DESIGN OF AN ENGLISH - OKUN PREPOSITIONAL PHRASE MACHINE TRANSLATION SYSTEM USING RULE BASED APPROACH

BY

TOLORUNTOMI, SAMUEL FEMI

CPE/13/1085

A Research Project Submitted to Department of Computer Engineering,

Faculty of Engineering,

Federal University Oye- Ekiti (FUOYE), Ekiti, Nigeria.

In partial fulfillment of the requirement for the degree of

Bachelor of Engineering (B.Eng) in Computer Engineering

MARCH, 2019.

DECLARATION

This is a result of my own work and has not been copied in part or in whole form from any other source except where duly acknowledged. As such all of previously published works (from books, journals, magazines, the internet and so on) has been acknowledged within the main report to an entry in the list of references.

I agree that an electronic copy or hardcopy of this report may be stored and used for the purposes of plagiarism prevention and detection. I understand that cheating and plagiarism constitute a breach of university regulations and will be dealt with accordingly.

Students Full Name ..... TOLORUNTOMI SAMUEL FEMI .....

Signature and Date..... SAM 10-07-19 .....

CERTIFICATION

This project with the title

DESIGN OF AN ENGLISH - OKUN PREPOSITIONAL PHRASE MACHINE TRANSLATION SYSTEM USING RULE BASED APPROACH

Submitted by

TOLORUNTOMI, SAMUEL FEMI
CPE/13/1085

Has satisfied the regulations governing the award of degree of

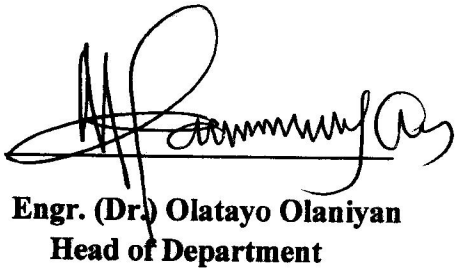
BACHELOR OF ENGINEERING (B.Eng) in Computer Engineering

Federal University Oye-Ekiti, Ekiti State, Nigeria.

10-4-19

Engr. (Mrs) Esan A.O
Project Supervisor

Date



10/4/19

Engr. (Dr.) Olatayo Olaniyan
Head of Department

Date

DEDICATION

I dedicate this project work to the Almighty God and to my parents Mr. and Mrs. Toloruntomi for their support financially and morally.

ACKNOWLEDGEMENT

My sincere thanks and appreciation to the Rock of Ages, words alone cannot express my appreciation to Him; He is a wonder working God.

A special appreciation goes to my Project supervisor, Engr. (Mrs) Esan for the time and energy expended on me during the process of putting together this project work, you are wonderful ma, I cannot forget you in a hurry.

My appreciation also goes to other members of staff of computer Engineering Department.

With a sincere heart and all humility, I appreciate deeply my parents Mr. and Mrs. T.F Toloruntomi for their financial, moral and spiritual support during my sojourn in this Institution.

My thanks and appreciations to my Uncle and Aunt; Mr. & Mrs. Moses Onivehu and also to my siblings, Mrs. Grace S. Adeyemi (nee Tomi), Engr. Stephen Tomi, Pastor Ayo Tomi of the Winners Chapel International, Miss Elizabeth O. Tomi, and of course my little nephews, King David Adeyemi, Asher and Tolu Tomi whose cries will always wake me up in the middle of the night to arise and work on my project work. God bless you all.

I cannot forget Engr. Adeyemi B., Mr Olubiyo T. Shawn, Engr. Ayeni Taiyelolu, Engr. Isiaq. F, Engr. Okunlola S., Engr. E.K Odole, May God bless you real good!

ABSTRACT

The predominance of English language over Okun language in almost all human endeavour is threatening the existence of Okun language and even Okun culture. It is therefore necessary to employ modern day processing tools for the language to catch up with the technological growth the world is experiencing.

This project work focuses on the design of a bidirectional English - Okun prepositional phrase machine translation system using rule based approach. The system was designed using Python Programming language and PyQt5 was used for the design of the user interface while Py2exe was used to compile the code into executable application. The system was evaluated using human judgement based on orthography accuracy.

Results show that the average orthography accuracy score for the expert is hundred percent (100%), experimental subject respondent score is about fifty eight percent (58%) and the score for the developed machine translator is ninety three percent (93%).

In conclusion, it was observed that most indigenous people are not good at writing Okun language and also find it hard to separate Okun dialects from the Yoruba language. Hence it is recommended that schools in the western part of Kogi state incorporate this software into their academic system to enable students learn more about the language, and the government should encourage researchers in the field of computational linguistics by funding researches.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

MT – Machine translation

DMT – Direct Machine Translation

RBMT – Rule based Machine Translation

SL – Source Language

TL – Target Language

SMT – Statistical Machine Translation

NP – Noun Phrase

PP – Prepositional phrase

VP – Verb phrase

ADJP – Adjectival Phrase

N - Noun

PRE - Preposition

ADJ - Adjective

DET - Determiner

APỌO - Àpólà ọ̀rọ̀ Orúkọ̀

APTK - Àpólà ọ̀rọ̀ Atọ́kùn

APỌI - Àpólà ọ̀rọ̀ ìṣe

APỌA - Àpólà Òrò Àpònlè

ATK - Ọ̀rọ̀ Atòkùn

ỌE - Òrò Orúkọ

AỌO - Arópò Òrò Orúkọ

ỌA - Òrò Àpònlé

AIỌO - Asàpejúwé Ìlò ọ̀rọ̀ Orúkọ

CHAPTER ONE

INTRODUCTION

1.1 Background Study

Language is an efficient medium of communication which expresses the human mind (Abiola, Adetunbi and Oguntimilehin, 2015). Human language is purposively to communicate ideas, emotions, feelings, desires, to co-operate among social groups, to exhibit habits etc. which can be translated along a variety of channels (Banjo & Jibowo, 2011) . There are over 6,800 living languages in the world which reflects the scope of linguistic and cultural diversity.

Access to information written in another language is of great interest and the means of sharing information across languages is translation, therefore creating tools for translating from one language to another is very crucial to human development. Without translation, there can be no communication, except among those who share a common language and many voices will not be heard without this critical function.

Translation is critical for addressing information inequalities. A study conducted by Common Sense Advisory on behalf of Translators without Borders finds that translation is critical for the public health, political stability, and social wellbeing of African nations (Banjo & Jibowo, 2011). However translation was originally carried out by human translators but its limitations such as high cost, lower speed of translation and insecurity of confidential information led to the development of machine translators.

Machine Translation (MT) can be defined as a subfield of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another. The history of machine translation can be traced back to the pioneers and early systems of the 1950s and 1960s, the impact of the ALPAC report in the mid-1960s, the revival in the 1970s, the appearance of commercial and operational systems in the 1980s, research during the 1980s, new developments in research in the 1990s, and the growing use of systems in the past decade. This research would develop a machine translator system for English-Okun using a rule-based approach.

1.2 Statement of the Problem

Okun culture is gradually going into extinction because the total dominance of English language over Okun language in almost all human endeavour is a challenge. Hence, there is need for modern day processing tools for the language to catch up with the technological growth the world is experiencing.

Also, translation was traditionally carried out by human translators but its limitations such as high cost, lower speed of translation, undependability and insecurity of confidential information as well as lack of in depth understanding of a language (Oladosu, et al., 2016) led to the development of machine translators.

Therefore, an English-Okun machine translator would be designed in this research to overcome the shortcomings of human translators.

1.3 Aim and Objectives

The aim of this project is to develop a ruled based system for English-Okun machine translation.

The specific objectives are to:

i. develop a machine translation system using rule based approach

ii. implement the system for English-Okun machine translation.

iii. Evaluate the system using Mean Score Opinion (Human Judgement).

1.4 Scope of Study

In this research, an MT system was developed using rule based approach. The system was bi-directional because English language was translated to Okun language. The data was obtained from locally spoken words and the performance of the system was evaluated using the Mean Opinion Score (tone mark and spellings).

1.5 Research Methodology

The Project methodology include:

i. Creating the Database: the data for this work was extracted from locally spoken words and stored in a database. The phrases were analyzed into their part of speech (POS) and the database designed, the parts of speech were classified into their different grammatical functions.
ii. System software training: the software was trained to understand the grammatical rules of translation.
iii. Designing of a parser: the Natural Language Toolkit (NLTK) parser was used to parse the data and test all the rules used as it affects each sentence.
iv. Python programming language: This is the core programming language that was used in designing the system.
v. Human Judgement was used to evaluate the system.

1.6 Significance of Study

The significance of Machine translation cannot be overstressed, with the world being faced with rapid growth and globalization. Thus the need for translation from and to different languages is of utmost necessity. MT systems has the capacity to translate with high speed and provides quality outputs, thus saving human the stress and time of poring on translating books or looking for human translator.

Also, machine translation is comparatively cheap and faster. Another factor which makes Machine translation favourable is the issue of Confidentiality. Giving sensitive data to a human translator might be unsafe while with machine translation your information is protected. With machine translation system with you one can access it anywhere.

In addition, machine translation has been integrated in many applications which include localization of software and speech translation. MT can be applied in software localization by making available supporting documentation for new software in many languages. Other applications includes: information retrieval (multilingual access to information) and extraction

(data and text mining). It can also be applied in government organizations for the translation of internal documents and assisting administrators in composing texts in non-native languages. Applicable also in Transliteration, Summarization, Question answering (cross-language) and authoring software.

Moreover, MT systems can be applied in broadcasting transcripts, distance education, language teaching, doctor-patient communication and emergency services. It can also be used in chatrooms, for social networking for example Facebook, Photocopying, Scanning for example pen scanner), Camera for example menus and road signs, Spectacles, robotics as in spoken commands and Universal Translator.

CHAPTER TWO

LITERATURE REVIEW

2.1 Language

Language is an efficient medium of communication which expresses the human mind (Abiola, Adetunbi and Oguntimilehin, 2015). Human language is purposively to communicate ideas, emotions, feelings, desires, to co-operate among social groups, to exhibit habits etc. which can be translated along a variety of channels (Banjo & Jibowo, 2011).

Language interacts with every aspect of human life in society, and it can be understood only if it is considered in relation to society. Because each language is both a working system of communication in the period and in the community wherein it is used and also the product of its history and the source of its future development, any account of language must consider it from both these points of view. (Robert & David , 2018).

2.2 Machine Translation

The term 'machine translation' refers to computerized systems responsible for the production of translations with or without human assistance (Hutchins, 1995).

Machine Translation can be defined as a subfield of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another (Arnold *et al.*, 1994). It is an area of applied research that draws ideas and techniques from linguistics, computer science, Artificial Intelligence (AI), translation theory and statistics. The goal of machine translation is to develop a system that accurately produces a good translation between human languages.

The translation of natural languages by machine was first dreamt of in the seventeenth century, but became a reality in the late twentieth. Computer programs now produce translations that is an ideal one when compared to human translator, this ranges from translations of technical manuals, scientific documents, commercial prospectuses, administrative memoranda, medical reports. Machine translation is not primarily an area of abstract intellectual inquiry but the application of computer and language sciences to the development of systems answering practical needs (Hutchins, 1995).

The history of machine translation can be traced from the pioneers and early systems of the 1950s and 1960s, to the impact of the ALPAC report in the mid-1960s, the revival in the 1970s, the appearance of commercial and operational systems in the 1980s, research during the 1980s, new developments in research in the 1990s, finally to and the growing use of systems in the past decade.

2.3 Approaches to machine translation

There are various approaches to machine translation. The approaches are classified into two: single and hybrid.

