# DESIGN & IMPLEMENTATION OF A NEURAL MACHINE TRANSLATION SYSTEM (LET’S TALK) FOR THE TRANSLATION OF HYAM TO ENGLISH

Thesis/Report submitted in partial fulfilment of the requirement For the degree of

# B.Sc.

In

# Software Engineering By

**DOGO** KYOM BU/16C/IT/2082

# Department of Computer Science Faculty of Computing & Applied Sciences

Baze University, Abuja

May 2021

# DECLARATION

This is to certify that this Thesis/Report entitled Design & Implementation of a Neural Network Translation System (Let’s Talk) for the Translation of Hyam to English which is submitted by Dogo Kyom in partial fulfilment of the requirement for the award of degree for B.Sc. in Software Engineering to the Department of Computer Science, Baze University Abuja, Nigeria, comprises of only my original work and due acknowledgement has been made in the text to all other materials used.

Date: May 2021 Name of Student: Dogo Kyom

**APPROVED BY** …………………

### Dean

Faculty of Computing and Applied Sciences

# CERTIFICATION

This is to certify that this Thesis/Report entitled Design & Implementation of a Neural Network Translation System (Let’s Talk) for the Translation of Hyam to English, which is submitted by Dogo Kyom in partial fulfilment of the requirement for the award of degree for B.Sc. in Software Engineering to the Department of Computer Science, Baze University Abuja, Nigeria, is a record of the candidate’s own work carried out by the candidate under my/our supervision. The matter embodied in this thesis is original and has not been submitted for the award of any other degree.

Date: Supervisor: Dr. Chollette Olisa

Date: Supervisor: Mr.Nasiru Aboki

# APPROVAL

This is to certify that the research work, Let’s Talk and the subsequent preparation by Dogo Kyom with BU/16C/IT/2082 has been approved by the Department of Computer Science, Faculty of Computing and Applied Science, Baze University, Abuja, Nigeria.

By

Dr. Chollette Olisa Date

1st Supervisor

Mr. Nasiru Aboki Date

2nd Supervisor

Dr. Chandrashekhar V. Uppin Date

Head of Department

Dr. Muhammed Baba Date

Dean, Faculty of Computing and Applied Science

[Full name] Date

External Examiner

# DEDICATION

This Project is dedicated to God Almighty for keeping me alive to see the end of my degree and to my father, Engr. David Dogo, and my mother, Dr. Mrs. Kande Dogo, and also dedicate this project to my siblings for the support and guidance they have given from day one to present. To my mentor Mr. Yashim, for the training you have been giving me, ever since I was a little boy. Their presence kept me motivated and encouraged to strive for my goal and to my relatives, lecturers Dr. Chollette, Mr. Charles and Mr. George for their great assistance and support, and also to my lovely classmates, for their amazing love and support and for their encouragement and support directly and indirectly.

# ACKNOWLEDGEMENT

I express gratitude toward God Almighty for his leniency and effortlessness that kept me all through my B.Sc. To my adored parents, my open appreciation for his assurance, direction and for giving the essential help, I value it a great deal and I cherish you.

I am additionally appreciative to Baze University, for giving me the chance and direction to experience this B.Sc. program. My appreciation goes to my Mentor, my supervisor, my lecturers Dr. Chollette Olisa, my special gratitude ma I really appreciate thanks one’s Again. My Pleasure goes to my Lecturers. Mr Charles, Mr. George, Dr Amit Mishra and I express gratitude toward them for their analysis, recommendation, advices and redresses.

I additionally need to thank everybody that added to the achievement of my B.Sc. program, my companions and my relatives. This unquestionably has been a magnificent encounter for me.

# TABLE OF CONTENTS

[DECLARATION](#_bookmark0) [ii](#_bookmark0)

[CERTIFICATION](#_bookmark1) [iii](#_bookmark1)

[APPROVAL](#_bookmark2) [iv](#_bookmark2)

[DEDICATION](#_bookmark3) [v](#_bookmark3)

[ACKNOWLEDGEMENT](#_bookmark4) [vi](#_bookmark4)

[LIST OF TABLES](#_bookmark5) [x](#_bookmark5)

[LIST OF FIGURES](#_bookmark6) [xi](#_bookmark6)

[ABBREVIATIONS & ACRONYMS](#_bookmark7) [xii](#_bookmark7)

[ABSTRACT](#_bookmark8) [1](#_bookmark8)

[CHAPTER 1: INTRODUCTION](#_bookmark9) [1](#_bookmark9)

* 1. [: Overview](#_bookmark10) [1](#_bookmark10)
	2. [: Background and Motivation](#_bookmark11) [1](#_bookmark11)
	3. [: Problem Statement](#_bookmark12) [3](#_bookmark12)
	4. [: Aim and Objectives](#_bookmark13) [3](#_bookmark13)
	5. [: Significant of the Project](#_bookmark14) [3](#_bookmark14)
	6. [: Risk Assessment for project](#_bookmark15) [3](#_bookmark15)
	7. [: Scope/Organization](#_bookmark18) [5](#_bookmark18)

[CHAPTER 2: LITERATURE REVIEW](#_bookmark19) [5](#_bookmark19)

* 1. [: Introduction](#_bookmark20) [5](#_bookmark20)
	2. [: Historical Overview](#_bookmark21) [6](#_bookmark21)
	3. [: Related Work](#_bookmark24) [10](#_bookmark24)
		1. [: Microsoft Translator](#_bookmark25) [10](#_bookmark25)
		2. [Google Translator](#_bookmark27) [11](#_bookmark27)
		3. [: iTranslate](#_bookmark29) [12](#_bookmark29)
	4. [: Architecture of the Work](#_bookmark31) [12](#_bookmark31)
	5. [: Summary](#_bookmark32) [14](#_bookmark32)

[CHAPTER 3: REQUIREMENTS, ANALYSIS & DESIGN](#_bookmark33) [14](#_bookmark33)

* 1. [: Overview](#_bookmark34) [14](#_bookmark34)
	2. [: Proposed Methodology](#_bookmark35) [15](#_bookmark35)
		1. [: Waterfall model](#_bookmark36) [15](#_bookmark36)
		2. [: Prototype model](#_bookmark38) [16](#_bookmark38)
		3. [: Iterative model](#_bookmark39) [17](#_bookmark39)
		4. [: V-model](#_bookmark41) [18](#_bookmark41)
	3. [: Approach to Chosen Methodology](#_bookmark43) [19](#_bookmark43)
	4. [: Tools and Techniques](#_bookmark44) [19](#_bookmark44)
	5. [: Ethical Consideration](#_bookmark46) [19](#_bookmark46)
	6. [: Requirement Analysis](#_bookmark47) [19](#_bookmark47)
		1. [: Responsibility Matrix](#_bookmark48) [20](#_bookmark48)
		2. [: Fact Finding Analysis](#_bookmark50) [21](#_bookmark50)
		3. [: Data Analysis](#_bookmark51) [22](#_bookmark51)
	7. [: Requirements Specifications](#_bookmark57) [24](#_bookmark57)
		1. [Functional Requirements Specifications](#_bookmark58) [24](#_bookmark58)
		2. [: Non-Functional Requirement Specifications](#_bookmark60) [25](#_bookmark60)
	8. [: System Design](#_bookmark62) [26](#_bookmark62)
	9. [: Summary](#_bookmark68) [30](#_bookmark68)

