https://dx.doi.org/10.4314/dujopas.v10i1c.6
ISSN (Print): 2476-8316
ISSN (Online): 2635-3490
Dutse Journal of Pure and Applied Sciences (DUJOPAS), *Vol. 10 No. 1c March 2024*

A Rule-based Approach to English-Okun Prepositional Phrase Machine Translation

[1]A. Esan, [2]A. Sobowale, [3]T. Adebiyi, [4]M. Adio and [5]S. Toloruntomi

[1,2,3,5]Department of Computer Engineering,
Federal University Oye-Ekiti,
Ekiti state, Nigeria

[4]Ajayi Crowther University,
Oyo, Nigeria

Email: adebimpe.esan@fuoye.edu.ng

Abstract
Okun culture is gradually going into extinction because the language is greatly dominated by English language. Therefore, this research developed a rule-based English-Okun machine translator using prepositional phrase. The dataset used were obtained from locally spoken words and the correctness of the re-write rules was tested using Java Formal Language Automata Package. Human judgment was employed in evaluating the system based on orthography accuracy. Results show that the developed translator is 95 percent accurate while experimental subject respondents have an accuracy of 58 percent. The developed machine translator's performance is closer to the Experts' but higher than that of the experimental subject respondents, this shows that indigenous people are not good at writing Okun language and find it hard to separate the dialect from Yoruba language. Hence, research in English-Okun machine translation should be intensified in order to prevent the language from going into extinction and future research should consider extending the full form bilingual lexicon of Okun language to achieve improved accuracy.

Keywords: Human Judgment, Okun Culture, Respondents, Prepositional Phrase, Translator

Introduction

Communication is an integral part of the culture and lifestyle of man and it is the major drive of progress and development globally (Esan et al. 2021). The absence of communication can lead to economic failure and the inability of a party to understand the language of the other can lead to miscommunication. Translation is important for communication to take place and existing methods of translation can be categorized into: human translation and technology aided translation (Ehab et al 2018). Translation of a source text from one language to another can be made possible through human translators and the individuals involved in translation are either professionals or non-professionals, the professionals are known as linguists. The use of human to carry out translation task can result in low speed of translation if the available text to be translated is of high volume, hence, technology was introduced to minimize this shortcoming. One of the ways by which technology can aid translation is the use of computer-assisted translation which are computer software packages employed in carrying out translation by the creation of translation memories (TMs) which contain the translator's choices (Espa˜na-Bonet and Ranta 2012). Machine translation is developed to translate one

**Author for Correspondence*

natural language into another by computers. It involves the automatic translation of words or sentences into a language other than the source (Ayogu et al 2018).

Many approaches such as: rule based (Esan et al 2018; Eludiora 2014), statistical based (Folajinmi and Omonayin 2015), phrase-based (Ayogu et al 2018), (Badr et., 2009) syntax based (Esan et al. 2021) and neural machine translation (Esan et al 2020) have been employed for machine translation. Rule Based Machine Translation (RBMT) deals with the morphological, syntactic and semantic information about the source and target language (Abiola et al, 2015). The objective is to convert source language structures to target language and build linguistic rules. Two types of RBMT are: transfer and direct based approaches (Oladosu et al 2016). Some of the advantages of the approach is the ability to deal with long distance dependencies and produce better translations syntactically. However, Okun culture is gradually going into extinction because English language has totally dominated Okun language in almost all human endeavor. Hence, this research developed an English-Okun machine translator using the rule-based approach.

Previous works have developed machine translation systems for foreign languages such as: German to French, Kazakh-Turkish, Catalan-Spanish, Croatian- Serbian and English-Indian among others. Torregrosa et al (2019) used a rule-based machine translation system for under-resourced neural machine translation. The research show that adding morphological information to the source language is as effective as using subword units. Charoenpornsawat et al., () proposed machine learning techniques, which help disambiguate word meaning. These methods focused on considering the relationship between a word and its surroundings, to automatically extract the context information, machine learning algorithms was applied which are C4.5, C4.5rule and RIPPER. The result shows that the accuracy of these three machine learning techniques is not quite different, and RIPPER gives the better results than C4.5 in a small train set. Costa-Jussa et al (2012) analysed the main differences between rule-based and statistical machine translation paradigms in the specific case of Catalan-Spanish pair. Results show that a correlation was found between the type of linguistic errors committed and the core methodology of the systems. Orthographic and morphological errors tend to be lower in the rule-based machine translation systems, while the performance at the semantic level is better in the statistical systems. Sevilay (2018) presented a shallow-transfer machine translation (MT) system for translating from Kazakh to Turkish. Linguistic components were developed, including a Kazakh-Turkish bilingual dictionary, Constraint Grammar disambiguation rules, lexical selection rules, and structural transfer rules. Results revealed that the RBMT system has reached performance comparable to publicly-available corpus-based MT systems between the languages. Sreelekha (2016) presented a case study between Statistical Machine Translation (SMT) and Rule-Based Machine Translation (RBMT) systems on English-Indian language and Indian to Indian language perspective. Results show that with a small amount of training corpus a statistical machine translation system has many advantages for high quality domain specific machine translation over rule-based systems. Klubiˇcka et al., (2016) developed a bi-directional rule-based Croatian- Serbian machine translation system based on the open-source Apertium platform, the project was developed with the aim of creating free linguistic resources. Sadiat (2013) built an Arabic-French phrase–based machine translation system using the rule based approach.