2.3.1 The Single Approach

Single approach to machine translation is the use of only one method to translate from one natural language to another. Under this approach are the rule-based, direct-based, corpus-based and knowledge based approaches to machine translation. (Oladosu, *et al.*, 2016)

2.3.1.1 Rule-Based Machine Translation (RBMT)

Rule Based Machine Translation (RBMT) has much to do with the morphological, syntactic and semantic information about the source and target language (Abiola, Adetunbi and Oguntimilehin, 2015). Linguistic rules are built over this information. Also millions of bilingual dictionaries for the language pair are used. RBMT is able to deal with the needs of wide variety of linguistic phenomena and is extensible and maintainable. However, exceptions in grammar add difficulty to the system. Also, the research process requires high investment. This approach is capable of handling word-order problems and tracing parse errors using linguistic knowledge.

The objective of RBMT is to convert source language structures to target language structures. The methodology could have several approaches as shown in figure 1:

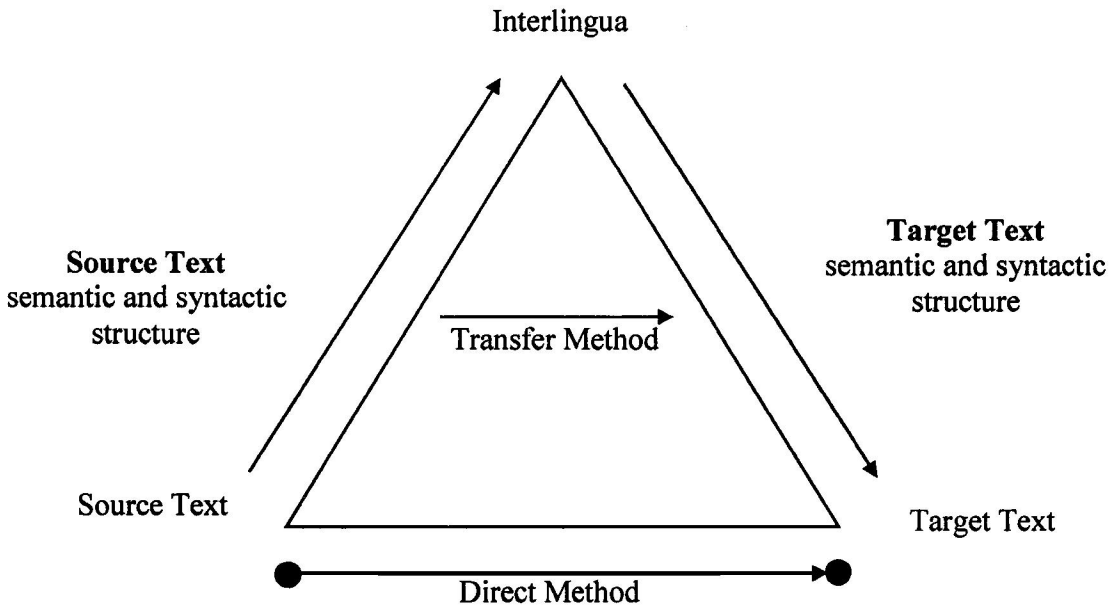


Figure 2.1: Different Method of Rule Based Machine Translation (Tripathi & Sarkhel, 2010)

- **Direct-Based Machine Translation (DBMT)**

This approach is the most elementary of all approaches to machine translation. Words of Source Language are translated without passing through an additional/intermediary representation.

- **Transfer based**

In this, source language is transformed into an abstract, less language-specific representation. An equivalent representation is then generated for the target language as well using bilingual dictionaries and grammar rules. There exists three main components for the system, they include:

Analysis: The source text is analyzed based on linguistic information such as morphology, part-of-speech (POS), syntax, semantics, etc. Heuristics as well as algorithms are applied to parse the source language and derive the syntactic structure (for language pair of the same family) of the text to be translated; or the semantic structure (for language pair of different families).

Transfer: The syntactic/semantic structure of source language is then transferred into the syntactic/semantic structure of the target language.

Synthesis: This module replaces the constituents in the source language to the target language equivalents. This approach, however, has dependency on the language pair involved. Thus, two independent monolingual dictionaries needed. Also, there are different representations for different languages.

- **Interlingua**

Interlingua is a combination of two Latin words Inter and Lingua which means between/intermediary and language respectively. It is an essential part of a branch called Inter-linguistics. Interlingua aims at stimulating linguistic homogeneity across the globe. In Interlingua, source language is transformed into an auxiliary/intermediary language (representation) which is independent of any of the languages involved in the translation. The translated verse for the target language is then derived through this auxiliary representation. Hence, only two modules i.e., analysis and synthesis are required in this type of system.

2.3.1.2 Corpus-Based Approach

Classified under this are the: example-based machine translations (EBMT) and statistical machine translation (SMT).

Example-based approach: also known as Memory based translation is based on recalling/finding analogous examples (of the language pairs), previous translation examples is used to generate translations for an input provided (Tripathi and Sarkhel, 2010).

Statistical machine translation (SMT): In this, statistical methods are applied to generate translated version using bilingual corpora. This approach incorporates the use of machine learning methods. It applies a learning algorithm on large body of previously translated text, known as a parallel corpus, and then the learner is able to translate previously unseen sentences.

2.3.1.3 Knowledge-Based Approach

Knowledge-based approach employs extensive semantic and pragmatic knowledge in machine translation. It involves the ability to reason about concepts (Sharma, 2011).

2.3.2 Hybrid Approach

This approach is a hybridization of the statistical approach and one or more MT approaches. Typically, hybrid MT approach combines the statistical method and the rule-based approach. The approach include: word based model, phrase based model, syntax based model and forest based model (Oladosu, *et al.,* 2016).

2.3.2.1 Word-based Models

Word-based models as designed by Brown *et al* (1993) models the lexical dependencies between single words. Under this approach is the IBM word-based models which were a breakthrough in pioneering the work in SMT, but their design is constraint to model the lexical dependencies between single words only. They fail when they are required to model long-distance re-ordering of words. Hence, phrase-based models were introduced (Oladosu, *et al.,* 2016).

2.3.2.2 Phrase-based Models

In Phrase-based model, Phrases can be any substring and they allow local re-orderings, translation of short idioms, or insertions and deletions that are sensitive to local context. They are thus a

simple and powerful mechanism for machine translation as a translation of reputable quality itself is never word for word. They were primarily introduced to alleviate the short coming of the word-based models by the introduction of phrases as the basic translation unit (Och, 2003). Though, phrase-based models help to alleviate shortcomings of the word based models, they still have their weaknesses one of which is the inability to model long distance reordering of source words. Thus the introduction of syntax-based models (Oladosu, *et al.,* 2016).

2.3.2.3 Syntax based model

To model long distance reordering of source words syntax based models were introduced.

Syntax is the hierarchical structure of a natural language sentence, syntax-based models can be divided into two broad categories based on the type of input: the string-based systems and tree-based systems (Oladosu, *et al.,* 2016).

2.3.2.4 Forest-Based Translation

Forest-based translation is a compromise between the string-based and tree-based methods because it combines the advantages of both methods. Forest based translation encourages faster decoding and alleviates parse errors. Informally, a packed parse forest, or forest in short, is a compact representation of all the derivations (i.e., parse trees) for a given sentence under a context-free grammar (Billot and Lang, 1989; Mi *et al.,* 2008). Forest based machine translation mainly extends the tree-string model to forest to string.

Forest-to-string translation is an extension of the tree-to-string model because it uses a packed parse forest as the input and outputs a string (Mi *et al,* 2008).

Research shows that the method provides improvement from 0.6 up to 1.0 point measured by (Ter–Bleu)/2 metric.

Figure 2.0 shows a summary of various approaches to machine translation.

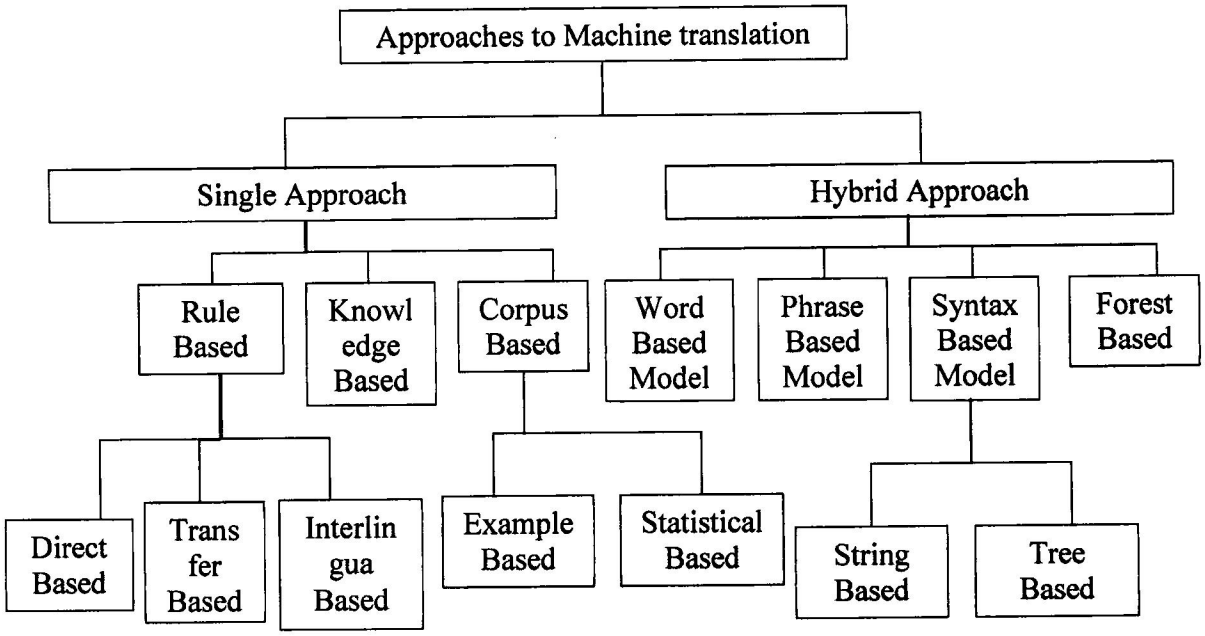


Figure 2.2: Approaches to Machine Translation [Adapted from: (Tripathi & Sarkhel, 2010)]

2.4 Evaluation of Machine Translation Systems

Traditionally, human judgment is used in evaluating MT systems based on two main criteria: 'adequacy' and 'fluency'. Human judgment of the MT output is expensive and subjective therefore, automatic evaluation measures are a necessity (Adeyanju *et al*, 2010). Outlined below are the various methods used in evaluating MT systems.

2.4.1 BLEU (bilingual evaluation understudy)

BLEU is used in automatic calculation of translation quality. It was introduced by Papineni *et al* (2002) to address evaluation problem by comparing the system output against a reference translation of the same source text. The effectiveness of the BLEU metric has been demonstrated by showing that it correlates with human judgment. Two properties of the BLEU metric are: the reliance on higher n-grams and the brevity penalty (BP). Given the precision Pn of n-grams, of size up to N (usually N=4), the length of the test set in words(c) and the length of the reference translator (r) in words, the equation is computed as in 2.1 and 2.2:

$$\boldsymbol{Bleu} = \mathbf{Bp}.\exp\left(\sum_{n=1}^{4} \boldsymbol{logPn}\right) \qquad \mathbf{2.1}$$

$$\boldsymbol{Bp} = \min\left(1, e^{1r/c}\right) \qquad \mathbf{2.2}$$

2.4.2 WER (Word Error Rate):

The word error rate is based on the Levenshtein distance. The WER is computed as the minimum number of substitution, insertion and deletion operations that have to be performed to convert the generated translation into the reference translation. In the case where several reference translations are provided for a source sentence, the minimal distance would be calculated to this set of references as proposed in (Nieben *et al,* 2000). WER measures translation errors. A shortcoming of the WER is that it requires a perfect word order.