[CHAPTER 4: IMPLEMENTATION AND TESTING](#_bookmark69) [31](#_bookmark69)

* 1. [: Overview](#_bookmark70) [31](#_bookmark70)
	2. [: Main Features](#_bookmark71) [31](#_bookmark71)
	3. [: Implementation Problems](#_bookmark78) [39](#_bookmark78)
	4. [: Overcoming Implementation Problems](#_bookmark79) [39](#_bookmark79)
	5. [: Testing Plans (Unit Testing, Integration Testing & System Testing)](#_bookmark80) [39](#_bookmark80)
		1. [: Test Suite](#_bookmark81) [40](#_bookmark81)
		2. [: Test Traceability Matrix](#_bookmark82) [40](#_bookmark82)

[Chapter 5: DISCUSSION, RECOMMENDATIONS & CONCLUSION](#_bookmark83) [41](#_bookmark83)

* 1. [: Overview](#_bookmark84) [41](#_bookmark84)
	2. [: Objective Assessment](#_bookmark85) [41](#_bookmark85)
	3. [: Limitations & Challenges](#_bookmark86) [41](#_bookmark86)
	4. [: Future Improvements](#_bookmark87) [41](#_bookmark87)
	5. [: Recommendations](#_bookmark88) [42](#_bookmark88)
	6. [: Summary](#_bookmark89) [42](#_bookmark89)

[REFERENCES](#_bookmark90) [43](#_bookmark90)

[APPENDICES](#_bookmark91) [44](#_bookmark91)

[Appendix A: Project Documentation/ Executive Summary](#_bookmark92) [44](#_bookmark92)

[Appendix B: Work Plan](#_bookmark93) [45](#_bookmark93)

[Appendix C: Gantt Chart](#_bookmark94) [46](#_bookmark94)

# LIST OF TABLES

[Table 1 Matrix for Risk Assessment 3](#_bookmark16)

[Table 2 Matrix for Risk Amendment 4](#_bookmark17)

[Table 3 Hardware Requirements 19](#_bookmark45)

[Table 4 Hardware Requirements 20](#_bookmark49)

[Table 5 Functional Requirements 25](#_bookmark59)

[Table 6 Non functional requirements 25](#_bookmark61)

# LIST OF FIGURES

[RBMT workflow 7](#_bookmark22)

[SMT workflow 8](#_bookmark23)

[Microsoft Translator 10](#_bookmark26)

[Google Translator 11](#_bookmark28)

[iTranslate 12](#_bookmark30)

[The Waterfall Model 15](#_bookmark37)

The Prototype Model **Error! Bookmark not defined.**

[The Iterative model 17](#_bookmark40)

[The Iterative model 18](#_bookmark42)

[Gender Chart 22](#_bookmark52)

[Age Chart 23](#_bookmark53)

[Current level of Education 23](#_bookmark54)

[Nationality of participants 24](#_bookmark55)

[Those interested in online learning 24](#_bookmark56)

[Application Architecture for Customer 26](#_bookmark63)

Application Architecture for Admin **Error! Bookmark not defined.**

[Activity Diagram for Customer 27](#_bookmark64)

Activity Diagram for Administrator **Error! Bookmark not defined.**

[Use Case Diagram 28](#_bookmark65)

[Sequence Diagram 29](#_bookmark66)

[Entity Relationship Diagram 30](#_bookmark67)

[A screenshot of the training of the model using google colab 32](#_bookmark72)

[A screenshot of the training of the model using google colab 32](#_bookmark73)

[A screenshot of the training of the model using google colab 36](#_bookmark74)

[A screenshot of the training of the model using google colab 37](#_bookmark75)

[A screenshot of the training of the model using google colab 38](#_bookmark76)

[A screenshot of the training of the model using google colab 38](#_bookmark77)

# ABBREVIATIONS & ACRONYMS

RNN - Recurrent Neural Network LSTM - Long Short Term Memory GRU - Gated Recurrent Unit

MT - Machine Translation

NMT - Neural Machine Translation

# ABSTRACT

In Southern Kaduna, the Hyam Community always welcome foreigners every year, due to their friendliness and hospitable nature. They have even welcomed Igbos in the past and when the Hausas wanted to kill them, they hid them and gave them a safe route back to the east. Now with this in mind, the Hyam people would find it hard to communicate easily, due to their little or lack of knowledge of the lingua franca, because of lack of exposure.

This project presents an overall model for the translation of natural languages (Hyam to English). It is a Neural Machine Translator, achieved using Recurrent Neural Network (RNN). The models are made up of large artificial neural neurons which is able to predict the probability of a sequence of words in a given sentence. This translation is achieved using hidden layers in both models, based on parallel documents provided for both languages. The models here are encoder and decoder.

# CHAPTER 1: INTRODUCTION

## : Overview

Right from the beginning, communication is a very key part of our lives. Language is seen or noticed as one of the representatives of culture be it for a group, society or a nation. But with this in mind, language can serve as a bane for communication. This happens when two or more people, who understand different languages, and not the same language. This unfortunately, hinders growth and development in the world today.

In developing countries like Nigeria, the nation is a living proof of the above statement, because ethnicity discrimination is done every day among people in places of work, appointments, just to name a few. One of the effective ways, or possibly the only way to communicate in a different language, is having a human translator as the person to ‘bridge the gap’ between them. But in some cases, the translator can be biased or be selfish, because he or she can alter information, so as to get more profit.

We came out with a web – based application let’s talk, which acts as a virtual, unbiased intermediary, to enable locals of Hyam ethnicity to understand English language as well as foreigners who understand only English language to understand Ham language, both written and spoken.

## : Background and Motivation

Natural Language Processing or NLP for short, is an area of specification under Artificial Intelligence that deals with automatic manipulation human language which is the natural language, using a software. It assists the computers or digital devices to interpret, comprehend and make use of the natural language. NLP gives way for people to interact with computer using natural languages. An example of this scenario is the Alan Turing test or Turing Test for short. In this example, a human judge would communicate with another human using computer as well as an unmanned computer, both behind a wall. If the human judge is not able to distinguish between the human and the unmanned computer, then the unmanned computer, is said to have passed the test.

What mainly NLP does, is that, it breaks down a natural language into smaller, less complex parts, which are identified as tokens. Tokens could be in form of words, characters, words, etc, and later on, attempts to comprehend the relationships of the tokens involved.

In the early 70s, the first signs of NLP emerged, when John Grinder (an associative professor of linguistics), and a psychology student, Richard Bandler observed that people who had the same amount of training, education, and years of experience all had different results at the end.

They later placed their focus on communication. How people who succeeded in their respective fields, interacted with each other, based on some factors, verbal language, movements of their eyes, their body language, gestures, etc. They studied successful people like Virginia Satir who

established Family Therapy as well as Fritz Perls, who is also a founder of Gestalt Therapy and Milton H. Erickson, who is well known for his hypnotherapist work.

When the study was already made, they were able to point out their patterns of thinking, which is what made the NLP successful. They made a theory stating that the brain can learn systematic procedures, which brings about positive as well as physical and emotional perks. The resultant study from all of this was identified as Neuro-Linguistic Programming.