Several methods have been employed by previous works to develop machine translation systems for English to Nigerian languages such as: English to Yoruba, English-Igbo, English-Igala among others. Esan et al (2018), Eludiora (2014), Eludiora et al (2015), Eludiora and Elufidodo (2016) developed English to Yoruba Machine Translation system using the rule-based approach. Ayegba and Osuagwu (2015) developed a rule-based machine translation system that translates English sentences to Igala language while Esan et al (2021) developed a syntax based English-Igbo statistical machine translator using the rule based and statistical machine translation approaches. Results from automatic evaluation revealed that the system performed better than google translate. Agbeyangi et al., (2015) developed a transfer Rule-Based Machine Translation system and results show that the system outperforms Google translator which was used as the baseline. Abiola (2015) proposed a hybrid approach to English -Yoruba machine translation, a word-sense model was incorporated in the system which disambiguate sentences before passing it to the translator to further improve the output. Ayegba et al., (2016) develop a rule based language processor that can accept input text in Igala language and automatically translate same to English language. Esan et al (2020) developed an English-Yoruba machine translation system using Recurrent Neural Networks, results show that neural machine translation outperforms the state of the art MT systems. This research developed an English-Okun Machine Translation system using the rule based approach.

Methodology

This project focuses on English-Okun prepositional phrase machine translation using the Rule-Based Approach. The architecture of the system as shown in Figure 1 includes four main components: a parser, an analyzer, a transformer and a generation component. Source text are either English or Okun, the parser produces a syntactic structure for the given input while the preprocessor counts the number of words in the given prepositional phrase and declare three arrays of the size of the number of words for use by the other modules. The tokenizer performs text processing operation which involves breaking up of raw text into words while 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. The bilingual dictionary is the database where all words are housed with their respective part of speech tags and the transformer consists of a set of transformation rules which was used to build the target language equivalent of the input source sentence. The syntactic/semantic structure of the source language is transferred into the syntactic/semantic structure of the target language. The system graphical user interface (GUI) is an interactive interface that connects the user and the translator and the database stores the dataset 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.

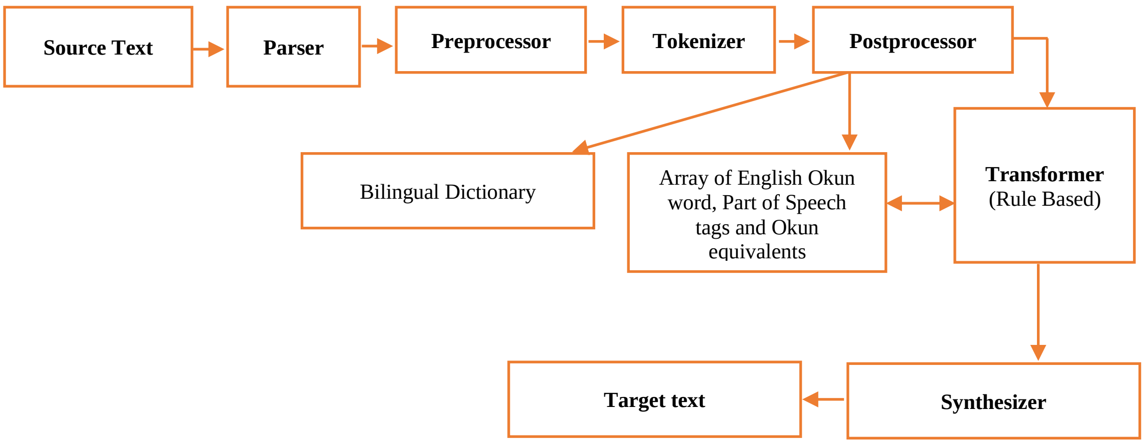


Figure 1: Architecture of the Rule-Based English-Okun Translator

System Design

The English-Okun machine translator was designed using Prepositional Phrase and Input texts were divided into tokens (lexemes) and the tokens were patterned according to the re-write rules. The output of the system (the translated text) was displayed through the Graphical User Interface (GUI). The design process is shown in Figure 2.