2.4.3 PER (Position-Independent Word Error Rate):

In order to overcome the shortcoming of the WER, the position independent word was introduced. Words that have no matching counterparts are counted as substitution errors, missing words are

deletion and additional words are insertion errors. The PER is a lower bound for WER and measures translation errors (Oladosu, *et al.,* 2016).

2.4.4 NIST (National Institute of Standards and Technology):

This evaluation method or approach got its name from the National Institute of Standards and Technology (NIST), is a slight modification of BLEU with the main difference between them being the method of their averaging n-gram scores. While BLEU relies on geometric mean, NIST performs an arithmetic mean. Another unique feature of NIST is its calculation of weights based on reliance upon n-grams which occurred less frequently, as this was an indicator of their higher informativeness.

2.4.5 ATEC:

ATEC is a metric which is claimed to have addressed word order issues. Here, the evaluation is based on explicit assessment of word choice and word order is done while comparing MT outputs and human references. Word choice is assessed by matching surface lexicons, stems and synonyms. Word order is quantified by calculating difference in word position and word sequence between MT output and reference translation. Like meteor, this metric also calculates precision, recall and F-Measure.

2.5 Structure of English and Okun Language

Illustrated below are the English and Okun re-write rules. The list of acronyms and Okun equivalents is shown in Table 1.1. These are the acronyms which would be used to replace the English acronyms in the Okun section.

Table 1.1: List of Acronyms and Okun Equivalents (Eludiora & Atolagbe , 2016)

English	**Okun**
NP	Àpólà ọ̀rọ̀ Orúkọ̀ (APỌO)
PP	Àpólà ọ̀rọ̀ Atọ́kùn (APTK)
VP	Àpólà ọ̀rọ̀ ìṣe (APỌI)
ADJP	Àpólà Òrò Àpònlè (APỌA)
PRE	ọ̀rọ̀ Atòkùn (ATK)
N	Òrò Orúkọ (OO)
PRN	Arópò Òrò Orúkọ (AỌO)
ADJ	Òrò Àpònlé (ỌA)
DET	Asàpejúwé Ìlò ọ̀rọ̀ Orúkọ (AIỌO)

The phrase grammar is used to describe the relationship between the sentence or phrase words. English and Okun sentence structures are explained below with the re-write rules rightly explaining how phrases are derived from noun and verb phrases. The two phrases are realized from the sentence.

The re-write rules according to the Okun sentence structure below showed that Okun is head-first in the Noun phrase (NP) structure while English is head-last in a Noun phrase (NP) structure, (this

holds except for when the determinant is "the"). For example, 'those boys' is (DetN),'Omode-kunrin ghoin', that is, (NDet).

Also in accordance with the Okun Prepositional Phrase structure below, Okun is head-last in the Prepositional phrase (PP) structure, English is also head-last in the Prepositional phrase (PP) structure. For instance, beside the girl is (PDetN), labu ghin omoburin, that is, (PDetN). The position of the preposition does not change in the two languages.

English Sentence Structure

Rule 1 S ===>NPVP
Rule 2 NP ===>DetN
Rule 3 NP ===>DetAdjP
Rule 4 NP ===> PP
Rule 5 AdjP ===>AdjNP
Rule 6 VP ===>VNP
Rule 7 PP ===>PrepNP

Okun Sentence Structure

Rule 1 S ===>NPVP
Rule 2 NP ===>NDet
Rule 3 NP ===>NAdjP
Rule 4 NP ===> PP
Rule 5 AdjP ===>NAdj
Rule 6 VP ===>VNP
Rule 7 PP ===>PrepNP

This is the same with Yoruba sentence structure. (Eludiora & Atolagbe , 2016)

Rule 7 explains the prepositional phrase which is the focus of the project. That is,

PP ===>PreNP.

This is derived from the whole sentence, the PP re-write rules were designed for the two languages.

The PP has six re-write rules each of the two languages, these will be shown below:

Rule 1 shows that PP is produced from noun phrase and PP can produce prepositional and noun phrase.

English Prepositional phrase structure

Rule 1 NP ===>PPNP

Rule 2 PP ===>PRENP

Rule 3 NP ===>ADJPNP

Rule 4 ADJP ===>ADJNP

Rule 5 NP ===>DETNP

Rule 6 NP ===> N

Okun Prepositional phrase structure

Rule 1 NP ===>PPNP

Rule 2 PP ===>PRENP

Rule 3 NP ===>NPADJP

Rule 4 ADJP ===>NPADJ

Rule 5 NP ===>NPDET

Rule 6 NP ===> N

For example, since that afternoon, láti (niati) ghin ọhọn where ni becomes l. This phrase can be tokenized as follows:

English [on the house]

PP===>Pre NP

NP===>Det N

PRE=➔since

DET=➔that

N➔afternoon

Okun [láti ghin ọhọn]

APTK::=ATKAPOO

APOO::=AIOOOO

ATK::=láti

AIOO::=ghin

OO::=ọhọn

2.5.1 Re-write Rules

The rules that will guide the system design are;

Rule 1: A prepositional phrase (PP) consists of a preposition and noun phrase (NP). In the case of Okun language, noun (QO) comes before determiner (AIQO). For example,

SL: since<PRE>this<DET>afternoon<N>.

TL: láti<ATK>ohon<QO>ghin<AIQO>

Rule 2: A determiner (if it does exist) must precede an adjective and a noun in English Language, but reverse is the case in Okun language. For example,

SL: This<DET>wicked<ADJ> boy<N>.

TL: ọmọkùnrin<OO>biburu<ỌA>ghin<AIỌO>.

2.6 Related Works

Using a single approach, machine translation has received considerable amount of research attention.

Oladosu and Olamoyegun (2012) developed a Yoruba-English language translator for doctor-patient mobile chat, with their motivation in rooted in the need to improve rural-urban health care by reducing communication barrier between semi- illiterate patients and highly educated medical personnel who are of different ethnic background. Results show that the application has a high degree of novelty and relevance with about 60% and 80% scores respectively.

Folajinmi and Omonayin (2012) developed a statistical machine translation (SMT) for English-Yoruba translation. With their motivation rooted in the need to find solutions to the problems of language barrier in the world and therefore provided tools to tackle the problem of language translation between Yoruba (a Nigerian language) and English language.

Kumar and Kumar (2013) proposed a Statistical Machine Translation Based Punjabi to English Transliteration System with their motivation rooted in the need to break communication barrier between Punjabi native speakers and those who do not understand Punjabi but English. Results revealed that the system has 97% accuracy when tested on 2000 words and trained on about 15000 words.

Eludiora (2014) developed an English to Yoruba Machine Translation system using rule-based approach, with his motivation rooted in the need to contribute to knowledge in machine translations by experimenting with an African language.

Agbeyangi, Eludiora and Adenekan (2015) developed a transfer Rule-Based Machine Translation system. With motivation rooted in the need to computerize Yoruba language due to its popularity.

Results revealed that the system outperforms Google translator which was used as the baseline. Although, these works contributed to efforts geared at finding solution to the problems of language barrier in the world, they are unable to achieve satisfactory performance because they are unsuitable for large scale application.

Sadiat (2013) built an Arabic-French phrase–based machine translation system with her motivation rooted in the need to build the first Arabic-French phrase–based machine translation system.

Sangeetha, Jothilakshmi and Kumar (2014) developed a Hybrid MT system for English to Tamil. They were motivated by the need to improve on English to Tamil machine translation. Experimental results show that the proposed approach competes with the machine translation methods reported in the literature and it provides the best translated output in each language with NIST score of 0.8963 and BLEU score of 0.7923.

Abiola, Adetunbi and Oguntimilehin (2015) proposed a hybrid approach to English -Yoruba machine translation. They were motivated by the need to improve English to Yoruba machine translation by building a hybrid MT system. Although, these systems improve machine translation greatly in that they are suitable for large scale application, their shortcoming is structural ambiguity in translation.

Ayegba and Osuagwu (2015) developed a rule-based machine translation system that translates English sentences to Igala language. A word-sense model was incorporated in the system which disambiguate sentences before passing it to the translator to further improve the output. Results show that the system is 68.2%adequate as a machine translator.

Eludiora, Agbeyangi, and Fatusin (2015) developed an English to Yorùbá Machine Translation System for Yorùbá Verbs' Tone Changing, the developed system employed rule based approach to handle the translation of tone changing verbs.

Eludiora and Elufidodo (2016) developed a system which considered the translation of English language verbs' group to Yorùbá language verbs' group. The study considered the verb group issue among different issues that affect English to Yorùbá machine translation (EYMT) system.

Ayegba, Ugbedeojo and Philip (2016) develop a rule based language processor that can accept input text in Igala language and automatically translate same to English language. Result shows that the system is 81.2% accurate.

(Esan, Omodunbi, Olaniyan, & Olaleye, 2018) developed a rule based system which translated adjectival phrases in English language to Yoruba language. Result shows that the system was 95.5% accurate.

Quite a lot of works were done so as to improve hybrid machine translation, this was achieved by combining more than two approaches. These works include:

Germann (2012) described a syntax-aware phrase-based statistical machine translation model for German-to-English translation. With his motivation rooted in the need to improve German-English translation, the experimental results for the system was a BLEU score of about 0.15.

Mehay and Brew (2012) presented a Combinatory Categorial Grammar (CCG) Syntactic Reordering Models for Phrase-based Machine Translation. They were motivated by the need to improve machine translation from Urdu to English with improvement when compared to a baseline LR system.

Zhang *et al* (2012) proposed a forest-based tree sequence to string model for Chinese- English translation. They were motivated by the need to integrate the strengths of the forest-based and the tree sequence-based modeling methods. Experimental results show that the system significantly outperforms a baseline system with improvement.

Hence, research revealed that a hybrid machine translation approach improves the fluency and adequacy of MT output greatly especially when more than two approaches are combined.

CHAPTER THREE
METHODOLOGY

3.1 Approach

This project focuses on English-Okun Prepositional Phrase translation using the Rule-Based Approach of Machine Translation (RBMT).

3.2 Requirement Analysis/Design Specification

The requirements and specifications of the software are as follow:

i. to present a simple application GUI to the user;

ii. to allow the user to enter Prepositional phrases in both English and Okun language;

iii. to formulate a grammar for the English and Okun prepositional phrase using phrase structure rewrite rule as long as the phrases are within the domain covered; and

iv. to translate and output the equivalent meaning of the sentences displayed in standard Okun/English language.

3.3 Development tools

The main tools utilized for the success of the project are:

i. Python programming language: this was used as the core programming language for the application development. Python Programming Language was chosen because it has API for natural language processing.

ii. PyQt5: this was used for the design of the application GUI.

iii. JFLAP: is a package of graphical tools which can be used as an aid in learning the basic concepts of Formal Languages and Automata Theory. This was used to test the re-write rule and grammar using parsing tree.

iv. py2exe: a Python Distutils extension which converts Python scripts into executable Windows programs, able to run without requiring a Python installation.This wasused to compile the python codes (.py) to an executable file (.exe).

3.4 Architecture of the System Designed

The architecture of the system as shown in figure 3.0 includes four main components, a parser, an analyzer, a transformer and a generation component.