NLP is applied in so many areas today. There are; machine translation, recognizing speech, opinion mining, question answering, automatic summary of information, chat bots, market intelligence, text grouping, recognition of characters and spell checking

Prior to the introduction of digital computers, there were tasks that needed intelligence, an example of such is the translation of a language to another language. This was a task that was never believed to be done by a machine. Machine translation, is the use of automated machines to translate a source language into a target language. Machine translation of a language can also be achieved by developing algorithms, which enable the computer or device to understand the semantics of a language, for quality translation without any help from humans when translating. In the ‘70s the building block for the premiered MT was put in place. Nowadays, MT is mainly based on a statistical system that can make a language rule-based on large corpus of already made translated text, rather than analyzing the rules or structures of every language and develop an algorithm to enforce the rules.

In communication, ICT has done so much from developing social media apps that can enable people communicate from far away as well as meeting up with loved ones or friends virtually, it has also helped in translating of languages. It is a quick and effective way to translate information received. This is known as machine translation. Machine translation is still a growing industry, given the fact of how languages are structured, and the vast amount of languages yet to be analyzed.

In Nigeria, language is very vast here. There are roughly over five hundred languages spoken in Nigeria. Now with the beauty of having so many languages to speak it is very hard to communicate easily. In Nigeria, this creates ethnic rivalry and diversity, which hinders our growth as a nation. A web application called ‘Let’s talk’, is developed with the aim of translating a language called Ham language. This is a language spoken by Hyam people who live in southern Kaduna. They are about 400,000 people who speak the language in Nigeria. It seeks to translate Hyam language to English and from English language back to Hyam language, this way, a local and a foreigner, can easily communicate without any problems.

Another point to note is that, since the language is not a major language, the probability of it being extinct by the newer generation, would be high. This would be to the fact that parents may not be educated, only their children, who would learn only English and may not learn Hyam, it would be an opportunity for them to learn it also, not just foreigners of the language.

The last point is that linguists, who plan on studying the language, can use the application as means of understanding the structure of the language.

## : Problem Statement

* + - Due to imperfection in this part of machine learning, it would be hard for the user to enable the system recognize his/her voice, due to noise, disfluencies, vocabulary size, and language perplexity.
		- The ability for the machine to translate effectively and correctly a language to another, is very bad. It may fail to recognize which synonym, collocation or word meaning should be used, given the lexicon it has inside. A study regarding machine translation, that is, from English to Lithuanian and vice – versa, shows that two -thirds of all the sentences were incorrectly translated, which means there is a more or less slim chance that mobile applications can translate accurately.
		- Quality issues – as opposed to a human translator, the computer software, cannot process the context in which a language is being used. For instance, a word may have many meanings and due to the way its pronounced mixed with emotions, the user may refer to a particular meaning of the word, which the computer software may not comprehend.

## : Aim and Objectives

* + - To create a suitable mobile application that can be used to translate Ham language to Nigeria’s to lingua-franca, and vice-versa
		- To enable the computer understand, as well as learn the structure, phonetics, and spelling of the words.
		- Preservation of the language, as part of the Ham culture.

## : Significant of the Project

This project is significant, in the sense that people who know the English language, would be exposed to the English language and it would enable them to understand the language better as well as those who are fairly conversant with the language(ham language), or possibly those who cannot speak the language but understand English.

## : Risk Assessment for project

Risks examination are performed to effectively manage the project in the case of unforeseen problems that may happen before, during or after the project has been developed. It is important to have risks expunged, as much as possible from the project in order to maintain project efficiency and reliability, from both developer(s) and customers alike.

Examination of risks however, cannot be precise. If it were, then it would be easy to predict the future of the app. But having a place to begin talking of the symptoms of the problems, as well as how it would have an impact in the system’s performance, is much better.

Table 1 Matrix for Risk Assessment

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Identified Risk | Rate ofOccurrence | Difficulty in timedetection | Potential impact | Overall risk to project |
| Surpassing fixedbudget amount | Medium | Medium | Medium | High |
| Testing and errordebugging | Medium | Medium | Medium | Medium |
| Incongruoustechnologies | Low | Medium | Medium | Medium |
| Serverunreliability | Low | Low | Low | Low |
| Hindrance to appointed delivery ofsystem | Low | Medium | Low | Low |
| Amalgamation of systemcomponents | Low | Medium | Medium | Medium |

Table 2 Matrix for Risk Amendment

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Riskidentification | Preventivemeasures | Contingentremedy | Trigger | Owner |
| Surpassing fixed budget amount | Oversee the funds being used | Get more funds | Expectations tocapable deliverables | Kyom Dogo |
| Testing and error debugging | Qualitative as well as qualified developers andtesters | Contracting and reusing of system units | Increase in system difficulty | Kyom Dogo |
| Incongruous technologies | Use non-incongruous technologies | Reliable framework | Brings about a quality system | Kyom Dogo |
| Server unreliability | The database for the system should be backedup frequently | Backup server should be readily available | High traffic in the server and/or illegalinfiltration | Kyom Dogo |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hindrance to appointed delivery ofsystem | The team be reminded about deadline date | More working hours for the team | Deliverables hindrance | Kyom Dogo |
| Amalgamation of system components | Amalgamation should be prepared at priorphase | System testers should test the system in priorphase | Inappropriate preparation for amalgamation | Kyom Dogo |

## : Scope/Organization

This describes the remaining chapters of the project are arranged. This report basically consists of five chapters that are outlined below:

Chapter 1: provides a general overview of what the whole project is all about. This chapter contains background, motivation, statement of the problem, aims and objectives, significance of the project and project risk assessment.

Chapter 2: this talks on the literature review- introduction, historical overview, related works and summary.

Chapter 3: States the requirement analysis and design. Chapter 4: This includes implementation and testing.

Chapter 5: Discussion, conclusion and recommendation are in the last chapter. Additional information include: references and appendix

# CHAPTER 2: LITERATURE REVIEW

## : Introduction

This chapter is centered on the review of literature on past works that are related to this dissertation. Section 2.2 is an overview of the history as well of technologies used in developing machine

translation in order to learn and understand languages wherever you go. Section 2.3 talks on some of the literature available, that are used to d

## : Historical Overview

Natural Language Processing (NLP) or machine translation started decades ago. Based on this fact, research on speech translation started to gain prominence from the late 1980s and early 1990s.

In 1970’s there were some indications for MT. the Rule-based Machine Translation was the focal point in research as of that time. In RBMT, what the linguist does is that he formalizes the knowledge he has about linguistics of a language and translates them into corpus and grammar rules. The knowledge which is now obtained by the system is used to comprehend the source language and translate these sentences to their equivalent in the target language. However, this method does not need any training lexicon and grants which would oversee any translations made by the system, as well as the process of encoding the knowledge of the linguistics which needs huge amounts of dexterity. The RBMT systems were put into three categories: direct systems, transfer RBMT systems and interlingua RBMT systems.

The transfer model that is used for RBMT technology, which is implied to huge collections of linguistic rules has three stages; analysis stage, transfer model and lastly generation model. When it comes to the analysis stage, linguistic examination is done on the sentence of the input source, so as to retrieve information, which is based on morphology, parts of speech, phrases, word sense break down and named entity. In the transfer of lexicons stage, two steps are done during the transfer. They are; word conversion and grammar conversion. For the word rendition part, the source language of the original word is being substituted by the original word that is in the target linguistics, which is feasible using a bilingual lexicon and in grammar conversion, with suffixes which are also converted.