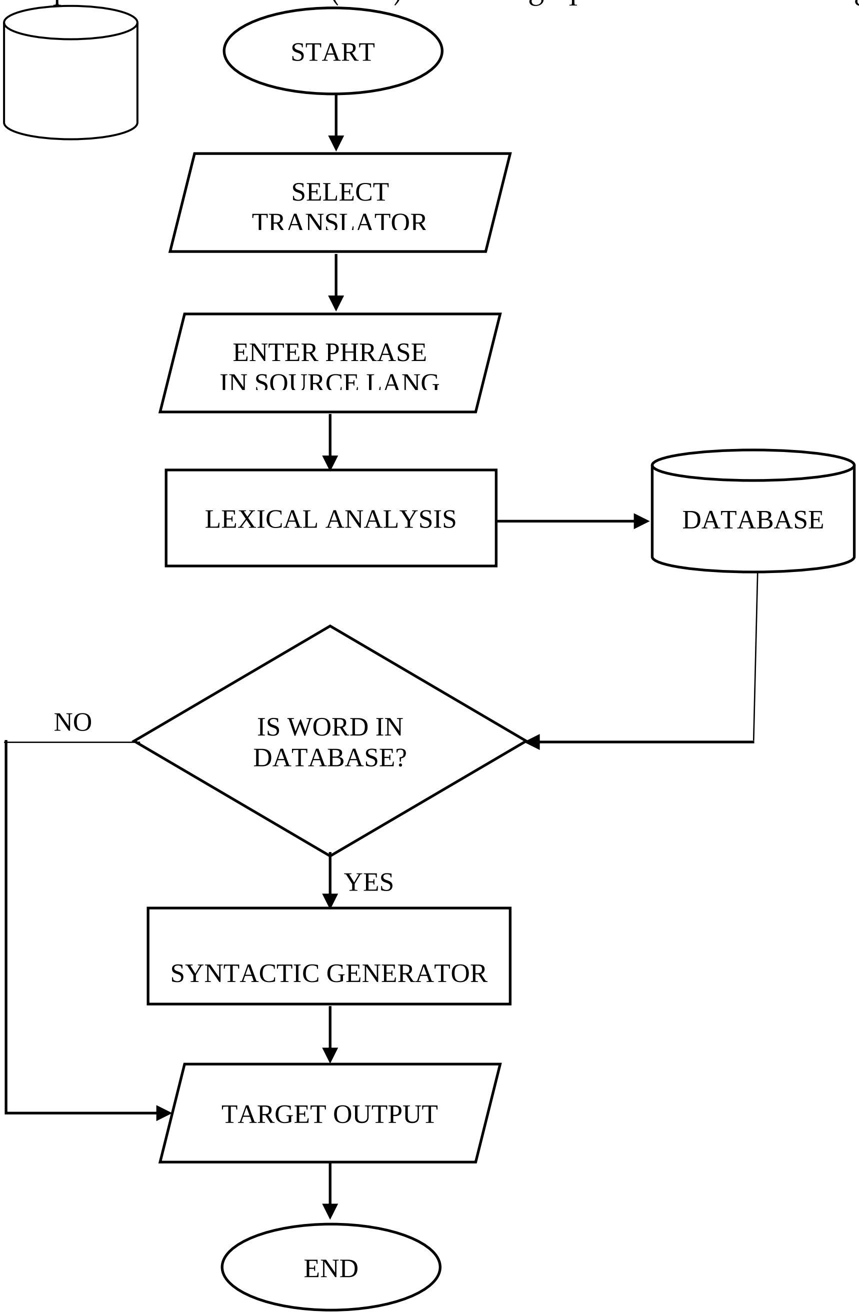


Figure 2: Flowchart for the system.

English and Okun Phrase Grammar and Re-Write Rules

The re-write rules according to the Okun sentence structure 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"). The list of acronyms for English and Okun re-write rules are shown in table 1.

Table 1. Lists of Acronyms

English	Acronym
NP	Àpólà ọ̀rọ̀ Erúkọ̀ (APỌE)
PP	Àpólàọ̀rọ̀ Atọ́kùn (APTK)
VP	Àpólà ọ̀rọ̀ ìṣe (APỌI)
ADJP	ÀpólàÒròÀpònlè (APỌA)
PRE	ọ̀rọ̀ Atòkùn (ATK)
N	ÒròErúkọ (ỌE)
PRN	ArópòÒròErúkọ (AỌE)
ADJ	ÒròÀpònlé (ỌA)
DET	Asapejuwello oro eruko (AIOE)

