**THESIS PROJECT SUMMARY**

Natural Language Processing (NLP) is an area of computer science research that entails enabling computers to understand and manipulate natural language text or speech. It involves understanding how human beings learn and use natural language so that technological tools and systems can be built to utilize natural language to perform desired tasks. It is heavily applied in a variety of fields such as: artificial intelligence, machine translation and speech recognition, to create technologies such as: customer or personal virtual assistants, text summarization systems and chatbots.

Natural Language Processing systems need a lot of human language data to be able to work effectively and efficiently. Some languages have a lot of already recorded and easily accessible data which can be used to build NLP systems. However, there are many other languages that do not have a lot of recorded data and thus do not have effective and scalable NLP systems built for them. These languages are referred to as “low resource languages” and native Ghanaian languages fall in this category.

This thesis work is going to explore the various approaches to building NLP systems for native Ghanaian languages and look at the different ways in which the data for developing these systems can be obtained.

In this work, one way in which data is going to be collected is using Bible and hymnbook translations and bilingual turn preaching. Another way in which data is going to be collected is using a limited amount of recorded audio with its respective transcriptions. The main question we will be attempting to answer is whether any of these approaches to data collection is efficient for the development of NLP systems on local languages.