For the generation stage, the genders of the converted words are checked assisted by a short distance and long distance accord done by intra-chunk and inter-chunk unit. These are done to guarantee the gender, amount and person of the local groups of phrases concur to the gender of the verb of the subject.

RBMT is very complex when learning the structures for certain languages, an example is English- Japan. A translation system was proposed that operates by using the analogy principle, this means the system makes use of a big dataset (corpus) of already provided sentences, and their translations, so as to learn the correspondence between English to Japanese words plus the language(s) structure(s). (Nagao, 1984). People describe various ways, in which a system can be developed and ways to organize data that is needed for the system. There are two sub-fields under MT namely: Statistical Machine Translation (SMT) and Neural Machine Translation (NMT).



Fig 1 RBMT workflow

*The diagram is a pipeline, which describes the workflow of the Rule Based Machine Translation. Within the first stage, which is analysis phase, the analysis of linguistics is done on the source input sentence, so as to bring out information, based on terms of speech, morphology, phrases, named entity , as well as word sense disambiguation (WSD). In the transfer of lexicons, a couple of steps are involved namely: conversion of both grammar and conversion of word. When it comes to the word conversion, the base word of the source language, is changed by the base word from the target language. This is achieved with the aid of a bilingual lexicon, as well as conversion of grammar. However in the grammar stage, the suffixes are being altered. Whereas in the generational stage, the genders of the words that have been translated, are then corrected, which would accompanied by short as well as long agreements done by the intra-chunk and inter-chunk modules.*

For Statistical Machine Translation, the methodology works depending on the statistical models that are gotten from bilingual text corpora which are aligned in parallel, that concludes that all the words found in the target language is a likely a conversion of the words in the source language. The higher the probability is for a word, the more accurate the translation would be. In SMT, some of the problems it faces are the languages have constant design of disagreement between them, also overseeing the rearrangement of divergence. Key phases involved in SMT are: planning of corpus, training, deciphering and analysis.

Planning of corpus, arrangement and its refurbishing would be done in advance, before the processing stage. Training is a means whereby, a statistical machine learning algorithm that is overseen or not by a human, is utilized to develop statistical tables based on the parallel corpora.

In SMT, a key component for the parallel corpus planning is the word by word and phrase based arrangement. In the course of learning, translation model, distortion model, language model and phrase table are designed. In MT, the most difficult task is the decoding stage for the models. However, it is a key method whereby the conversions based on the target linguistics, are decoded with the help of the linguistics model, conversion model and phrase table being generated. The two main shortcomings of SMT are; decoding complexity and target linguistics rearrangement.



Fig 2 SMT workflow

*In the workflow above, there are four stages namely: preparation of corpus, training, decoding and finally testing. In the first step, we prepare the corpus, which is a lexicon of bilingual words in different documents, the alignment as well as the correctness would be achieved in the pre- processing phase. Training is a method whereby, either a manned or unmanned machine of statistical learning algorithm, which is used to develop the statistical tables, gotten from the parallel lexicon. In SMT, the word by word as well as alignment of phrase based, which is salient during training. Within the training, the language, translational models, and distortion tables, the phrase table are also modelled. The most complex task falls on the decoding stage, where the training models would be decoded.*

SMT, was brought about by Brown et al. in 1990. The theory assumes that every sentence let’s say S, in a source language, would have a feasible translation T that is in the target language. With this

in mind, SMT based methodologies gives to each (S, T) sentence pair the likelihood of P(T|S), which is interpreted as the likelihood that the sentence T is the translated version that is of the target language, of sentence S, written in the source language.

In NMT however, makes use of artificial neural network to be able to foresee the probability of a series of words, constructing whole sentences in one integrated model. Perpetual depictions of linguistic units, which can be in form of sentence, word document or character, has brought about encouraging results on several processing tasks. In one of the pioneer works, which was proposed by Bengio et al (2003), they model words with perpetual fixed aspect word vectors making use of neural network and achieve ground breaking results on the language modelling tasks. Again, Collobert et al (2008), have proved that repetitive modelling of words, capture the semantic, syntactic and morphological properties of the words.

A major goal of NMT, is to provide complexity. The neural networks can possess a huge amount of parameters that contain weights and biases, that are flanked by nodes, so as to provide the non- rigidness to match data that’s is very complex, as well as train the complex models.

NMT technology is used in any pairing of language, mostly bilingual. These languages however, can be either new languages or languages that can be understood by a specific amount of people. NMT can be tailored to match specific kinds of style and language types. The training data are the main focus which is used to train the neural networks, as it learns to copy the data being trained with.

The standard of for machine translation is measured based on the parallel translation of humans. This means that, the whole system for translation, is not just translating the individual words which makes the whole text. The context of the sentence, should also be put in mind also. Deep knowledge for the grammar is needed, as well as semantics and sentence structure of both the target and source languages.

Development of approaches for MT quality evaluation is a daunting job. Ilya U. (2012) states that assessment of translation correctness, is fully achieved as a result of the complex structure of different languages. This means that, two words in different sequences, can mean the same thing, and also two sentences with the same structure, but varying with only a word can give distinct definitions.

In the evaluation for MT quality, there are only two factors namely: adequacy and fluency. In adequacy, the translation that has been provided must match the original text in terms of meaning or definition. The other one which is fluency, the translation must match the original one in terms of grammar.

## : Related Work

The difficulty faced by when developing technological solutions for let’s talk is the ability to make sure that ICT solutions, for instance communication is well suited to behave and operate in the particular environment. Failure to customize adequately the product would undermine the ease of access and its relevance to the users, which would hinder the language translation.

According to both Hinton et al (2012) and Dahl et al (2012), deep neural networks, over the years have shown great amounts of improvements in computer vision and speech recognition tasks. Since then, neural networks have been effectively used to solve various natural language processing tasks like recognition of paraphrases, according to Socher et al (2011) and also word embedding extraction Mikolov et al (2013). Neural networks have also been used to improve the technology that is in SMT, its predecessor.

In the following subchapters, we talk on some systems or applications that are popular that makes use of Neural Machine Translation:

* + 1. : Microsoft Translator

Fig 3 Microsoft Translator

Microsoft Translator, is an application developed by Microsoft. It is a multilingual machine cloud service that is implemented within several consumer, developer and government BLEU -owned products. The methodology for the translation, is done using parallel texts, which is understood and deciphered by the inbuilt algorithms, rather than writing down the overt rules of the languages. The accuracy of the translation is measured by a score, which is a metric widely renowned and inexpensive.

* + - 1. : Pros of Microsoft Translator
				* The service can support 54 languages
				* It can translate up to 44 languages offline
				* All the shared vernaculars are being interpreted
				* The user interface for the Microsoft translator is friendly
			2. : Cons of Microsoft Translator
				* Microsoft translator is not always accurate- the BLEU rating rates the accuracy of the language, but if the algorithm to decipher the language is inaccurate, the translation would also be inaccurate, as it does not make use of the language rules to translate.
				* It contains only one kind of version
				* The languages in Microsoft translator are limited – the service covers only 54 of the languages, as opposed to google translator. This reduces the market of the users using it albeit the languages are major languages of the world
		1. Google Translator



Fig 4 Google Translator

Google Translator, is an application developed by Google. It translates text and websites from a source language into target language, using a gratis multilingual statistical in addition to neural machine translation service. Google translate, does not make use of rules of grammar, because the algorithms used are developed on statistical or pattern investigation as opposed to the traditional based investigation.