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

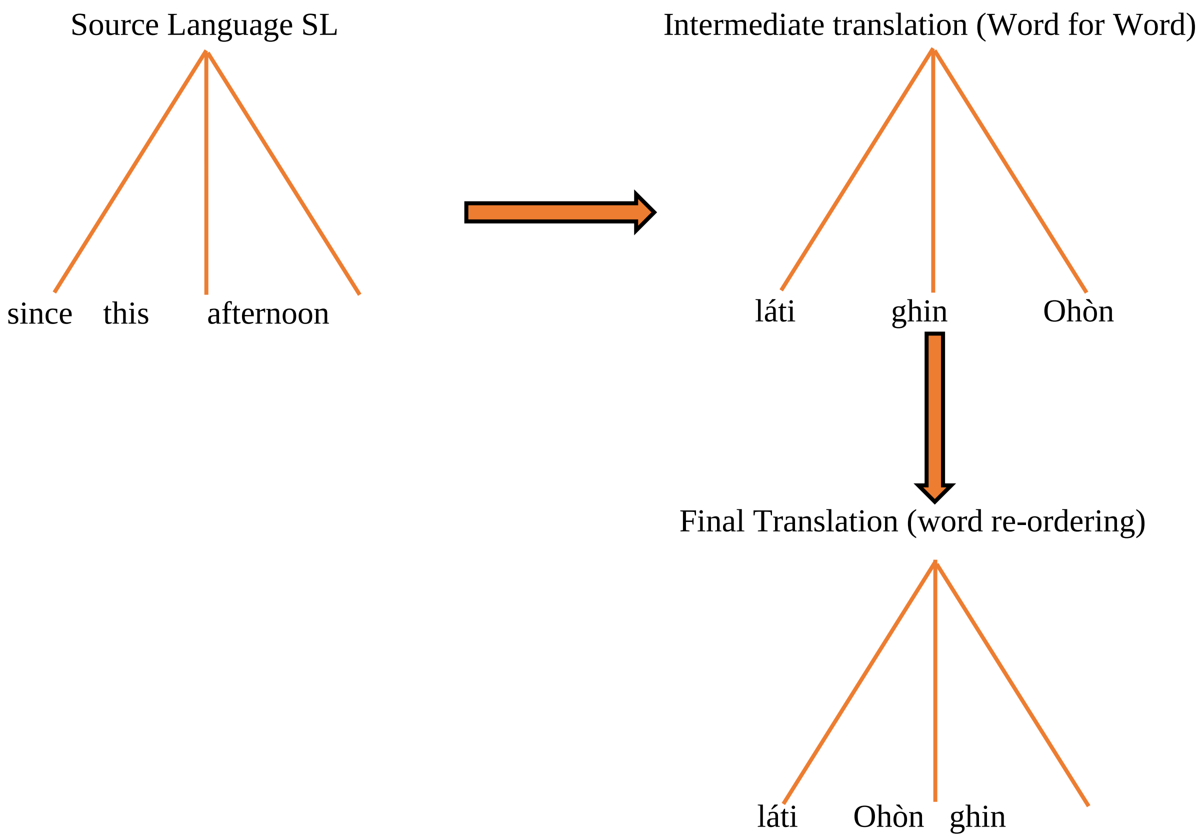
Re-write Rules

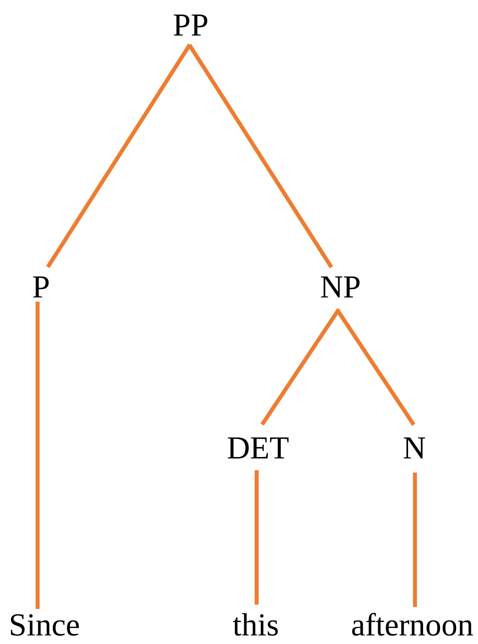
The following rules guide the design of the machine translator:
Rule 1: A prepositional phrase (PP) consists of a preposition and noun phrase (NP). In the case of Okun language, noun (ỌE) comes before determiner (AIỌE). That is;
SL: since<PRE>this<DET>afternoon<N>.
TL: láti<ATK>ohon<ỌE>ghin<AIỌE>
Rule 2: A determiner must precede an adjective and a noun in English Language, but not in Okun language. That is;
SL: This<DET>wicked<ADJ> boy<N>.
TL: ọmọkùnrin<ỌE>biburu<ỌA>ghin<AIỌE>.

Prepositional Phrase Translation Process Abstraction

The mode of translation was based on the grammar designed for both English language and Okun language. The parse trees which shows the pictorial view of the prepositional phrase is shown in Figure 3. JFLAP was used to test the rewrite rules as shown in figures 3 and 4.

Figure 3: Prepositional Phrase Translation Process Abstraction





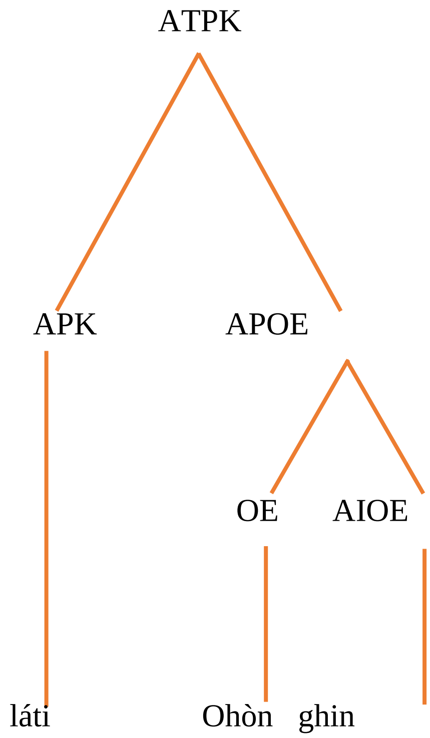


Figure 4a: Parse tree for an English prepositional phrase Figure 4b. Parse tree for an Okun prepositional phrase

Re-write Rules Testing

The correctness of the re-write rules for English and Okun were determined using Java Formal Language Automata Package (JFLAP). Figures 5 and 6 show the outputs of the SL and TL prepositional phrases.