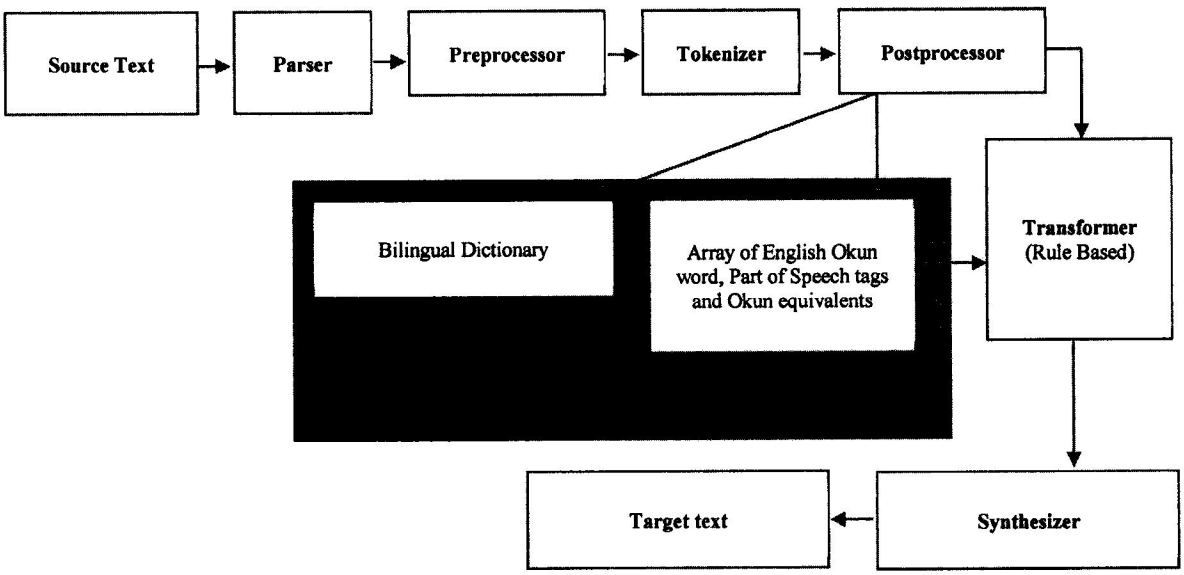


Figure 3.1 Architecture of the English-Okun Translator

The system is made up of the following;

1. **Source Text**: this refers to the original text to be translated.
2. **Parser**: (algorithm) will produce a syntactic structure for a given input, it formats the text in a way that makes it useful for the system. (Parsing simply put, means analyzing, explaining or describing) is used to understand the syntax and semantics of any source language sentence confined to the grammar. Parsing entails the automatic process of analyzing of texts according to any grammar. A Parser would parse the words in the database.
3. **Preprocessor**: the pre-processor counts the number of words in the given prepositional phrase input and declare three arrays of the size of the number of words for use by the other modules.

4. **Tokenizer**: the Tokenizer performs the text processing operation which involve breaking up of raw text into words and it recognizes a word whenever a space is encountered which signifies the end of the word as well as put each of the tokens (words) into one of the arrays created by the preprocessor.
5. **Postprocessor**: The Postprocessor opens the full-form bilingual dictionary for each of the tokens in the array, retrieves its part of speech (pos tag) and target language equivalent and store the retrieved pos tags and target equivalents in the remaining two arrays respectively. Thus arrays of source word (the tokens), POS tags and target equivalents will be generated by the postprocessor.
6. **Bilingual Dictionary**: this is the database where all words are housed with their respective part of speech tags.
7. **Transformer**: The transformer consists of a set of transformation rules which will be used to build the target language equivalent of the input source sentence.
8. **Synthesizer**: The syntactic/semantic structure of source language is then transferred into the syntactic/semantic structure of the target language.
   This module replaces the constituents in the source language to the target language equivalents.
9. **System GUI**: this is an interactive interface that connects the User and the translator, it provides space for the user to easily type-in what they intend to translate.
10. **Database**: Database is the housing unit used in storing the data used for the translation. The phrases were broken down into their part of speech (POS). The database was designed by classifying all the parts of speech into their different grammatical functions.

3.5 System Design

The developed system was designed using the following: Phrase Grammar, rewrite-rules, system design, database design, and system software design.

3.5.1 System Design

The system design was implemented using all principles and rules that govern the translation of the source language to the target language.

In conformance to requirement and design specification, the system users would be allowed to enter a text in the source language, after which the texts are broken into token (lexemes). The re-write rules are designed and developed. The token is then patterned according to the re-write rules. The lexemes are fetched from the database. The outputs of the system (the translated text) was displayed through the Graphical User Interface (GUI).

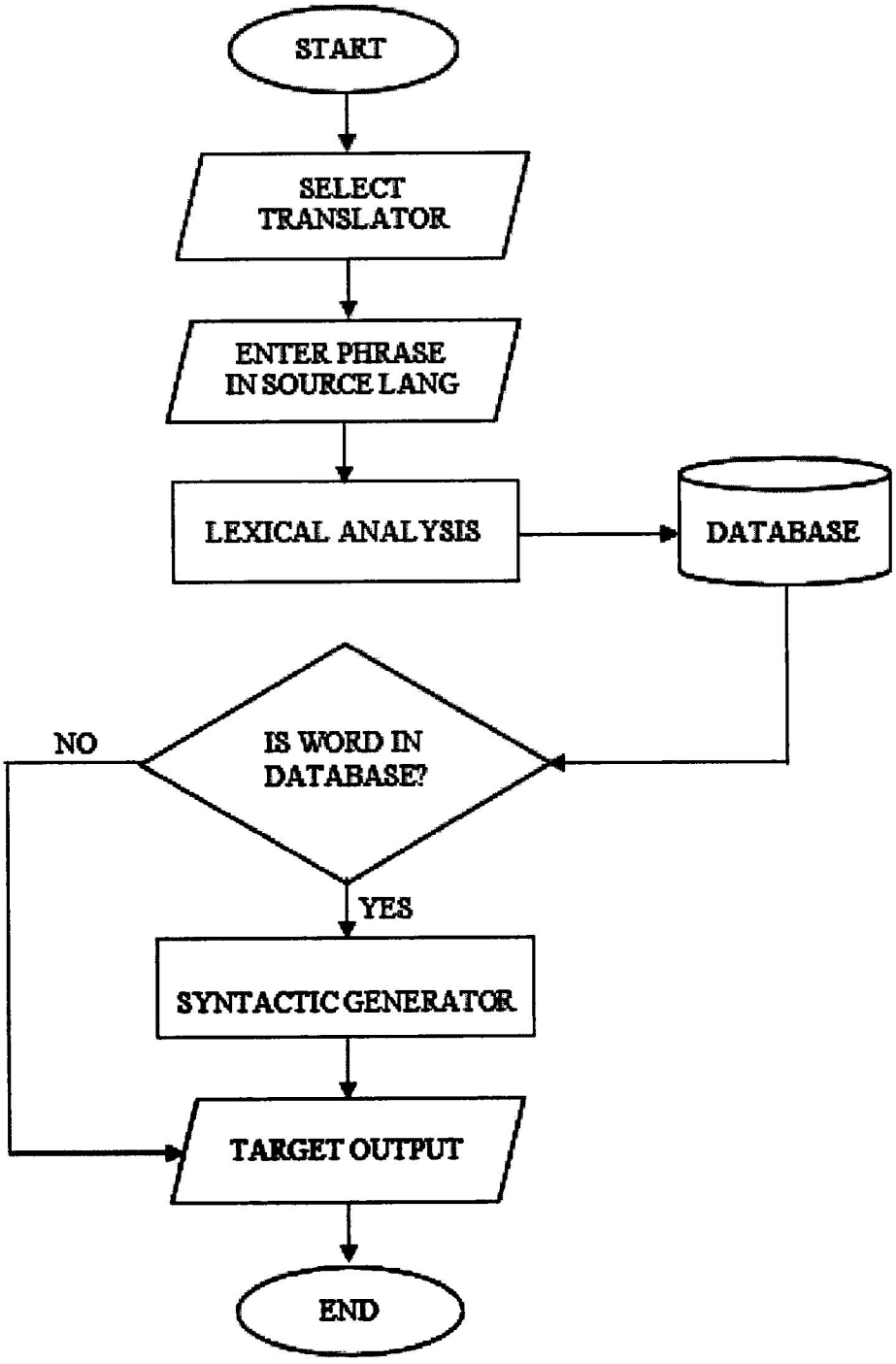


Figure 3.2: Flowchart for the system.

3.5.2 Parse Tree

The parse trees show the pictorial view of the prepositional phrase. The essence of re-write rules is to provide structural view of the given PP before the real coding. One important thing to note is that, the noun (OO) and Preposition (APTK) swapped with the determiner. It shows that Okun language is head first and English language is head last (this holds except for when the determinant is "the").

