INSTANT BIBLE TRANSLATION

Posted on July 14, 2009 by Eddie

Category: Methodology

Mark Brown of New Zealand has just given a fascinating talk at the Codec: Christianity in the digital space symposium. I wasn't there, but I was able to follow comments from people who were on Twitter as well as watch the broadcast of his talk over the web. The talk has now been published as a pdf and it is something that I think most readers of this blog will want to get to grips with. Download it here.

The paper is too long and has too many interesting observations for me to think of commenting on everything he says, but I did want to pick up on one aspect of his talk. Mark mentioned that Bible translation is a slow process and that progress is very slow (I'll leave aside the fact that he only mentioned one Bible translation organisation - and not the one doing the most translation!). He then went on to suggest that by using online translation tools translation could be speeded up enormously.

> *We may see a day in the not to distant future when translation, whether written or spoken, is a near instantaneous process. So rather than 20 years for a full Bible translation, it takes 20 minutes.*
>
> *Right now companies like Google have the grand plan of collating and indexing all information, so that it can be searched and referred to. An example is Google Books where there aim is to have a digital copy of every book, magazine etc.. on the planet, with the ability to search every word and image. There are now some 7 million books that have been digitized, with agreements made with more than 20,000 publishers and authors. They have also partnered with a number of University Libraries including Oxford Universities Bodleian Library.*
>
> *What if their vision included documenting each of the 6,909 language groups? The text, cultural nuances, idioms, and so on. And where there is only an oral language the spoken equivalent is documented.*
> *Well this has started with the Google translate project which presently has some 34 language pairs. Their style of translation doesnâ€™t use a rule based approach, with emphasis on grammar etc.. but they, â€˜feed the computer billions of words of text, both monolingual text in the target language, and aligned text consisting of examples of human translations between the languages. Google then applies statistical learning techniques to build a translation model.â€™*

There are a number of significant problems with what Brown is saying here. Firstly, he clearly underestimates the time and effort that it takes to gather data in unwritten languages. We can do clever, corpus based computing with English and other languages precisely because there is an ubelievable amount of information written and digitized in those languages (this post is adding to the available information in English). But if you take most unwritten languages around the world, there is

very little available - perhaps even nothing. To capture the millions of words of data, of all sorts of different genres, which would be needed to allow the sort of analysis Brown is talking about is a huge task. It can't be done without people with digital recorders spending vast amounts of time meeting people, recording and transcribing data. We are talking about years of work here for each language. Brown seems to suggest that in some way we will have access to automatic transcription, digital voice recordings will be accurately transcribed by a computer into a written format. Well, anyone who has tried to use the products which claim to do that in English, knows how much the technology needs to advance before it is fool proof and could handle the amounts of data we are talking about here. And that is in English? I don't see anyone putting the effort into making a system that will accurately identify advanced tongue root, lexical tone, prosodic nasalisation and the other fun things that we had to wrestle with in Kouya.

But, let's just imagine that some does do all of this data collection and analysis in the foreseeable future, couldn't we still use computers to do Bible translation? Well, let's reread a little bit of what Mark Brown said.

> *â€˜feed the computer billions of words of text, both monolingual text in the target language, and aligned text consisting of examples of human translations between the languages. Google then applies statistical learning techniques to build a translation model.â€™*

In other words, the model for doing translation is built on translations which have previously been done by humans. What is the text that is most often the first to be translated into minority languages? Yes, you've got it, the Bible (or sometimes the UN declaration of human rights). So, in order to do an instantaneous translation of the Bible, we will need to load up the system with a translation of the Bible. I don't actually see much of a short-cut here. It is Wycliffe's vision that a translation project would start in every language that needs it by 2025 (and we are not far off track). I don't realistically see the sort of data collection that Mark is talking about happening till well after then.

Don't get me wrong, there is a huge role for computers to play in Bible translation. If you have an interest in IT, then there is plenty of work out there. But let's be realistic in our expectations.

But, even if we could do a Bible translation in twenty minutes, would we really want to? As Lingamish pointed out today, the problem is not getting the word translated, it is getting people to read it and base their lives on it. One of the most important aspects of any translation project is the Godly lives of the translation team demonstrating the reality and relevance of the message long before the words emerge from the printer. As one African church leader put it, 'we want to see the Holy Spirit in the lives of the translators long before we see the words Holy Spirit on the page'. There is a value in speeding up Bible translation, but not at the cost of losing the human and community touch. This is something we westerners with our task orientation find very hard to understand - but

the value ofÂ long term relationships trump the speed of Google in the Kingdom!