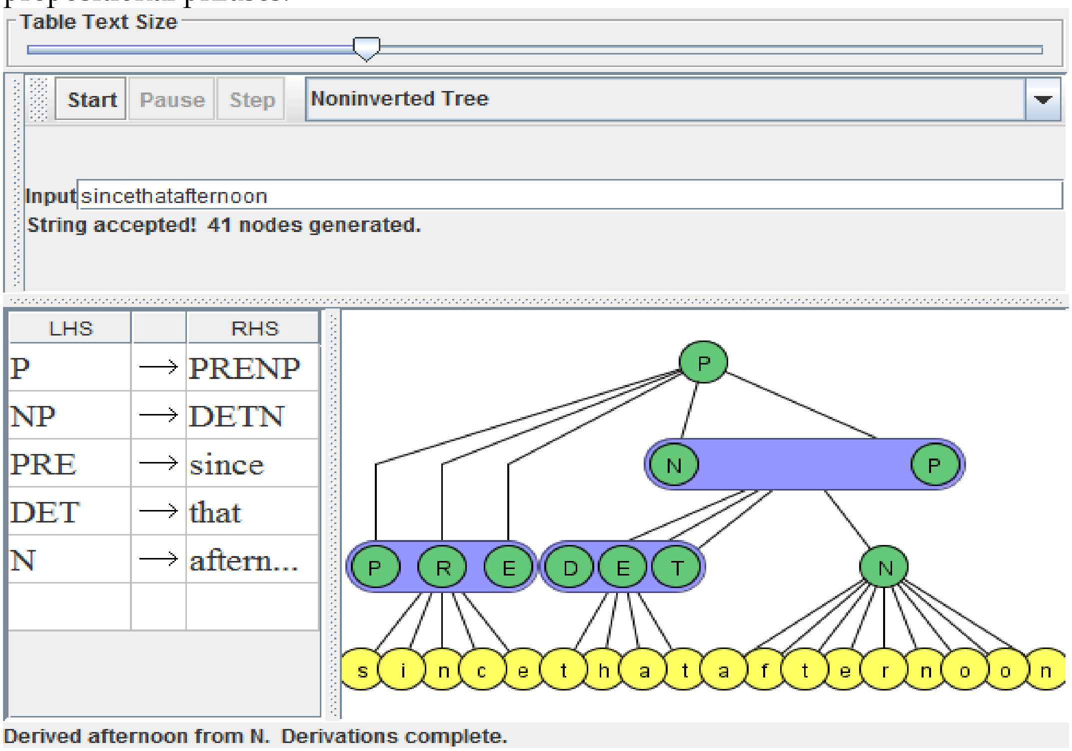


Figure 5: English Prepositional Phrase Rewrite Test

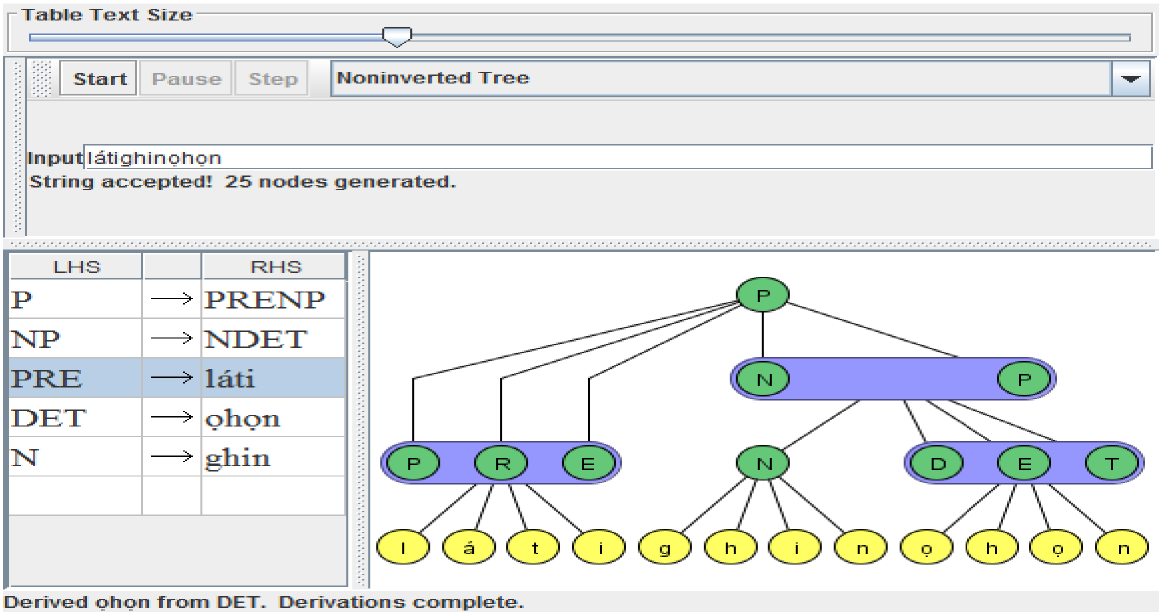


Figure 6: Okun Prepositional Phrase Rewrite Test

The Prepositional Phrase Translation Process Model

PREDETN and PREDETADJN are possible translation combinations for the English preposition phrase. They are: ATKỌEAIỌE and ATKỌEỌAAIỌE. The noun (ỌE) and adjective (ỌA) are swapped with the determiner and this shows that Okun language is head first and English language is head last. Figure 6 is the state diagram for the English PP translation process model and Figure 8 is the state diagram of the Okun language PP translation process.

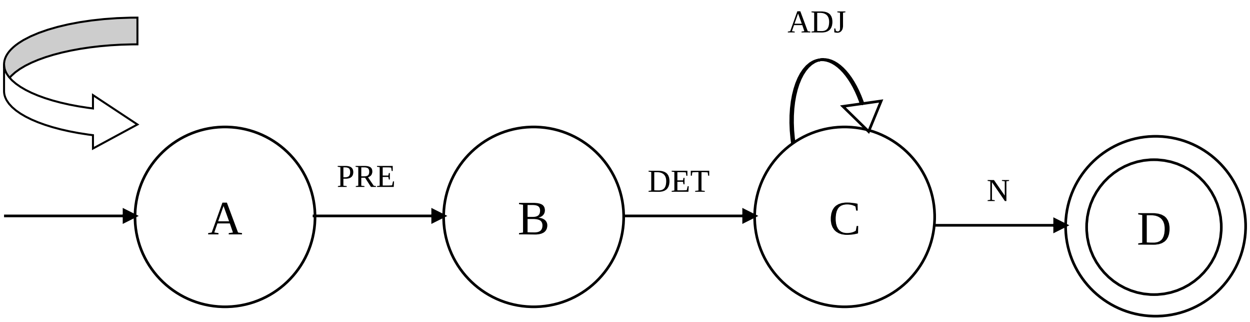


Figure 7: State diagram for the English translation process

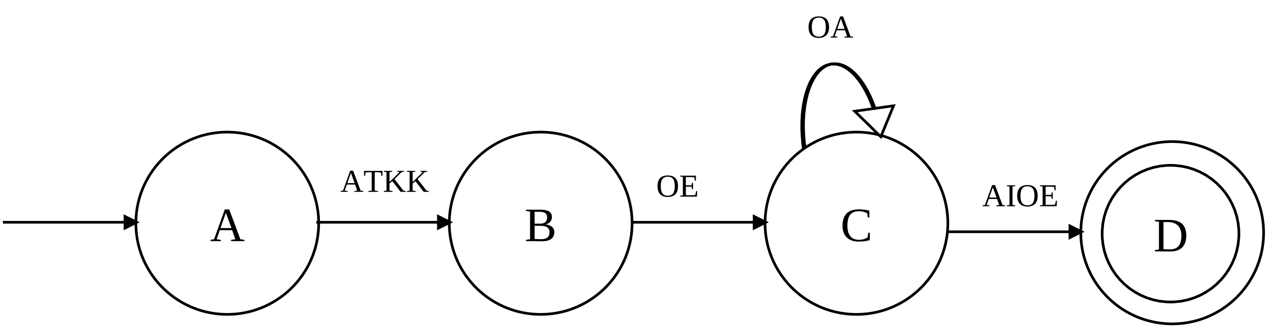


Figure 8: State diagram for the Okun translation process

Evaluation
The developed system was evaluated using Mean Opinion Score approach and questionnaires were designed and distributed among an Expert and subject respondent. The administered questionnaire contains simple English prepositional phrases, designed to test the experimental subject respondents on the ability to translate simple English language sentences to Okun language. The questionnaire contain (20) simple phrases. The phrases in the questionnaire are meant to test the respondents' translations' accuracy in Okun language orthography and the syntax. The questionnaires were administered in Kabba Town, Kogi State, Nigeria (the capital of the Okun people) and distributed among the Okun speakers in the Community. This area was chosen because there are literate Okun speakers widespread from all the villages and towns collectively referred to as "Okun land".

Results
English language prepositional phrases were successfully translated to Okun language and prepositional phrases in Okun language were successfully translated to English language. A pictorial view of the developed system is shown in Figure 9 and 10.