* + - 1. : Pros of Google Translator
				* The service/application supports 109 languages, including hausa, Yoruba and igbo
				* It can translate half of the languages supported offline
				* Google translate services are gratis
				* Google translator is fast.
			2. : Cons of Google Translator
				* The meaning can be altered, in the translation phase, due to the fact that context cannot be incorporated -
				* The language pair determine the standard of the translation – here, the ‘developed’ language pairs like French-English would give more accurate translations, since both a major tongues of the world would be more analysis, due to the high chance that people would make use of the pairs, unlike other language pairs like Swedish - Mandarin, for example.
				* Google translator generates translations, which have high amount of grammatical blunders.
				* There is no system whereas users can contribute, to increase the accuracy of translation – unlike Microsoft translator, people cannot make suggestions regarding a language, either on spelling or grammar. This is bad and reduces the accuracy of translation of that language.
		1. : iTranslate



Fig 5 iTranslate

iTranslate is an app, developed by iTranslate GmbH. It is available on both android and iOS. It is very useful for conversations, like extended conversations, because it keeps the previous translations it has made, whether voice or text. It also contains a dictionary, which is an added bonus, for those who want to increase their vocabulary in a certain language.

* + - 1. : Pros of iTranslate
				* It supports at least 90 languages.
				* An application that does free translation with both support from all kinds of devices as well as apple watch support
				* Offline mode for only eight languages. They are English, Spanish, German, French, Italian, Chinese, Portuguese and Russian
				* It has an outstanding mode for making conversations in voice.
			2. : Cons of iTranslate
				* In order to make use of the best features, you need to pay for it, including the whole access to all languages and its pairs

## : Architecture of the Work

Let’s talk is an Recurrent Neural Network architecture (RNN) with Long-Short Term Memory, which is unlike the Convolutional Neural Network (CNN), which is a feed forward network. The major reason why RNN is able to perform better is that is has a memory.

In RNNs, there are three kinds of neural networks, vanilla neural network, Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). The difference between the vanilla neural network and LSTM as well as GRU is the architecture or structure of the memory. For vanilla neural network, the network is ingenious, because the input is a little varied or unstructured unlike GRU and LSTM.

For example human brains, can understand as well as keep memory on past images, so as to understand the next sequence of pictures. CNN cannot perform this because of lack of memory which would be needed to understand the next sequence of either pictures, text or symbols. However, the memory can be flawed, if there is vacuum between the connected information and

the point it is required. RNN which is an advanced of CNN, has a feedback, meaning it is like a loop that would form directed graphs along a temporal sequence. It is used to process varying lengths of data. Some examples are handwriting recognition, translation of natural languages, recognition of speech, etc.

LSTM is a system from the deep learning world that avoids the problem of the disappearing slope. LSTM is been increased by the recurrent gates. Long Short Term Memory avoids training mistakes from either disappearing or exploding.

The RNN, has two models, decoder and encoder, which is like a sequence to sequence model. The model contains hidden vectors, which enable smooth transition, between the input vectors, of the encoder to the output vectors of decoder. The hidden vectors are also within the encoder and decoder nodes. In the encoder, the hidden vector, encoder hidden vector, whose goal is to hide the information that is in the input elements, which would make the other sequence called decoder to preserve the accuracy of translation, by morphing this hidden vector into the first hidden state of the other sequence.

The decoder, which is a stack of a lot of recurrent elements which gives an output of y\_t where t is time step. Every recurrent elements gives a hidden state from the encoder, as well as the later hidden states from the decoders, which then produces the output as well as its own hidden state of the decoder.

Within the input and output vectors have both weights and biases. These are the learnable measures that enable the learning model to be accurate when translating input to output. In a network, nodes are connected to one another, which means that the information is been preserved when translating text from input to output. Within the nodes are the weights and biases, they change constantly to improve the quality of translation for the input, as well as output. When the inputs are been applied to the nodes, the weight matrix is multiplied to the input. The weight matrix is responsible for how much impact the input will have on the resulting output. After that, before going to the next stage which is the bias, the network checks if it matches the specifications or not, where if it does goes to the bias else stays in the weight matrix. In the bias, it is a constant which is added to the result gotten from the previous stage. They act is increments of the input needed for the next layer. The equation for input, weight and bias is : Y = ∑(w \* i) + b. Where w = weight matrix; I = input; b = bias.

* Offline mode cannot be contextually correct – human languages are much unstructured, meaning it would be hard for computers to decipher the appropriate context of a word as of that point in time. Also, the online mode is more accurate, due to the app being up-to- date with the database that may change. Now having translations offline may not give you accuracy, because of the rigidness and obsoleteness of the database.
* The conversations done with voice cannot be done in offline mode.

## : Summary

The literature review that has been provided in this chapter, shows that Neural Machine Translation, which uses artificial neural networks, is the better way of developing an app or a service that helps to translate a language to another.

Therefore, it is very salient to point out that the predecessor, the Statistical Machine Translation, may not be very reliable, because it makes use of predetermined rules, which are not the same for each language involved.

# CHAPTER 3: REQUIREMENTS, ANALYSIS & DESIGN

## : Overview

This chapter will expand the key phase of the development of let’s talk as well as other findings done in the literature review. Furthermore, this chapter will include the research methodology, techniques and tools, feasibility studies and also the fact finding techniques that includes questionnaires and interviews.

This chapter expatiates on the requirement analysis design, the functional aspect of the application. Furthermore, in the chapter, we also talk about the techniques used to gather information for the software’s design. Analyzation required for the data, was done carefully

## : Proposed Methodology

A system design methodology is the process of outlining the possible components that form together a system for a unique situation. Methodologies have been developed over the years, with each of them having its advantages as well as disadvantages. One methodology, is favored for certain types of projects, as a result of several industrial, hierarchical, project and group considerations. In methodologies, we discuss some hypothetical methods which ensure that requirements are met, but not fully expatiate the methodologies.

* + 1. : Waterfall model

Waterfall model is a progressive and linear development method, whereby the development steps change from one to another steadily downwards, just like a waterfall. The phases involved are, requirements collection, design, implementation or development, testing, deployment and lastly, maintenance of the system. The outcome of the previous stage forms the input for the incoming stage. These phases however, cannot be revisited again whether it is completed or not, with no form of overlapping for the stages involved.



Fig 6 The Waterfall Model

* + - 1. Advantages of the waterfall model
				* The stages are self-explanatory
				* It is very easy to use
				* A great choice for small projects, where the requirements are very understandable
				* The process and the result of the stages, are listed in the documentation.
				* The resources utilized would not be much because the project is small with untestable requirements
			2. Disadvantages of waterfall model
				* High risks and high ambiguity
				* Requirements cannot be altered in any of the stages
				* Since testing would be done in the future stage, a slim probability that risks at the previous stages may not be known.
				* The system cannot be revised when a stage has been completed
		1. : Prototype model

In the prototype model, we make use of prototype – a version of a product, which is not the final product, but has the features end elements required like the final product. This prototype is now developed as many times as possible, so as to obtain the ideal product which is the ultimate product.