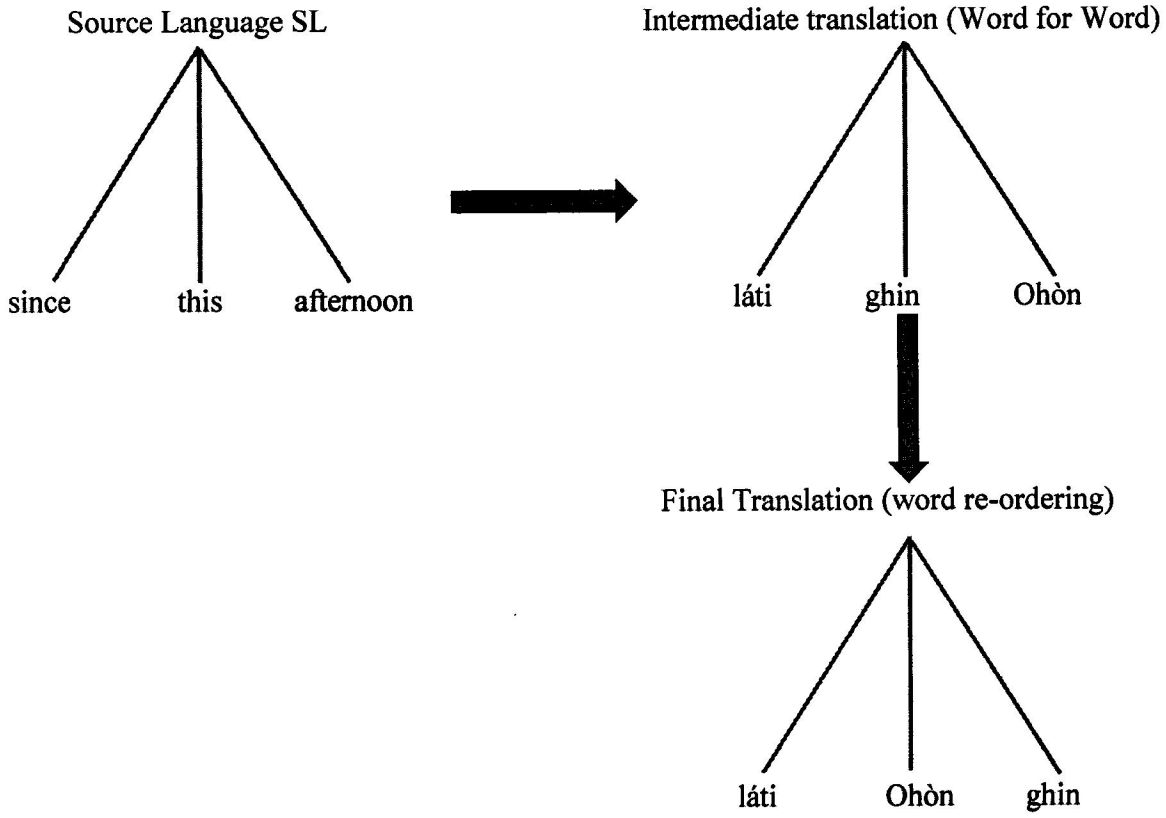


Figure 3.3: Prepositional Phrase Translation Process Abstraction

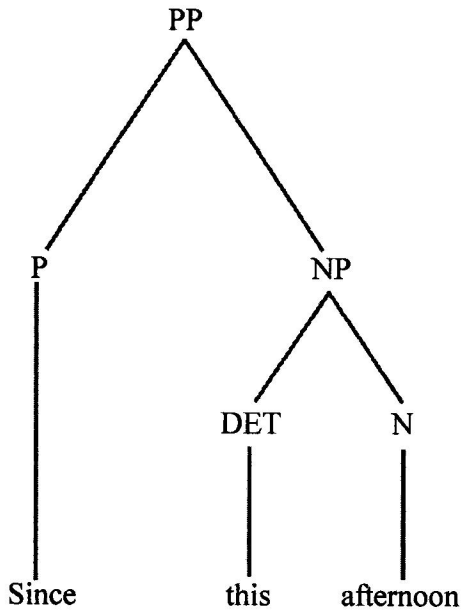


Figure 3.4: Parse tree for an English prepositional phrase

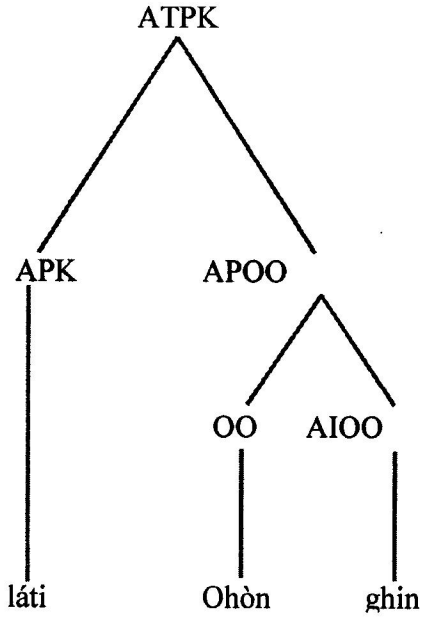


Figure 3.5. Parse tree for an Okun prepositional phrase

3.5.3 Re-write Rules Testing

JFLAP was used to determine the correctness of the re-write rules designed above. Figures 3.6 and figure 3.7 show the outputs of the SL and TL prepositional phrases. The two language grammars re-write rules was used for the JFLAP environment. This is to determine whether the PP strings will be accepted or not. The nodes will then be generated following the re-write rules provided. If the re-write rules follow the sequence of the language grammars it will accept the string being tested, otherwise it will be rejected. Many PP can be tested for the two languages.

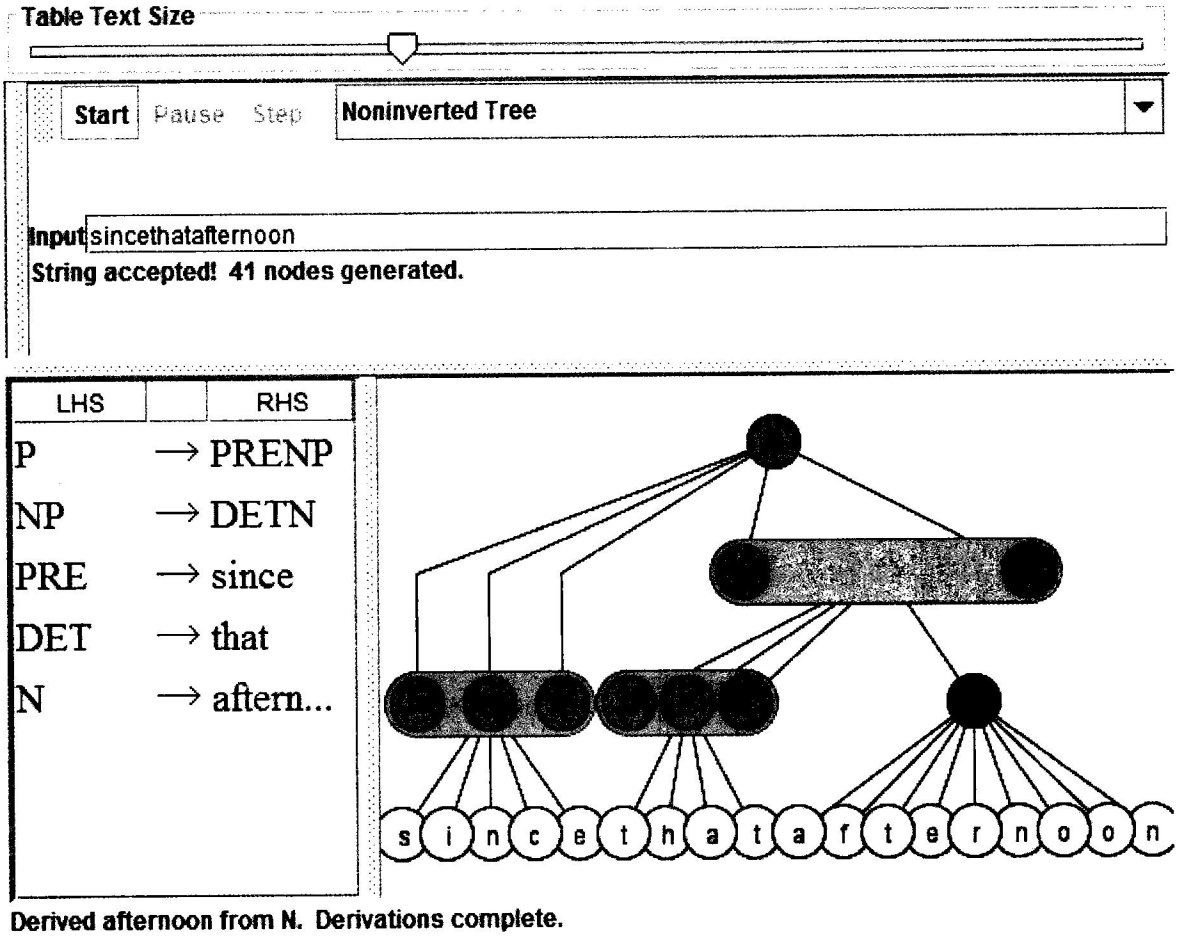


Figure 3.6: English Prepositional Phrase Rewrite Test

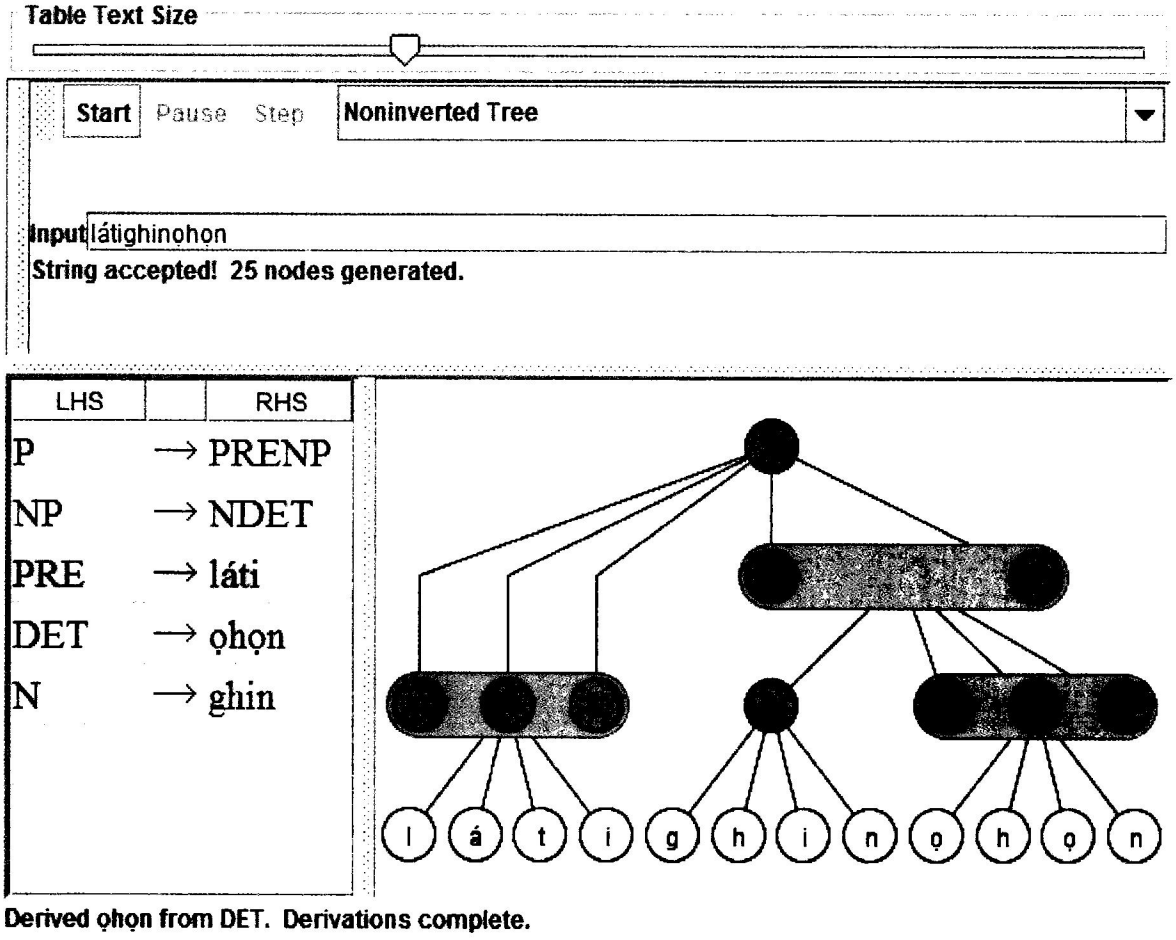


Figure 3.7: Okun Prepositional Phrase Rewrite Test

3.5.4 The Prepositional Phrase Translation Process Model

Figure 3.8a is the state diagram for the English PP translation process model. Figure 3.8a describes possible phrases that can be translated from the source language (SL) to the target language (TL). These are possible translation combinations for the English preposition phrase. They are: PREDETN and PREDETADJN. It means that there can be,for example,before the church and before the Catholic Church. Figure 3.8b is the state diagram for the Okun language PP translation process. Figure 3.8b shows possible combinations of prepositional phrases that can be accepted by the TL. They are: ATKOOAIOO and ATKOOỌAAIOO. One important thing to note is that, the noun (OO) and adjective (ỌA) swapped with the determiner. It shows that Okun language is head first and English language is head last.