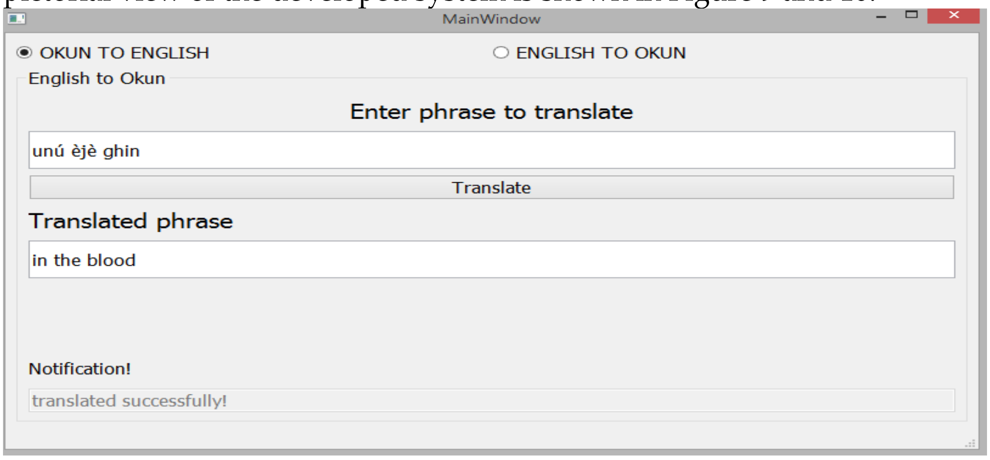


Figure 9: System Output for English language to Okun language translation

Figure 10: System Output for English language to Okun language translation

Result from evaluation of the Developed Rule-based English-Okun Machine Translator
The accuracy of the developed system was verified using word orthography (tone marking and spellings). Expert translated phrases were compared to the developed machine translator and the experimental respondents using the mean opinion score (MOS) technique. Results

show that the developed machine translator has ninety three percent (93%) accuracy while experimental subject respondents had 57.85 percent as shown in Table 4 and Figure 11.

Table 4: Results from Evaluation of the Developed Rule-based English-Okun Machine Translator

Phrases	Expert (%)	Respondents result (%)	Developed system's result (%)
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
Total	**100**	**57.85**	**93**

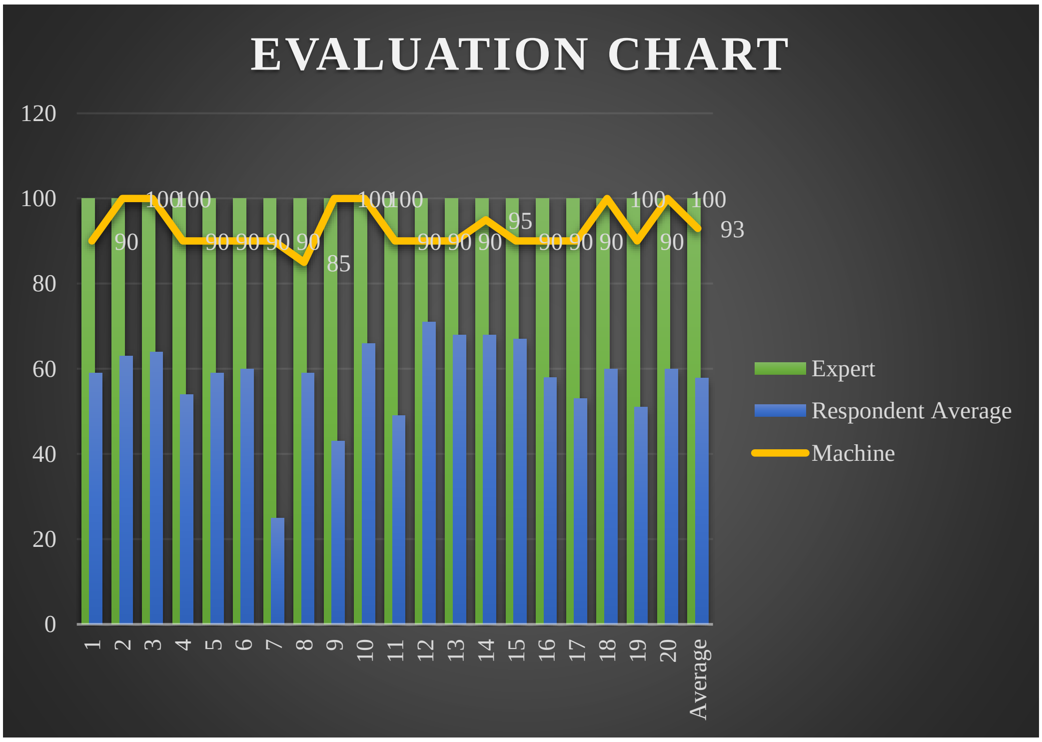


Figure 11: Translated phrases orthography accuracy

Discussion

The results obtained show that the experimental subject respondent's accuracy is significantly lower than the developed machine translator. This contradicts the research by Eludiora and Odejobi (2016) where the experimental subject respondent score is greater than that obtained from the machine translator. It therefore implies that most native Yoruba speakers can write Yoruba language than native Okun Speakers can write Okun language. In addition, the accuracy obtained from the developed system is lower than previous research by Esan et al. (2018) up to 2.5 %. This shows that rule-based approach performs better in translating English to Yoruba language than it does in English to Okun language translation. This research also revealed that rule-based system is mainly suitable for word-level translation but not effective for sentence-level translation as affirmed by Torregrosa et al. (2019). Also, the accuracy of the developed Rule-based system is high despite the inadequacy of parallel text used in the research, this shows that rule-based approach is good in handling languages with rich inflectional morphology like Yoruba language than statistical machine translators according to Macketanz et al. (2017) and Vičič (2013).