Customer Suggestion Acceptance by Customer

Customer evaluation of prototype

Requirements Gathering

Refine Requirements

Build Prototype

Quick Design

Design

Implementati on

Test

Maintenance

* + - 1. : Advantages of Prototype model
				* The users or customers, are immersed keenly in the development process
				* Errors can be discovered in the initial stage(s)
				* The features are not in the system can be easily identified
				* The feedback is faster and available which leads to great remedies
			2. : Disadvantages of Prototype model
				* The requirements in the prototype model may be altered
				* Large amount of resources, as well as time is required
				* Unfinished or poor error analysis in the previous prototypes can result to poor system design
				* The endorsement steps as well as control is not stern
		1. : Iterative model

Iterative model is an approach whereby larger software development phases are broken down into smaller segments. The main focus is on the early development and design and overtime, gain complexity with the requirements, which would continue to be improved or developed, until the ultimate product is fully built or designed.



Fig 7 The Iterative model

* + - 1. : Advantages of Iterative model
				* Some of the features of the system, can be developed in the preliminary stages
				* Implementation of this features in the preliminary stages, can enable developers as well as testers, to find any possible fault in the features early and rectify them. This would improve the quality of the product, as they go to the next stage(s) of development
				* For agile organizations, it is advisable to use this approach
				* There is less time spent for documenting, which creates more time for making designs
				* Very cost effective in terms of making alterations for the requirements
			2. : Disadvantages of Iterative model
				* Iterative model utilizes a lot of resources
				* It is not appropriate for small projects
				* Dexterous resources are needed for skill study
				* The progress of the project is vastly reliant on the risk analysis stage
				* The model is not appropriate, if requirements would be altered at any point
		1. : V-model

V-model (verification or validation) contains a testing stage synonymous to the development phase, which is in the waterfall model. Basically, it is an augmentation of the waterfall. But the steps does not move downwards and linearly like the waterfall model, but rather turned upwards at the end of coding stage, giving the ‘V’ shape.



Fig 8 The Iterative model

* + - 1. : Advantages of V-model
				* It is uncomplicated to use
				* It is very appropriate for small projects, with clear requirements
				* Errors or defects are easily detected at the preliminary phase
				* The downward movement of the errors are eluded
			2. : Disadvantages of V-model
				* The model is very rigid
				* A huge amount of technical resources, as well as technical dexterity, is suitable for the V- model

## : Approach to Chosen Methodology

In this project, we implement the waterfall methodology. Based on a critical investigation of the system, we investigate the software as a complicated system, due to the fact that the requirements from the users collected at the early stages of the project are ambiguous from the developers’ perspective. So, we make use of this approach, in which we make sure a phase is completed before moving to the next stage.

## : Tools and Techniques

The technicality required to develop the software is made up of resources which are updated to match the current technology. We consider some factors to ascertain if we have the needed requirements to develop the software. The technical feasibility is divided into two groups: hardware and software.

Furthermore, for the hardware we have two groups: developer hardware and customer hardware.

Table 3 Hardware Requirements

|  |  |  |
| --- | --- | --- |
| S/N | Items | Specification |
| 1 | Processor | Intel Core i7 4510U |
| 2 | Memory(RAM) | 8 GB |
| 3 | Display(Graphics Card) | Intel HD Graphics Family @2GB |
| 4 | Hard Disk | 1 TB |
| 5 | I/O | Mouse, keyboard and cd drive |
| 6 | Network |  |

## : Ethical Consideration

We make sure to keep the information provided by the intended users are safe and they kept incognito, so as for them to provide honest answers without duress or feeling threatened when supplying the information.

## : Requirement Analysis

When it comes to requirement analysis, there are made up of different steps. The first step is the initial investigation. In this phase, retrieving data is key here. For this reason, fact finding techniques are very important, so as to get the specific kind of answer(s) based on the proposed project. The following methods are used:

* Questionnaires: This are prepared set of questions, with the main purpose of retrieving information or data from the potential target market. The target market in this case are mainly locals from the Ham land. The questions asked are drawn from fixed format questionnaires, which is basically predetermined options that the individuals would select from, as well as free format questionnaires which would also be used. The pros of using questionnaires are; it saves and or reduces the resources that are used, compared to other techniques and there is no prior training when it comes to dispatching them and also answering them.
* Interviews: In this type of gathering technique, the author would oversee interview sessions, so as to get the analysis required for the research. The interviews are a means of retrieving precise data from the target market regarding the current system. In the course of the interview, the interviewer would be able to reveal areas where there is misunderstanding, as well as descriptions of activities and the problems of the systems from the interviewee(s), with resistance to the proposed system. The pros of using interviews is that; questions not understood can be further explained, and the feelings or reactions of the users are captured.
	+ 1. : Responsibility Matrix

The goal of the responsibility matrix, is to ascertain the people who would be assigned to complete a particular task or activity. This is done so as know if a project has not been completed it can be easily be traced to the individual(s) or department who erred to complete the task. It defines the length of the responsibility as well as the relationships among groups.

The following table will give an illustration of the key responsibilities of the project and the person or department who would handle the activity.

Table 4 Hardware Requirements

|  |  |  |
| --- | --- | --- |
| Member | Position | Responsibility |
| Kyom Dogo | Project Manager | Oversees the entire project |
| Kyom Dogo | Graphic designer | Designs the graphicalcomponents |
| Kyom Dogo | System analyst | Designing of the elements of the system, the modules andtheir system interrogation |
| Kyom Dogo | Programmer | The functions are beingscripted |

|  |  |  |
| --- | --- | --- |
| Kyom Dogo | Database administrator | Designing and also implementing the system’sdatabase |
| Kyom Dogo | Software tester | Testing the whole system todetermine it is free from errors |

* + 1. : Fact Finding Analysis

We first talk on the questionnaires which has two parts: Part 1 and 2.

The first part of the questionnaires is based on demographic questions. This is used when the person taking the survey wants to sort out the wants of the participants involved, again it assists the organization to understand what the individuals and satisfy them.

The questionnaires were further divided into sub-categories namely:

### Gender.

The goal here is to determine the ratio of male to female who took part in the survey as of that time.

1. **Age**.

The goal is to ascertain the age groups that took part in the survey as of that point in time.

### Highest level of education completed.

The goal is to ascertain the highest or current level of education attained by the participants in the survey in that moment. This is done to determine the complexity of the proposed system to the potential users, so that if matches their level of intellect.

### Nationality

The goal is to determine the nationality of the participants involved in the survey

The second part consists of both fixed and non-fixed format questions, so as to retrieve user requirements and views.

* + Would you be keen on learning Hyam language via computer means?

**The objective:** the author would want to know if they are interested in the computer method, or if they still prefer the human translator

* + Would you like to learn Hyam language using let’s talk?
	+ Have you been recommended to learn languages online?
	+ Would the app enable you to understand English and/or Hyam language?
	+ Which of the language (Hyam or English) is easy to learn?
	+ Is Hyam your favorite language?

**Interviews:** A more in depth analysis when gathering user requirements and views, since it requires a face to face dialogue.

* + In what way would the app help you?
	+ How long does it take to learn a new language?
	+ If you were to use the app per day, how many times would you use it?
	+ What are the steps taken when you want to learn a new language?
	+ What slangs do you know that can be added into the system?
	+ How long does it take you to learn a new language?
	+ Are there any ways let’s talk can be improved?
	+ If you could add some words or sentences to the system, what would they be?
	+ Do you think that let’s talk is user friendly?
		1. : Data Analysis

We make use of Microsoft Excel to generate the reports which would be used in the research analysis, due to the fact it’s user-friendly as well as it reducing the monotonous means of calculating manually. In the figures below, pie charts have been made to represent as well as sum up the results that have been retrieved.



