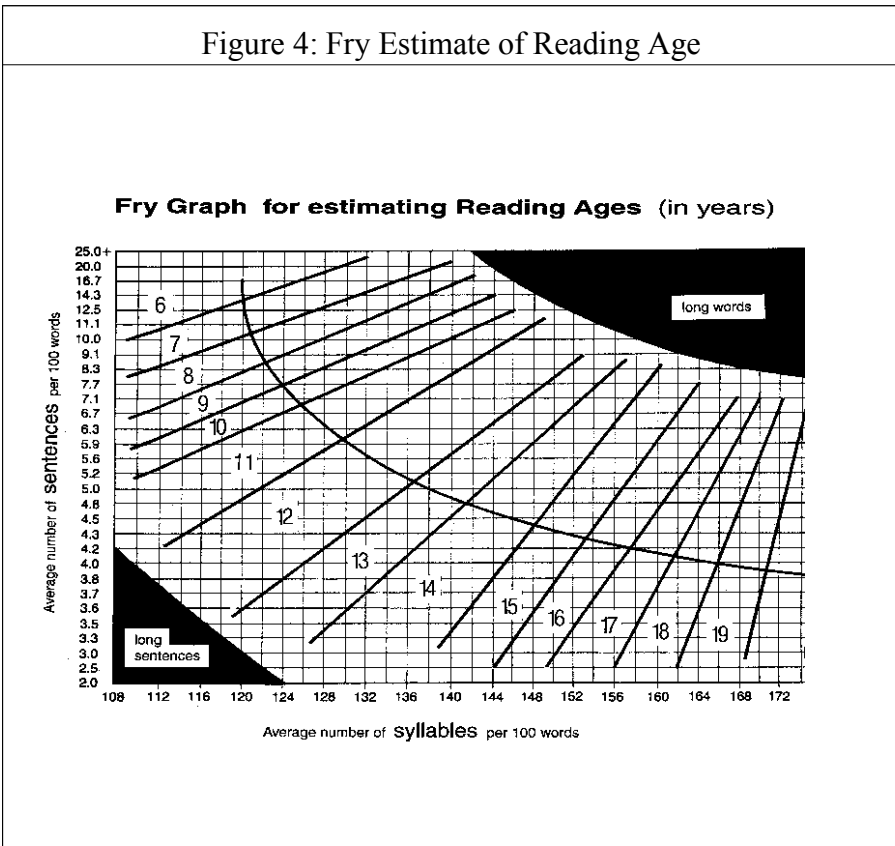
There are a number of tests that have been used for determining the



reading level of English material and which can be easily calibrated for other languages. In American English, the most commonly used tests for readability are Fry<sup>6)</sup>,

Fleishmen<sup>7)</sup>, and SMOG<sup>8)</sup>. The basic principle is that long sentences and long words add to the complexity of a passage and conversely to the level of education required to understand the material. A simple Perl program can quickly measure the readability of passages from a Bible. It can also be used to highlight passages that may be too difficult for the target audience to read.

Figure 4: Fry Estimate of Reading Age



6) The Fry Readability Method is to determine the average number of sentences and number of syllables that can be found in consecutive 100-word samples. The reading level can be looked up on the Fry Readability chart. Method published in 1978 by E. Fry *Fry's readability graph: clarifications, validity and extension to level 17*, Journal of Reading, vol 21

Understanding of a passage is a little harder to determine as it requires testing the comprehension of readers. The cloze test is the most popular way to test for understanding. These tests are generated by replacing keywords of a passage of Scripture with a blank that the reader fills in. These tests are easy to produce and can accurately measure whether one version of a translation is better understood by the target audience.

#### 4. Preparation of Bible Manuscripts for Publication

The most challenging and time-consuming part of translation is developing a manuscript that accurately captures the translation being developed. The manuscript forms the key resource to arise from a translation project, as it would be used dynamically to develop a wide range of Scripture products. However, a Bible manuscript can take 3 man-months to proofread and in that time it is hard for a single reader to remember all aspects of the manuscript. The following check are a list of computer checks that have been designed to facilitate the checking process.