Conclusion

An English-Okun prepositional phrase Machine translation system based on rule based approach was developed in this research. The system was designed to enhance the learning of Okun language which is a dialect under Yorùbá language. It is user-friendly and enable learners learn the language at ease. The system was evaluated using human judgment based on orthography accuracy and it was found ninety three percent accurate. The research established that most indigenous people are not good at writing Okun language and find it hard to separate Okun dialects from Yoruba language. It was also deduced that rule based approach is mainly suitable for word-level translation and efforts should be geared at hybridizing the approach with machine learning or deep learning models to improve the accuracy.

References

Abiola, O.B, Adetunmbi, A.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.

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

Ayegba, S. F., and 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., and 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*.

Ayogu, I. I., Adebayo, O. A., and Bolanle , A. O. (2018). Developing Statistical Machine Translation System for English and Nigerian Languages. *Asian Journal of Research in Computer Science*, 1-8.

Badr, I, Zbib, R and Glass J. (2009)Syntactic phrase reordering for English-to-Arabic statistical machine translation. EACL '09: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics Pages 86–93

Bayatli, S., Kurnaz, S., Francis, M. and Tyers (2018). Rule-based machine translation from Kazakh to Turkish Proceedings of the 21st Annual Conference of the European Association for Machine Translation, p. 49–58 Alacant, Spain.

Charoenpornsawat, P., Sornlertlamvanich, V. and Charoenporn, T. (2002). Improving Translation Quality of Rule-based Machine Translation. Association of Computational linguistics.

Costa-Jussa`, M., Marino˜ , F., and Fonollosa J. (2017). Study And Comparison Of Rule-Based and Statistical Catalan-Spanish Machine Translation Systems

Ehab, R., Amer, E. and Gadallah, M. (2018). Example-Based English to Arabic Machine Translation: Matching Stage Using Internal Medicine Publications ICSIE '18: Proceedings of the 7th International Conference on Software and Information EngineeringMay 2018 Pages 131–135https://doi.org/10.1145/3220267.3220294

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

Eludiora S.I. and Odejobi O.A. (2016). Development of an English to Yorùbá Machine Translator. International Journal of Modern Education and Computer Science. (11) 8-19.

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

Esan, A., Oladosu, J., Adeyanju, A., Okomba, N. and Oforkansi, S. (2021): Development of a Syntax-Based Model for English - Igbo Statistical Machine Translation. *LAUTECH Journal of Computing and Informatics (LAUJCI).*

Espa˜na-Bonet, C. and Ranta, A. (2012). Rule-Based Machine Translation Proceedings of a Workshop Held in Gothenburg 14-15 June, 2012

Esan,, A.O, Omodunbi, B. Olaniyan, O. Odiase, P. and Olaleye, T. (2018). Development of Adjectival Phrase- Based English to Yorùbá Machine Translator. *International Digital Organization for Scientific Research Journal of Applied Sciences,* 3(1) 29-42.

Esan, A.O., Oladosu J.B., Oyeleye, C. A., Adeyanju, I. A., Olaniyan, O. M., Okomba, N. S., Omodunbi, B.A. and Adanigbo, O. (2020). Development of a Recurrent Neural Network Model for English to Yoruba Machine translation. International Journal of Applied Computer Science and Applications.

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

Klubi˘cka, F., Ram´Irez-S´Anchez, G., and Ljube˘, Si´ C, N. (2016). Collaborative development of a rule-based machine translator between Croatian and Serbian. *Baltic J. Modern Computing*.

Macketanz, V., Avramidis, E., Burchardt, A., Helcl, H. and Srivastava, A. (2017). Machine Translation: Phrase-Based, Rule-Based And Neural approaches with Linguistic Evaluation.
Bulgarian Academy of Sciences Cybernetics and Information Technologies; 17(2).

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

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.

Sreelekha, S., (2016). Statistical Vs Rule Based Machine Translation; A Case Study on Indian Language Perspective. World Journal of Computer Application and Technology. 4(4): 46-57.

Torregrosa, D., Pasricha, N., Chakravarthi, B., Masoud, M. and Arcan, M. (2019). Leveraging Rule-Based Machine Translation Knowledge for Under-Resourced Neural Machine Translation Models Proceedings of MT Summit XVII, volume 2Dublin, Aug. 19-23, 2019 pp 125-133.

Vičič, J. (2013). A Fast Implementation of Rules Based Machine Translation Systems for Similar Natural Language. Informatica; 37:455-456.