Fig 9 Gender Chart



Fig 10 Age Chart



Fig 11 Current level of Education



Fig 12 Nationality of participants



Fig 13 Those interested in online learning

## : Requirements Specifications

* + 1. Functional Requirements Specifications

Functional requirements are the requirements specified by customers which describes what the system must do.

Table 5 Functional Requirements

|  |  |  |
| --- | --- | --- |
| **Req.****No.** | **Description** | **Type** |
| FR-101 | Any device that can connect to the internet can be used | Configurational |
| FR-102 | The application shall have a simple user interface | Functional |
| FR-103 | Customers can view their favourites | Configurational |
| FR-105 | The minimum size for the RAM memory should be 512MB | Configurational |
| FR-106 | Customers can be able to take in user input for English language(reverse translation) | Functional |
| FR-107 | Customers can be able to type in user input for Hyam language | Functional |
|  |  |  |
| FR-109 | The system shall have a voice command (for English only) | Functional |
| FR-110 | The system shall work on all smart phone devices | Functional |
| FR-111 | The past translations done by the customer would be saved | Functional |
| FR-113 | Customers can give suggestions where the translations are wrongor incomplete | Functional |

## : Non-Functional Requirement Specifications

Table 6 Non functional requirements

|  |  |  |
| --- | --- | --- |
| **Req.****No.** | **Description** | **Type** |
| NFR- 01 | When the app is being launched, the application would still be running in the background, unless no internet connectivity or theuser closes the tab | Performance |
| NFR-02 | A better component design, which will be used to initiate betterperformance of the app at peak time | Performance |
| NFR-03 | How customers can have trust in my app and the design | Integrity |

|  |  |  |
| --- | --- | --- |
| NFR-04 | Navigation for the app by users would be made easy | Usability |
| NFR-05 | The app shall keep safe confidential data of the user’s details | Security |
| NFR-06 | The quality of the application’s performance | Reliability |
| NFR-07 | Only the admin should be able to edit the app’s corpus for bothlanguages | Integrity |

## : System Design

In this chapter, there’s demonstration about the system design and also Unified Modelling Language (UML), which is a standard representation implemented for the designing real world objects and systems. There are many UML diagrams, which were implemented as a part of the development process, examples are the Data Flow Diagram, Activity Diagram, Use Case Diagram, Application Architecture and Entity Relationship Diagram.



Fig 14 Application Architecture





Fig 15 Activity Diagram



Fig 16 Use Case Diagram

*In the use case diagram above, the user after entering the web application, would insert the desired text he or she wants to be translated. The text can either be a word or a phrase, after which a click of a button automatically translates the Hyam text to English text, which can be later viewed after a successful translation. Other sections in the app are the share app info, settings page, how to use and view page (this gives a brief knowledge and history of the Hyam people).*



Fig 17 Sequence Diagram

*In the sequence diagram, it has to do with a list of sequential activities that a user does when carrying out a particular task in the system.*

*Here we talk on the translation of hyam to English and back, which is a task on the system. The user firsts enters the desired sentence or word in hyam, clicking translate button, which would prompt the system to translate the system, to English based on the training the system has received. If the system has been adequately translated, would generate the adequate translation in English.*



Fig 18 Entity Relationship Diagram

*In the above diagram, many users can submit many feedback, generating a many-to-many relationship between user and feedback entities. The name of the branch entity is summary, which contains both PKs of feedback and user as well as the feedback and summary.*

*Again, many users can translate many times, producing a many-to-many relationship between users and translation entities. The name of the branch entity is translation details which have the PKs of both entities plus translation and reverse translation.*

## : Summary

The web based system, would be developed and tested using the waterfall methodology. A user- friendly interface would be implemented for novices to be well accustomed to the app. The functional requirements gleaned, would be altered, based on the wants of the customer to improve the application’s efficiency. The programming language that is used in the application development is python, because it is a web app.

# CHAPTER 4: IMPLEMENTATION AND TESTING

## : Overview

This is the penultimate chapter that deals with the implementation of Let’s Talk. Furthermore, it also shows the processes that are needed to attain the processes of the system. This chapter, presents the testing with various testing domains, plus the results after testing is done. The summary of the implementation is: developing the new system, readily present to the target market, supporting as well as maintaining the system. Deployment of the system involves execution of all processes, which are salient to educate the target market, which in this case are locals of the Hyam community, putting the application into production, and all the information are impeccable and available.

After the system is developed, a handful amount of briefs must be accomplished, before installation and deployment. We refer to this process as system implementation. These steps involve the procurement of both hardware equipment, as well as the required software(s) needed to develop the program, getting the users acquainted with the rudimental of the system, setting up of data for the program, installation, testing, start up and finally delivery of system on time.

The salient variance, that is between the system execution and the remaining stages involved in the lifecycle of the program, is that the functions or tasks that were carried out, were done effectively in secure and safe domains, meaning that the issues that would arise from the project, would have infinitesimal to insignificant influence on the operations of the software.

## : Main Features



Fig 22 The Home Page



Fig 23 The about page of the web app



Fig 24 The How to use page of the web app



Fig 25 The settings page of the web app



Fig 26 The share app info page of the web app



Fig 27 The send feedback page of the web app



Fig 28 A screenshot of the training of the model using google colab



Fig 29 A screenshot of the training of the model using google colab



Fig 30 A screenshot of the training of the model using google colab



Fig 31 A screenshot of the training of the model using google colab

## : Implementation Problems

The key banes in this project were time and the coding. The time given to the project by the author was limited, as a result of attending other courses, doing assignments, tests plus examinations. Since the means of collecting information is a questionnaire, given the large amount or the target market, some people may not answer, or answer truthfully the questions, which would influence the data analysis negatively. Again, the system would be unclear and ambiguous to the target market, reducing the efficiency and reliability of the system. Furthermore, they are some functionalities that were not properly implemented, or not in the system, which would appear later in subsequent versions of the program.

## : Overcoming Implementation Problems

Despite the problems encountered, we were able to find remedies to them. We were still able to develop and deliver a program, worthy enough to meet the users’ requirements that were stated in the preliminary stage. We started first, by making time management, so we would be able to work adequately on the other resources to deliver a good program.

For the coding aspect, the bugs found were expunged, by first sifting the project into modules, so as to be able to review the faulty modules and vet the codes in the module(s).

* 1. : Testing Plans (Unit Testing, Integration Testing & System Testing) Fig Test Plan

Users

|  |  |
| --- | --- |
| 1 | Users can translate sentences from Hyam to English |
| 2 | Users can send feedback about the system |
| 3 | Users can reverse translation (i.e. English to Hyam) |

Admin

|  |  |
| --- | --- |
| 1 | Admin can edit the corpus being used in the system |
| 2 | Admin can view the feedback from users |
| 3 | Admin can view |

* + 1. : Test Suite

|  |  |  |
| --- | --- | --- |
| Requirement Number | Description | Type |
| R1 | When launched, the application shall stay running unless there is an intentional shutdown of the application or the platform. | Performance |

* + 1. : Test Traceability Matrix

# Chapter 5: DISCUSSION, RECOMMENDATIONS & CONCLUSION

## : Overview

In this last chapter of my project, the goals of the project would be elaborated here. There are restrictions that have been encountered, when undertaking the project, as well as recommendations and future improvements that would be done to enhance the efficiency, and quality of the project.

