

Initial Report on the Parallel Corpus Translation Challenges and Process

This paper proposes to discuss to some of the issues emerged in the initial stages of the translation of parallel corpora material from the Princeton Corpus (?) and thereby caused a delay in sending it at the previously scheduled deadline.

At the outset all the sentences of the 100,000 word text were separated and aligned with their corresponding Sinhala translations.

Initially it was found that Parallel texts contained some POS tags which we ourselves failed to notice. We then instructed our translator who was new to the concept of POS tags to ignore them in the translation.

Some examples are given below:

' Furukawa said the purchase of the French and German plants together will total about 40 billion yen -LRB- \$ 280 million -RRB-.

-LRB- Woodcliff Lake , N.J .

-RRB- -- William G. Kuhns , former chairman

The first translator we hired had been a senior person in the trade. He had been a translator for the Legal Draftsman's Department with long years of service behind him. He, after a few days came back to us with a confession that he has decided to give up our assignment with the complaint that it is 'Difficult to achieve an output'. The following are some of the issues revealed in the discussions some of us had with him. He then unwillingly and apologetically withdrew himself from translation project.

1. It was very difficult to find equivalent vernacular terminology for some American slang found in the texts as pointed out by him such as;

e.g. taken a bad rap – slang;. to take the blame and punishment for a crime committed by another: He took the rap for the burglary. (Random House Webster's Unabridged Dictionary Version 3.0 – RHW)

e.g. screw up = 1. a mistake or blunder: *The package was delayed through an addressing screw-up.* (Slang RHW)

e.g blip up = *Blip* was found as a separate word not as a prepositional phrase *blip up*

We too agree that some of these expressions can cause a problem for a translator finding exact equivalent in the vernacular.

American idiom was prevalent in the texts selected and we site some examples of such expressions below.

e.g. One of the fastest growing segments of the wine market is the category of super premiums—wines limited in production , page 77

e.g. For years , this group included a stable of classics page 77

2. The text contained highly subject related vocabulary.

e.g. a consensus manager,

3. American place names *Elsevier*, the pronunciation of Donoghue not found in the Random House Webster's unabridged edition 3.0. The pronunciation of some of the place names could be found in the RHW dictionary and we were able to use the transliterated forms in the Sinhala version.

4. Abbreviations: *N.V.*, especially those of States in the USA paused a problem.

5. Then sudden change of subject of the texts made translation cumbersome and time and time consuming. Translators generally do not like this.

6. Another comment = Structure is alien to us = e.g page 21 * *The turf recently has ranged from Chile to Austria to Portugal.*

As all this contributes to a time consuming job translators who work on piecemeal basis for payment will find a task of this nature unyielding. Hence the comment 'no output'

So it is natural for them ask for more time.

With the withdrawal of the former translator we hired a second person with more experience in journalism. He found the context of the texts somewhat familiar.

The present situation is all the translated texts are now being tagged our team members.