- **Completeness and consistency:** Editors of Bible manuscripts are well acquainted with the warning of Rev 22:18-19 that warns that Scripture must be complete and correct. There are various programs around that automatically check the sequencing of chapters and verses in the manuscript according to various traditions of canonical order. Missing verses or empty verses (verses with no text) are reported to the computer

---

7) Flesch-Kincaid Formula is a US Government standard test used by the Department of Defense. The grade level is calculated by determining L, the sentence length (average number of words per sentence) and N, number of syllables per word. The grade level is equal to  $(L \times 0.39) + (N \times 11.8) - 15.59$  URL: <http://www.nist.gov/itl/div894/ovrt/people/sressler/Persp/Views.html>

8) SMOG (Simplified Measure of Gobbledygook) test is another quick, consistent, and easy to use tool to determine reading level of written materials. However, it is considered to not be as accurate for materials written at a less than sixth grade reading level. To conduct this test, a sampling is done by collecting 10 consecutive sentences near the beginning of the work, 10 sentences in the middle and 10 sentences near the end (30 sentences in total). From this sample, N (the number of words in this sample with three or more syllables) is determined. The grade level is equal to 3 plus the square root of N. Described in McLaughlin, H. (1969), SMOG grading: A new readability formula. Journal of Reading, 2 (8) 639-646.

user. In addition, sequence variations are compared to known variants (such as multiple endings of Mark) and discrepancies are reported.

At the same time, it has been found that for each genre in the Bible, there appears to be a general ratio between the length of the source and target text. Utilities are available to compare the target text to the source text and determine which verses appear to contain significantly more or less words than expected. This test has been useful to identify verses that were not fully translated, passages that were difficult to translate concisely, sentence fragments that were orphaned by cut and paste operations as well as interesting features of a target language that efficiently render a Bible phrase.

- **Titling, headings, footnotes and reader helps:** A century ago, most Bibles had few section headings and other reader helps. However, it was been found that readers find it easier to read a Bible that has been divided into meanful sections. Section breaks and headings can easily be inserted into unmarked Bible version automatically and the results can easily be compared against model texts using tools that work with Paratext. In addition, extract tools exist to enable the translator to view headers and titles in isolation to better focus on the readability and relevance of the reader helps as well as to compare against those in model text to ensure completeness.
- **Paragraphing, Indentation and Line breaks:** Using utility software designed to work with Paratext, it is possible to compare paragraphing, indentation and line breaks against that used in model text. In addition, the appearance of the final product can be simulated to allow translators to double check the accuracy of paragraphing, indentation and line breaks. This is particularly important in poetic sections that use indentation to display the underlying structure of the section.
- **Punctuation:** Computers can use comparison against model text to check the location, pairing and use of quotation marks in a translation. In addition, the text can be checked against known text bases to provide markings needed for extracting the discourse into a format needed to drive captioning systems used for audio production of dramatized readings of the Bible, and video typesetting of captions for Bible films like the

Jesus Film.

- **Spelling, Orthography, and Hyphenation:** Spelling is a major challenge for any publication project but for a Bible translation project extra special care must be taken to ensure accuracy and correctness of the words given. This is especially true with Biblical names like Babylon that occur in low frequency across the length the volume. It is not uncommon to discover multiple renderings for such names as the words are infrequent enough that the misspellings go undetected. New utilities have been developed to address this problem using a variety of approaches: identification of abnormal character sequences, matching of words that sound alike or that differ by a single typo or transposition, and removal of affixes to identify the underlying root forms.

Because a Bible translation project can span a number of years, major changes in orthography can be expected before the Bible translation is completed. Using dictionary or rule-based approaches, it is possible to automatically update a Bible manuscript. The computer can also analyze the impact of a proposed orthography change by listing all words that would be affected.

Within Asia-Pacific there are numerous minority groups that span political borders. While individuals on either side of the border speak the same language, local political pressure has created situations where multiple scripts are used to represent this language. For example, the Kamut of Cambodia, Laos, Thailand and Myanmar speak the same language but use 4 separate writing systems. In these cases, it is often possible to use dictionary and rule-based processing to create software that supports both orthographies from the same manuscript. This makes it easier to develop Scripture products that can be used on either side of the border.

