CHAPTER IV

CONCLUSION

Corpus analysis can be performed with the assistance of a number of different tools, with their own advantages and disadvantages. UAM Corpus Tool (2012) has the capabilities to annotate a corpus using a variety of annotation schemes automatically and manually. However, its current maximum capacity is 31,734 words / 164KB text file. Meanwhile, CQPweb is powerful with the query search with large and small corpora. However, it is an indexing tool which by default does not have any annotation function.

In this project, the author has combined the automatic SFL annotation function with the indexing function by creating a converter that can transform the annotation output from UAM Corpus Tool to CQPweb format. Thus, a large-scale text analysis based on SFL-Transitivity scheme will be possible.

Python programming language is used to create the converter. In this project, ten contemporary novels are used as a corpus data. This study consists of five stages: (1) the automation of SFL analysis using the UAM Corpus Tool; (2) the analysis of the UAM output; (3) the development of converter; and (5) the indexing and analysis using CQPweb.

There are six different query methods available for searching a corpus in CQPweb, namely basic word form, matching parts of speech, matching simplified POS tags, lemma queries, word sequences, XML tags, and proximity queries.

Most of the methods can be combined with wildcards to perform under-specified query.

With this research, the reader will be able to comprehend how useful it is to use two corpus tools to save time in annotating and achieve better results when expanding corpus analysis, particularly in transitivity analysis.

Future research based on these findings may focus on ways to optimize this converter. Since the program is being executed on the Google Colab site, the author can now focus on creating a friendly user interface. The next step is to refine the instructions and warnings. Moreover, in terms of tokenization, the process of separating texts into individual analytic units, has to be enhanced.