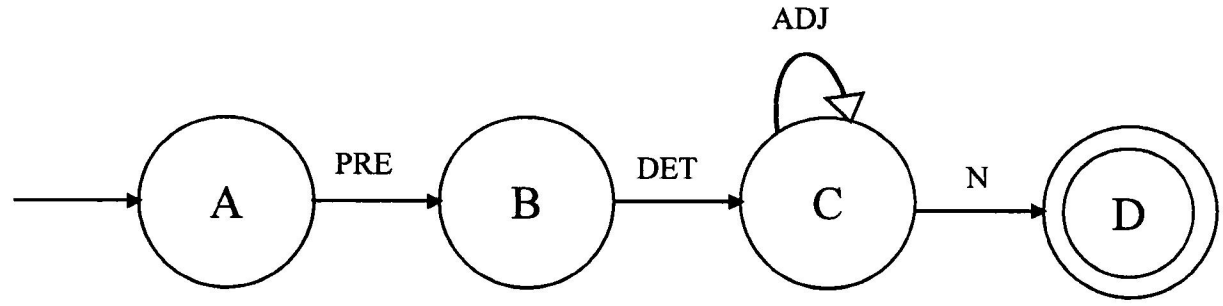


Figure 3.8a: State diagram for the English translation process

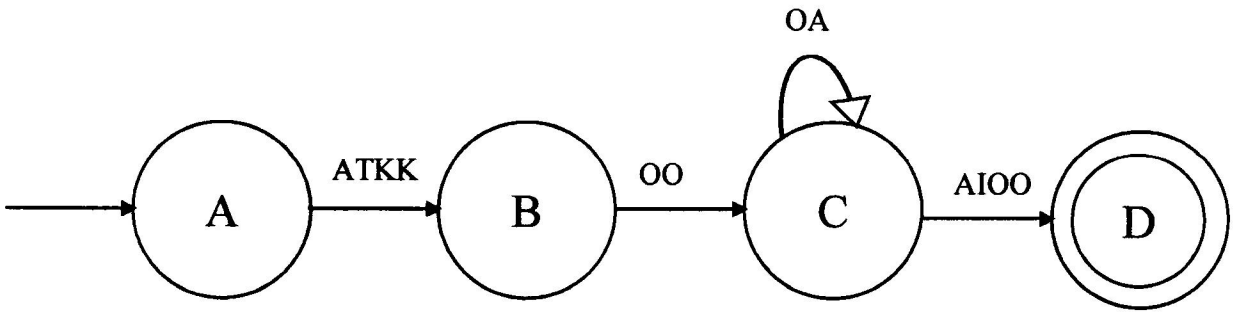


Figure 3.8b: State diagram for the Okun translation process

3.6 Database Design

The System database consists of about three hundred (300) nouns, fifty (50) prepositions and about twelve (12) determinants. Phrases from the two languages were broken down into their part of speech (POS). The different parts of speech were stored in pairs. Table 3.1 and 3.2 shows list of English pronouns and their Okun equivalents, and English Determinants and their Okun equivalents. Figure 3.9a, and Figure 3.9b, show the Noun and Prepositions respectively in the database.

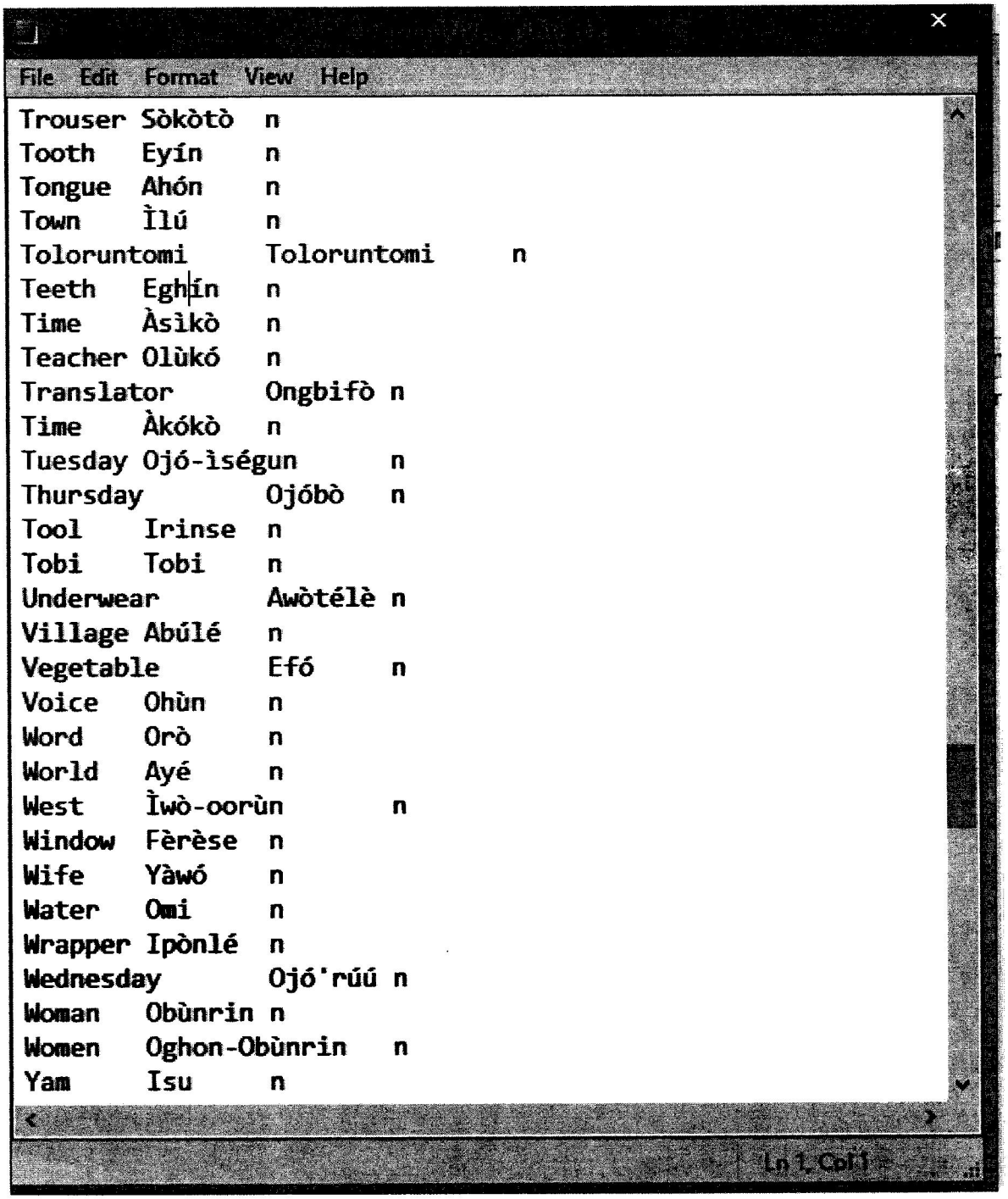


Figure 3.9a: Database showing nouns.

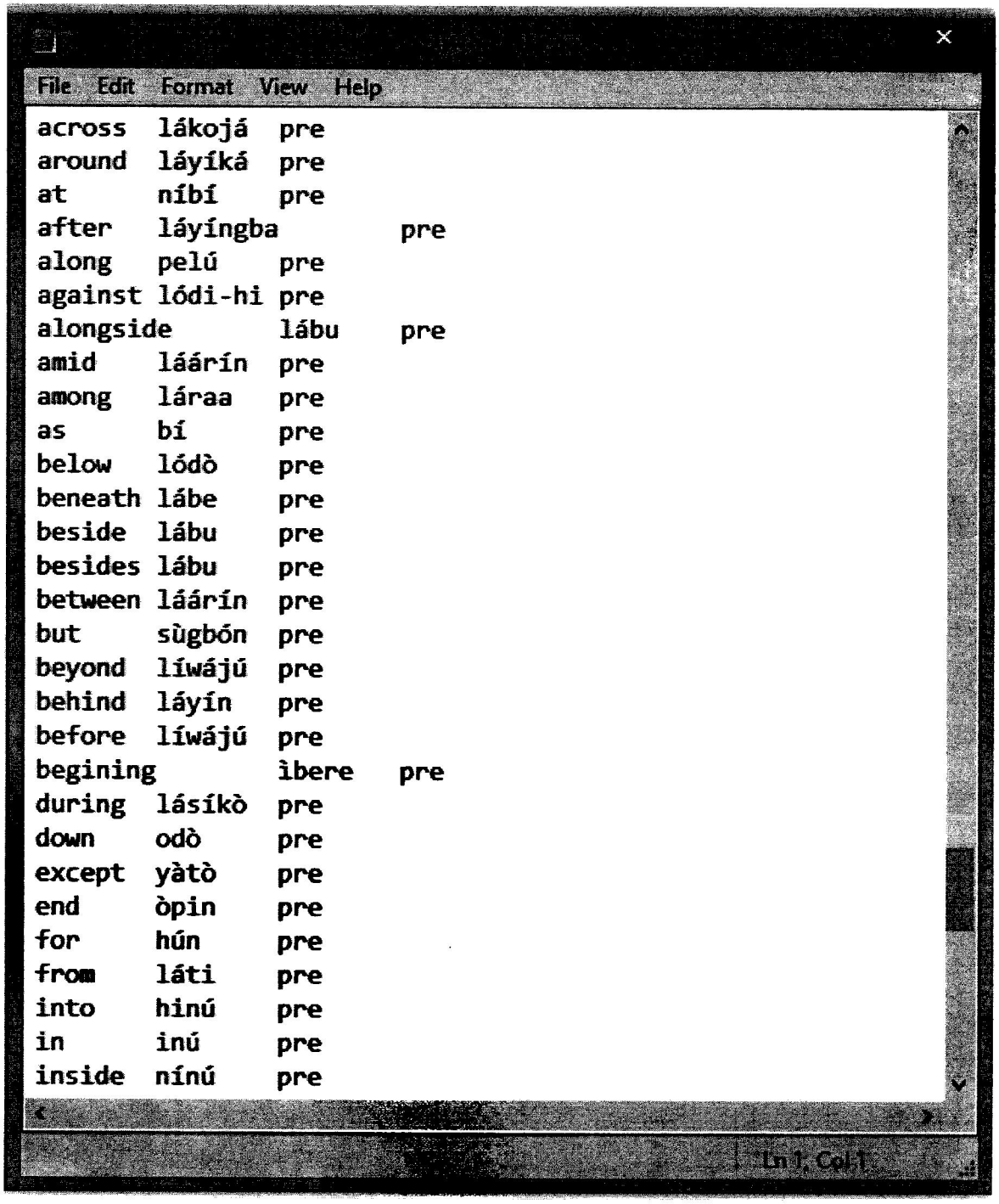


Figure 3.9b: Database showing Prepositions

Table 3.1: List of English pronouns and their Okun equivalents

English	**Okun**
She/he/it	Óun
They	Oghọn
You	Ìwọ
We	Àgha
Them	Ghọ́n
Those	Ghòìn
These	Eléìn
This	Èìn
That	Tédẹ́

Table 3.2: List of English determinants and their Okun equivalents

English	**Okun**
A	Kan
An	Kan
Some	Díé
The	Ghin

3.7 System Software Design and Implementation

The software design contains the Graphical User Interface (GUI) which was designed using PyQt5 and implemented using python programming language. The GUI features four phases, the first phase features two radio buttons which help to select either of the languages to be either the source or as target language, the second phase is a space where users can enter the Prepositional phrase. The third phase displays the translated prepositional phrase. The fourth phase displays whether the translation was successful or not.

On entry of the text in source language, the translator module of the code began to execute. The phrase was broken into lexemes, it was then tagged into different parts of speech. The translator module accepted input sentence from the GUI module then broke it down, and sent it to the database module to confirm that the lexemes are in the database. However, if the lexemes are not in the database an error message will be generated. The translated sentence to target language was then displayed by the GUI. Python programming language was used in the software coding and the interface of the machine was designed using PyQt5. The lexemes are manually tagged and each word is categorized according to its parts of speech. The Natural Language Tool Kits (NLTKs) was used as the parser module. The translation process is based on the phrase grammar rules built in the source code which implements the re-write rules. The machine translation system has the capability to translate sentences that contains a Prepositional phrase from the English Language to Okun language in its textual form and vice versa.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Evaluation of System

The developed system was evaluated using Mean Opinion Score approach. Questionnaires were designed and distributed amongst an Expert and subject respondents. Results from the experimental subject respondents were recorded.

4.2.1The Mean Opinion Score

A Mean Opinion Score (MOS) is a numerical measure of the human-judged overall quality of an event or experience (Paul, 2018). The Mean opinion score (MOS) is a subjective measurement of people's opinion. The Expert who is a professional translator translates the sentences from English language to Okun language and this serves as the baseline for this research.

4.2.2 Questionnaire Design

The administered questionnaire contains simple English prepositional phrases, designed to test the experimental subject respondent on the ability to translate simple English language sentences to Okun language. The questionnaire has twenty (20) simple phrases. The phrases in the questionnaire are meant to test the respondent translations' accuracy considering Okun language orthography and the syntax of the language which is described in term of tone marks and diacritics (dotted vowels and consonant).