## : Objective Assessment

The goals that were achieved for the project are as follows:

* + - The ability to change the project to its quality.
		- Being able to develop the let’s talk application to match this quality
		- Ability to know how to use web applications
		- Skills on developing a software realistically
		- Testing done which on a scale was 90% accurate or successful

## : Limitations & Challenges

* + - **Funds:** Resources in terms of funds when it comes to hosting of the application, registering a domain name plus laptop issues.
		- **Security issues:** No matter how great an app is, if it is not successful is prone to be a target for hackers especially, if sensitive information are required and stored in the app.
		- **Not working with the best performance:** This time, speed is a huge factor in developing apps, so I have to vet everything of the app, including the speed.

## : Future Improvements

The goals that were achieved in this project, are as follows:

* + - Let’s talk would be developed as a mobile application for iOS, android blackberry OS and also the OS for the windows mobile platforms.
		- Let’s talk would be able to ‘reverse translations’. A simple term which means that the input language, can be the output language and vice-versa. This is only for the two languages, however, Hyam and English.
		- The quality for the training, would be improved, by increasing the size of both documents, which would be useful in the long run.
		- Implementation of the translation services offline, at least 30% of the translation.
		- Offline translation of words that are mostly translated on the app.
		- The app would be implemented as a web app. As of now, only the model that has developed.

## : Recommendations

Given the outcomes made when developing the project, we can conclude that the world is morphing metaphorically, into a global village, given the constant improvement(s) in the field of technology. With this in mind, the below recommendations are highlighted:

* + - People should be kept abreast of English to some extent, so when they translate from Hyam to English, would have an idea of the word.
		- Knowing how to use the app, effectively to help the locals with foreigners and vice-versa.

## : Summary

This project has given us the ability to understand, as well as translate words and/or sentences from Hyam to English, which was done using computerized means. The work of the middle man is limited or eradicated, with more transparent translations for both languages.

# REFERENCES

* + - Avramidis, E. & Koehn, P. (2008). Enriching Morphologically Poor Languages for Statistical Machine Translation. *Proceedings of ACL 08HLT: 763-770.*
		- Hsu, J. A. (2014) Error Classification of Machine Translation A Corpus-based Study on Chinese-English Patent Translation. *Translation Studies Quarterly* 18: 121-136
		- Vilar, D., Xu, J., d’Haro, L. F., & Ney, H. (2006). Error Analysis of Statistical Machine Translation Output. *Proceedings of the 5th International Conference of \n Language Resources and Evaluation (LREC). Genoe, pp. 697-702.*
		- Gubler, K. 2015, inWhatLanguage, accessed 20 October, 2020,

<<https://www.inwhatlanguage.com/the-problems-with-machine-translation/>>

* + - Expert System Team 2016, Expert System Team, accessed 13 June, 2020,

<<https://expertsystem.com/natural-language-processing/>>

* + - Colah 2015, Colah, accessed 13 June, 2020, <<http://colah.github.io/posts/2015-08-> Understanding-LSTMs/>
		- Jain, S., 2018, Analytics Vidhya, accessed 12 June, 2020,

<https:/[/www.analyticsvidhya.com/blog/2018/02/natura](http://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-)l[-language-processing-for-](http://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-) beginners-using-textblob/>

* + - No author, Python Machine, accessed 01 August, 2020, <https://www.python- course.eu/neural\_network\_weights.php>
		- Shaikh, F. 2019, Analytics Vidhya, accessed 01 August, 2020,

<https:/[/www.analyticsvidhya.com/blog/2019/01/fundamentals](http://www.analyticsvidhya.com/blog/2019/01/fundamentals-deep-learning-recurrent-)-[deep-learning-recurrent-](http://www.analyticsvidhya.com/blog/2019/01/fundamentals-deep-learning-recurrent-) neural-networks-scratch-python/>

* + - Hayab, J. P. , African Storybook, accessed 19 August, 2020,

<<https://www.africanstorybook.org/reader.php?id=20506>>

* + - Hayab, J. P., 2016, *Basic Hyam Grammar, with ethnographic notes,* Beltina Digital Press, Abuja.
		- Phi, M., 2018, towardsDataScience, accessed 19 August, 2020

<[https://towardsdatascience.com/ilustrated-guide-to-lstms-and-gru-s-a-step-by-step-](https://towardsdatascience.com/ilustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf2) [explanation-44e9eb85bf2](https://towardsdatascience.com/ilustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf2)>

* + - Lanners, Q.,2019, towardsDataScience, accessed 16 June, 2020, < <https://towardsdatascience.com/neural-machine-translation-15ecf6b0b>>

# APPENDICES

## Appendix A: Project Documentation/ Executive Summary

**Student Name:** Kyom Dogo

**Matric Number:** BU/16C/IT/2082

**Project Title:** Design & Implementation of a Neural Machine Translation System (Let’s Talk) for the Translation of Hyam to English

### Introduction:

In Southern Kaduna, the Hyam Community always welcome foreigners every year, due to their friendliness and hospitable nature. They have even welcomed Igbos in the past and when the Hausas wanted to kill them, they hid them and gave them a safe route back to the east. Now with this in mind, the Hyam people would find it hard to communicate easily, due to their little or lack of knowledge of the lingua franca, because of lack of exposure.

The web app Let’s Talk comes in to bridge the gap of interaction between people of Hyam and foreigners who know English but not Hyam. The app does not require any certain kind of phone (e.g. iPhone), made very easy to use, and both parties would have a glimpse and idea of how to speak the other language, as well as understanding it.

### Basic Features:

* + - The app would be a web based app.
		- The system would be able to translate effectively, a sentence from Hyam to English.
		- Functionalities like share app info, settings, would be implemented on the app.

### Project Goals and Objectives:

The main objective is to enable indigenes of the Hyam community to understand as well as communicate English with foreigners and vice versa. Again it would be used a preservation of the language for the later generation of Hyam people.

## Appendix B: Work Plan

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Task Number** | **Task Name** | **Expected Start Date** | **Expected Completion Date** | **Expected Completion in Duration (Days)** | **Deliverables** |
| 1 | Background Study | 7/12/2021 | 7/13/2021 | 1 day | Project Topic(s) |
| 2 | Introduction | 7/13/2021 | 7/17/2021 | 4 days | Chapter One (1) |
| 3 | Review of Literature | 7/13/2021 | 7/21/2021 | 8 days | Reviewed Literature Document |
| 4 | Design Methodology (Analysis/Design) | 7/22/2021 | 8/14/2021 | 23 days | Required Document, Use Case, Application Architecture, ERD |
| 5 | Implementation and Testing | 8/16/2021 | 8/26/2021 | 10 days | Source Code, Testing Reports. |
| 6 | Collation of Draft Project | 8/27/2021 | 8/29/2021 | 2 days | Spiral-Bound Document |
| 7 | Final Submission | 8/29/2021 | 8/31/2021 | 2 days | Hard Cover Bound Document |

## Appendix C: Gantt Chart

12-Jul 17-Jul 22-Jul 27-Jul 1-Aug 6-Aug 11-Aug 16-Aug 21-Aug 26-Aug 31-Aug

Background Study

Introduction

Literature Review

Design methodology (Analysis and Design)

Implementation and Testing

Collation of Draft Project

Final Submission