Automated hyphenation can be developed from a variety of approaches that can be tailored to the needs of a particular language group. The emphasis not attempting to identify every place that can hyphenated in a word but rather that every computer-generated hyphen appear in a correct position. Currently, computer-generated hyphenation for the personal computer includes a combination of the following approaches that can be used as required on any language in the Asia/Pacific region:

- Uses of a fully hyphenated dictionary to auto generate hyphenation rules and exceptions.
- Use of fully hyphenated list of long words. (Words not in the list are not hyphenated).
- Rule based hyphenation that identifies syllable boundaries.
- Rule based hyphenation that identifies morpheme/affix boundaries
- Exceptional word process

## 5. Improved Bible Search Engines and Indices

Commitment to helping readers understand and get the most out of their Bibles is one of the major themes of the UBS worldwide program. This includes helping readers look up things in the Bible. Initially, Scripture search engines were only designed to able to deliver exact matches. For example, a search for the English word *love* was unable to find the word *loving*. However, this simple dictionary-like look up does not work well in a world of decreasing Biblical literacy because this approach requires prior knowledge of the vocabulary used in the Bible.

In addressing this problem, Bible publishers have been developing new generation of search engines that can provide information and relevant Scripture to those unfamiliar with the text. Here is a list of approaches that have been used:

- **Parsing surface forms to lexical roots:** This technique maps words as they appear in the Bible to standard dictionary form without affixes. If user types the English word *love*, the system is able to locate all derivative forms of the word such as *loves*, *loving*, *beloved*, *loved*, and *lovable*. This makes it easier to search for all occurrences of a common root word without working through all the various possible combinations.
- **Semantic mapping of lexical roots using a thesaurus:** This technology links together words of similar meaning into semantic domains. A traditional search of the RSV for the English word *affection* yields only 7 matches. A modern search of this word would also suggest several

hundred references for *love*, and *charity* as well.

- **Topical mapping:** This technology links together major themes and concepts found in the Bible and attempts to provide additional information that is relevant to the search. There two major applications of this method:
  - grading the relevance of search results by specific topics: A traditional search of the RSV for the English name *John* would return a mixture of references to *John the Baptist*, *John the Apostle*, and *John Mark*. However, a topically oriented search would give preference to additional references to specific person currently under consideration.
  - identifying related themes: A traditional search of the RSV for the English word **despondency** will not succeed, as the word does not occur in the Bible. However, a topical search of the Bible for depression would yield over a dozen passage that illustrate this human condition.<sup>9)</sup> In fact, clinical psychology instructors in some medical schools have used 1Ki 19 as an excellent case study of depression.

All of these approaches are attempts to provide relevant information to those unfamiliar with the text of a Bible translation. This technology will require extra linguistic and Biblical input to be tagged to each word used in a Bible translation. The best Web and CD examples of this technology are products of close collaboration of translators and computer specialists to develop applications that use IT to capture the linguistic and Biblical information into a practical and useful knowledge base.

## 6. Conclusion

Computer technology opens new doors at an astounding rate. However, most translation project teams do not fully exploit the potential either because they are unaware of the possibilities or they have not learned how to harness the power of this technology. The best results have come from project teams that encourage and foster collaboration between translators and Computer-

---

<sup>9)</sup> Despondency turns out to be a fairly common human condition that is recorded in the Bible: Gen 4:13-14, 21:15-16, Exo 4:1,10,13, 6:9, 14:15, Num 11:15, Jos 7:7-9, 1Ki 19:4, Act 27:20

Assisted Publishing (CAP) specialists. An enlightened programmer can develop a simple linguistic tool in minutes and hours that could save a project teams days and weeks of hard work. Likewise, with proper training and technical support, a translator can complete a new translation in a fraction of the time and with fewer errors. A little synergy between CAP staff and translation project teams can make a huge difference in the progress and outcome of a translation project.

## Abstract

In the past two decades, process of translating Scripture has been greatly facilitated by the use of personal computers and has driven their development. This paper is an attempt to explore practical and useful ways translators can use their computers effectively in the development of new translations of Scripture.