4.2.3 Questionnaire Administration

The questionnaires were administered in Kabba Town, Kogi State, Nigeria (the capital of the Okun people). This area was chosen because there are literate Okun speakers widespread from all the villages and towns collectively referred to as "Okun land". The questionnaires were distributed among the Okun speakers in the Community.

4.3 Result

The result of the system is shown below, English language prepositional phrases were successfully translated to Okun language, and also prepositional phrases in Okun language were successfully translated to English language. A pictorial view of the System Output are shown in figure4.1 and 4.2.

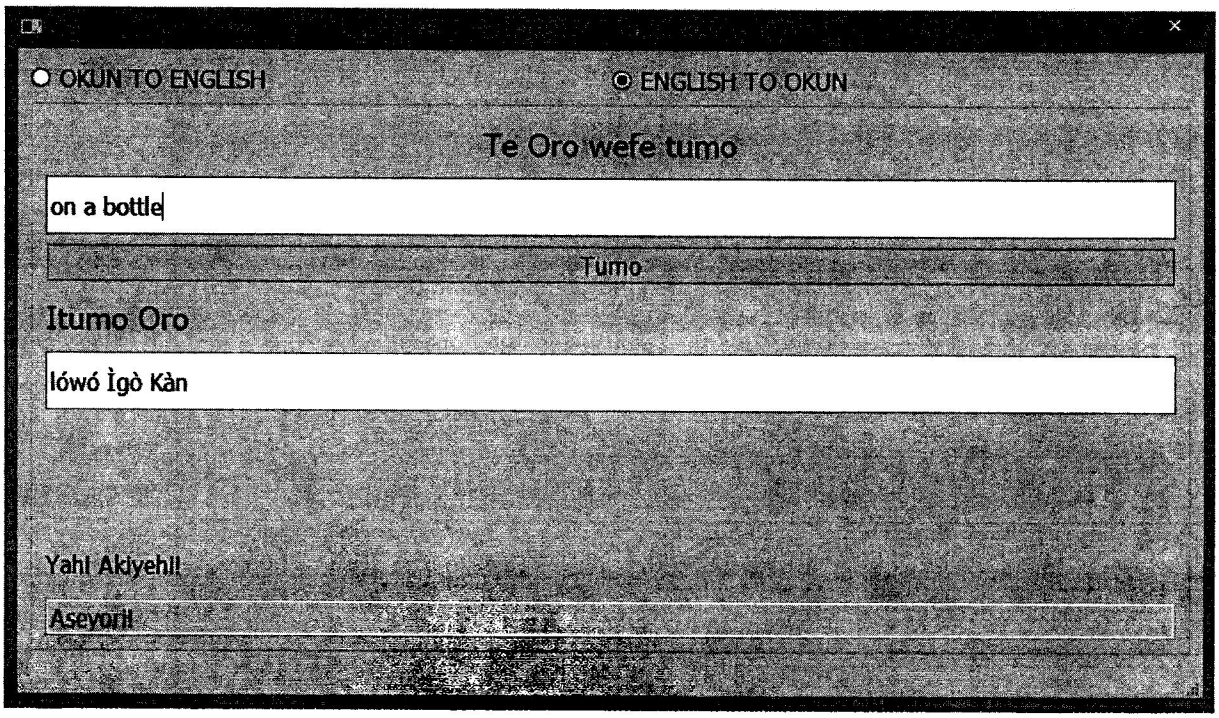


Figure 4.1. System Output for English language to Okun language translation

Figure 4.1 above shows the output for English to Okun translation, shown is the English prepositional phrase (on a bottle) inputted as the source text with the Okun language equivalent of the text displayed as the target output.

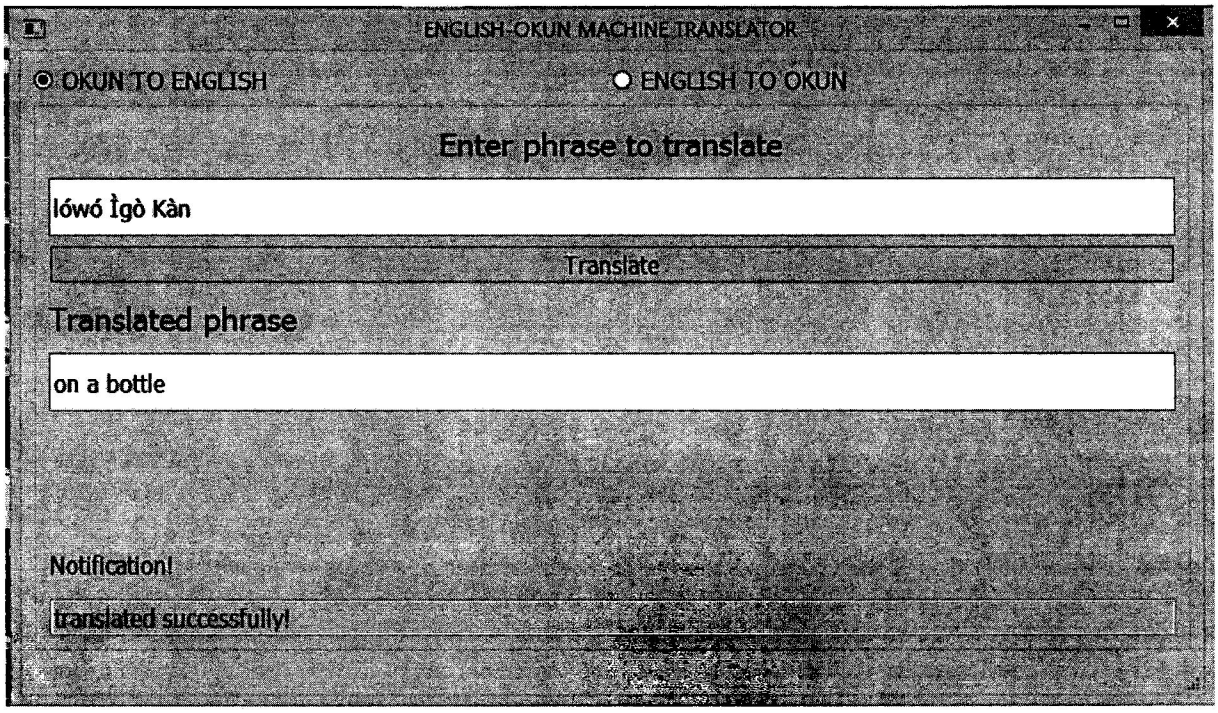


Figure 4.2. System Output for Okun language to English language translation

Figure 4.2 above shows the output for Okun to English translation, shown is the Okun prepositional phrase (lówó ìgò kàn) inputted as the source text with the English language equivalent of the text displayed as the target output.

4.4 Discussion of Result

The system was evaluated to determine the performance of the system designed. The accuracy of the developed system was verified using word orthography (tone marking and spellings). Most of the experimental respondents got the translation, thus the need for the metric above as many do not know how to tone mark words. The evaluation was done by comparing the Expert translated phrases with the one translated by the machine and the experimental respondents using the mean opinion score (MOS) technique. The result of the evaluation is shown in Table 4.1 and figure 4.3. It was observed from the bar chart that the machine scores are higher than the average score of the experimental subject respondents and the expert has the highest.

Table 4.1: Analysis of Evaluation Results

Phrases	Expert	Respondent Average	Machine
1	100	59	90
2	100	63	100
3	100	64	100
4	100	54	90
5	100	59	90
6	100	60	90
7	100	25	90
8	100	59	85
9	100	43	100
10	100	66	100
11	100	49	90
12	100	71	90
13	100	68	90
14	100	68	95
15	100	67	90
16	100	58	90
17	100	53	90
18	100	60	100
19	100	51	90
20	100	60	100
Average	**100**	**57.85**	**93**

The expert average was 100 percent and that of the experimental subject respondents was 57.85 percent while that of the developed machine translator was 93 percent. The Chart below depicts that the machine correctness is close to that of the Expert and more accurate than that of the average experimental subject respondents.

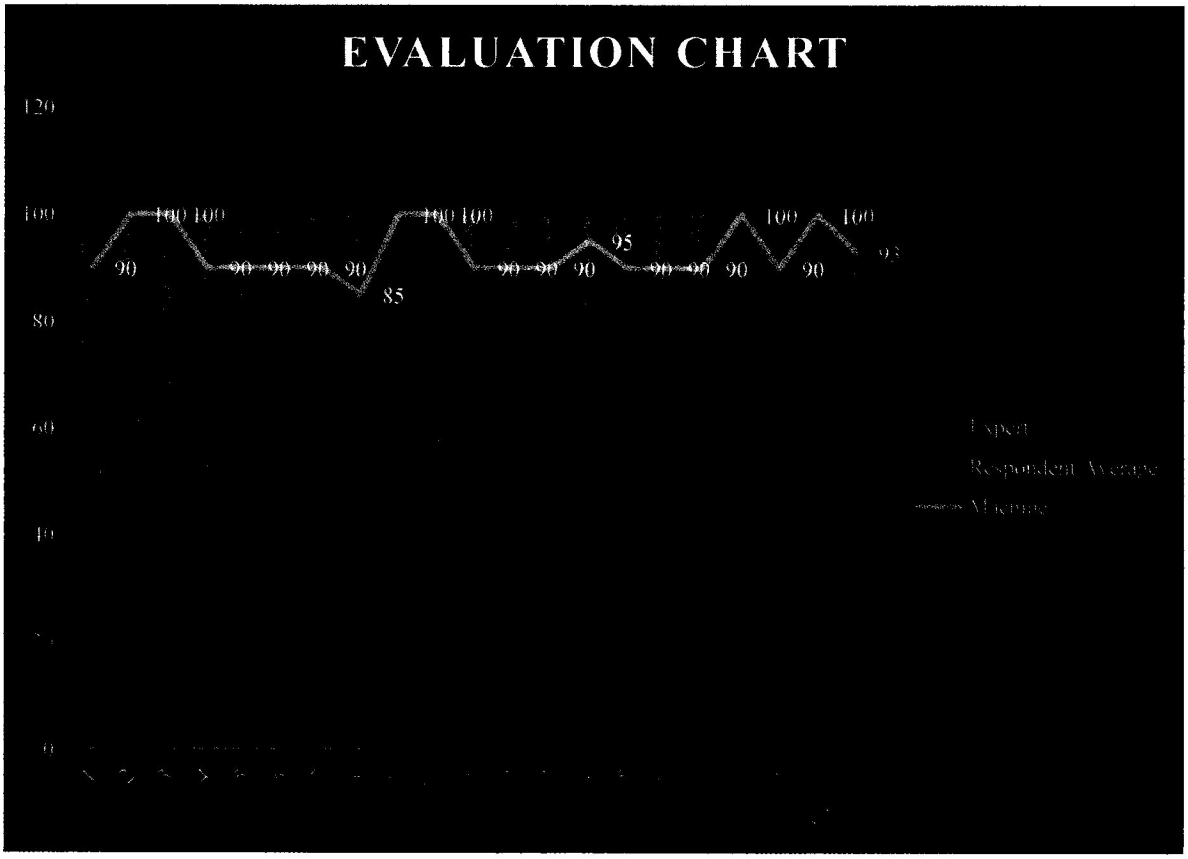


Figure 4.3: Translated phrases orthography accuracy

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 **Conclusion**

The need for Machine translation system cannot be overemphasized in this era of rapid globalization, especially for indigenous languages in Nigeria. An English-Okun prepositional phrase Machine translation system based on rule based approach was developed in this research. The system was evaluated using human judgment based on orthography accuracy. Results show that the system is ninety three percent accurate when compared to that of the experimental subject respondents which is just about fifty eight percent. The results show that most indigenous people are not good at writing Okun language and also find it hard to separate Okun dialects from the Yoruba language. This reveals the extent to which Okun Language or dialect is going into extinction. Hence, the system was designed to enhance the learning of the Okun language which is a dialect under the Yorùbá language. It is user-friendly and allows learners to learn the language at ease.

5.2 **Recommendations**

The accuracy of the developed system can be enhanced by improving and extending the full form bilingual lexicon. This work only handles the translation of prepositional phrases which is part of a complete sentence. The system can be developed further to produce acceptable translation of complete sentences and hosted on the internet for public use. The system would be of immense benefit to the Okun people as it will help to further develop the language and elevate it to the level of languages of developed nations.

Also, the result obtained from this research work reveals that most speakers of Okun language have low proficiency in writing it and some find it hard to differentiate it from Yoruba language. In view of this, it is recommended that schools in the western part of Kogi state incorporate this software into their academic system to enable students learn more about the language, tones, hyphenation, capitalization, word breaks, punctuations, sounds and spellings.

Finally, it is recommended that schools and the government encourage researchers in the field of computational linguistics by funding researches.

REFERENCES

Abiola, O.B, AdetunmbiA.O and Oguntimilehin A. (2015). Using hybrid approach for English-to-Yoruba text to text machine translation system (proposed)". *International Journal of Computer Science and Mobile Computing, 4(8)*, 308-313.

Adeyanju I., Wiratunga N., Lothian R. & Craw S. (2010). Applying Machine Translation Evaluation Techniques to Textual CBR, Proceedings of Eighteenth International Conference on Case Based Reasoning (ICCBR'10), LNCS 6176, 21-35, Springer press, Berlin.

Agbeyangi, A.O., Eludiora, S.I. &Adenekan, O.A. (2015). English to Yorùbá Machine Translation System using Rule-Based Approach. *Journal of Multidisciplinary Engineering Science and Technology, 2(8)*.

Arnold, D., Balkan, L., Meijer, S., Humphreys, R. L., & Sadlery, L. (1994). *Machine Translation: An Introductory Guide*. London: NCC Blackwell.

Ayegba, S. F., & Osuagwu, O. E. (2015). Improving Adequacy in a Rule-Based English-to-Igala Automatic Translation System through Word Sense Disambiguation . *British Journal of Mathematics & Computer Science* .

Ayegba, S. F., Ugbedeojo, M., & Philip, N. (2016). Design and Implementation of a System for Automatic Translation of IGALA to English Language. *International Journal of Scientific Research in Science and Technology*.

Banjo, A. E., & Jibowo, A. V. (2011, December). The use of principles and Theories of Translation in languages: A case study of Yoruba. *Journal of Communication and Culture; International Perspective Vol. 2 No 3.*

Bayo Ijagbemi (1996)"O-Okun Yoruba in Yoruba Art Historiography: *History, Problems And Prospects".PhD Thesis*. The University of Arizona

Billot, S. & Lang, B. (1989). The structure of shared forests in ambiguous parsing: In Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics, pp 143- 151

Billy , W., & Chunyu, K. (2009). *Automatic evaluation of machine translation via word choice and word order.* Massachusetts: Kluwer Academic Publishers Hingham, MA, USA .

Brown, P.E., Della, V. J., Della, S. A., Mercer, R. L., (1993). The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics, 19(2),* pp. 263-311

Cmejrek, M. (2014). Reordering Model for Forest-to-String Machine Translation. In proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 227–232

Doddington, G. (2002). *Automatic evaluation of machine translation quality using n-gram co-occurrence statistics.* San Diego, California, United States of America: Morgan Kaufmann Publishers Inc.

Eludiora, S.I. (2014). Development of an English to Yoruba Machine Translation System. Unpublished PhD thesis, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria.

Eludiora, S. I., Agbeyangi, A. O., & Fatusin, O. I. (2015). Development of English to Yorùbá Machine Translation System for Yorùbá Verbs' Tone Changing. *International Journal of Computer Applications.*

Eludiora, S., & Elufidodo, G. (2016). Development of a Verb Group Machine Translation System. *International Journal of Computer Applications.*

Eludiora, S., & Atolagbe , R. (2016). Development of a Prepositional Phrase Machine Translation System. *World Journal of Computer Application and Technology*, 46-57.

Esan, A., Omodunbi, B., Olaniyan, O., & Olaleye, T. (2018). Development of Adjectival Phrase-Based English to Yorùbá Machine. *International Digital Organization for Scientific Research*, 29-42.

Folajinmi, Y. & Omonayin, I. (2012). Using Statistical Machine Translation As A language Translation tool for understanding Yoruba. *EIE's* 2nd International Conference on Computing Energy, Networking, Robotics and Telecommunications, pp 86-91

Galley, M., Graehl, J., Knight, K., Marcu, D., DeNeefe, S., WeiWang, & Thayer, I. (2006). Scalable Inference And Training Of Context-Rich Syntactic Translation Models. Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, pp. 961–996

Germann, U. (2012). Syntax-aware Phrase-based Statistical Machine Translation: System Description. In proceedings of the 7th Workshop on Statistical Machine Translation, pp. 292–297

Graehl, J. & Knight, K. (2004). Training tree transducers. *In Proceedings of HLT-NAACL* (pp. 105–112). Boston: MA

Huang, L., Knight, K. & Joshi, A. (2006). Statistical syntax-directed translation with extended domain of locality. In Proceedings of the 7th Conference of the Association for Machine Translation in the Americas, pp. 223-231

Hutchins, J. W. (1994). MACHINE TRANSLATION: A BRIEF HISTORY. London: Academic Press.

Hutchins, W. J. (1995). Concise history of the language sciences: from the Sumerians to the cognitivists. In W. J. Hutchins, *Concise history of the language sciences: from the Sumerians to the cognitivists (pp. 431-445).* Oxford: Pergamon Press.

Ijagbemi, B. (1996). O-OKUN YORUBA IN YORUBA ART HISTORIOGRAPHY: HISTORY, PROBLEMS AND PROSPECTS. Arizona: The Ubiversity of Arizona.

Kumar, P. & Kumar, V. (2013). Statistical Machine Based Punjabi to English Transliteration System for Proper Nouns. *International Journal of Application or Innovation in Engineering and Management (IJAIEM). 2(8)*, pp. 318-321.

Li Z., Callison-Burch, C., Dyer, C., Ganitkevitch, J., Khudanpur, S., Schwartz, L., Thornton, W. N., Weese, J. &Zaidan, O.F., (2009). Joshua: An Open Source Toolkit for Parsing-based Machine Translation. Proceedings of the 4th EACL Workshop on Statistical Machine Translation, pp. 135–139

Liu Y., Yajuan, L. & Liu, Q. (2009). Improving Tree-to-Tree Translation with Packed Forests. In Proceedings of Association of Computational Linguistics, pp. 84-91

Mehay, D.N. & Brew, C. (2012). CCG Syntactic Reordering Models for Phrase-based Machine Translation. Proceedings of the 7th Workshop on Statistical Machine Translation, pp. 210–221

Melamed, I. D., (2004): "Statistical machine translation by parsing". In Proceedings of the Association for Computational Linguistics (ACL), pp. 654–661

Mi, H., Liang, H. & Liu, Q. (2008): "Forest-based translation". In Proceedings of Association of Computational Linguistic, pp. 192–199

Nießen, S., Och, F.J., Leusch, G. & Ney H. (2000). An evaluation tool for machine translation: Fast evaluation for MT research. In Proceedings of 2nd International Conference on Language Resources and Evaluation, pp. 39–46

Och, F.J. (2003). Minimum error rate training in statistical machine translation. Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, pp. 160–167

Oladosu, J.B. & Emuoyibofareh, J.O. (2012). A Yoruba-English Language translator for doctor-patient mobile chat application. *International Journal of Computers and Applications, 34(3),* pp. 149- 156.

Oladosu, J., Esan, A., Ibrahim, A., Benjamin, A., Olatayo, O., & Bolaji, O. (2016). Approaches to Machine Translation: A Review. *FUOYE Journal of Engineering and Technology*, 120-126.

Papineni, K., Roukos, S., Ward, T., Zhu, W., (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (pp. 311-318). Stroudsburg: PA, USA

Paul, K. (2018, February 23). *Mean Opinion Score (MOS).* Retrieved from Twilio: https://www.twilio.com/docs/glossary/what-is-mean-opinion-score-mos

Quirk, C. & Corston-Oliver, S. (2006). The impact of parse quality on syntactically-informed statistical machine translation. In proceedings of Conference on Empirical Methods in Natural Language Processing, pp. 56-63

Robert , H. R., & David , C. (2018, June). Language. Retrieved from ENCYCLOPÆDIA BRITANNICA: https://www.britannica.com/topic/language#accordion-article-history.

Sadiat, F. (2013). Towards a Hybrid Rule-Based and Statistical Arabic- French Machine System. In proceedings of Recent Advances in natural Language Processing, pp. 579-583.

Sangeetha, J., Jothilakshmi, S., & Kumar, R. (2014). An Efficient Machine Translation System for English to Indian Languages Using Hybrid Mechanism. *International Journal of Engineering and Technology (IJET).*

Sharma, N. (2011). English to Indi Statistical Machine Translation System, *Unpublished Masters Thesis, Tharpa University, India*

Tripathi, S., & Sarkhel, K. J. (2010, December). Approaches to machine translation. *Annals of Library and Information Studies Vol. 57*, pp. 388-393.

Zhang, H., Zhang, M., Aiti, H.L., Chew, A. & Tan, L. (2009). Forest- based Tree Sequence to String Translation Model. In proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, pp,172-180,Suntec, Singapore.

QUESTIONNAIRE

FEDERAL UNIVERSITY OYE-EKITI, EKITI STATE NIGERIA, DEPARTMENT OF COMPUTER ENGINEERING.

PROJECT TOPIC: DESIGN OF AN ENGLISH - OKUN PREPOSITIONAL PHRASE MACHINE TRANSLATION SYSTEM USING RULE BASED APPROACH

Dear Sir/Madam,

I am a final year undergraduate from the Department of Computer Engineering at the Federal University Oye-Ekiti, Ekiti State. I am undertaking a research for my dissertation which aims to design a rule based system to translate prepositional phrase in English to Okun. Your expertise in the educational and linguistic field is extremely valuable in assisting me to learn more about both languages and the project proper. Please be assured that your responses will be kept confidential and will only be used for academic purposes.

I sincerely thank you for your valuable time!

PART ONE: BACKGROUND INFORMATION

Please provide the information as required below by ticking (√) in the appropriate box.

AGE: 18 and below ( ) 19-24 ( ) 25-40 ( ) 41-60 ( ) 60 and above ( )

GENDER: Male ( ) Female ( )

OCCUPATION: Self-Employed ( ) Civil Servant ( ) Others ( )

HIGHEST ACADEMIC QUALIFICATION:

NCE ( ) ND/HND ( ) BSC ( ) M.Sc ( ) Ph.D ( )

INSTITUTION:.....

PART TWO: START OF QUESTIONNAIRE

Please give the appropriate translation to the following prepositional phrases in the spaces provided.

	ENGLISH PHRASES	OKUN TRANSLATION
1.	Across the Road	
2.	Inside a Bottle	
3	On the Bridge	
4	Besides the Table	
5	Into the Market	
6	Among the boys	
7	As the student	
8	Amidst the girls	
9	Except a Banana	
10	Between two women	
11	Over the Car	
12	In the blood	
13	Beneath the table	
14	Before the Statue	
15	Across the bridge	
16	Beyond those boys	
17	Behind the mirror	
18	Since that afternoon	
19	Opposite the mosque	
20	Outside the house	