ESTIMATING SEMANTIC SIMILARITY IN YORUBA SENTENCES USING PATH-BASED METRICS

Adenike ADEGOKE-ELIJAH[1], Theresa OJEWUMI[2], Kudirat JIMOH[3]
[1,2]Redeemer's University, Ede, Nigeria
[3]Osun State University, Osogbo, Nigeria
[1]adegoke-elijaha@run.edu.ng, [2]ojewunmit@run.edu.ng, [3]kudirat.jimoh@uniosun.edu.ng

Keywords: Yoruba, edge-counting, sentence similarity, ontology, low-resource language

Abstract: *Measuring semantic similarity among texts is an important task in many Natural Language Processing applications such as information retrieval, text summarization. However, there is dearth of work in the development of the tool for Yoruba Language, and this has therefore limited the advancement of Yoruba Language Engineering. This study addressed the gap by using a knowledge-based approach based on lexical resources. A total number of 434 nouns were collected from home-domain. The nouns were grouped into hypernym semantic classes. The classes were thereafter organized in hierarchy to form a taxonomy for Yoruba nouns and concepts. The model for the measurement of semantic similarity in Yoruba sentences was thereafter developed using path-based similarity measurement between the concepts represented in the taxonomy. Using the model, the system was implemented using python programming language. The developed system was evaluated using accuracy mean opinion score, and a score of 73.2% was achieved.*

1. INTRODUCTION

Semantic similarity is a computational task that estimates the similarity between texts or documents using predefined metrics. This task is essential in natural language applications such as information retrieval [1]. In word sense disambiguation, semantic similarity is used to distinguish between senses based on the hypothesis that similar words occur in similar contexts [2]. In the task of text summarization, semantic similarity is used to ensure that the summarized text is semantically related to the original text [3] etc. Given that words are the smallest unit of documents or sentences, the similarity between sentences or documents mostly depends on the similarity between words [4]. While very few studies have been done in developing semantic similarity tools for many resource- scarce languages, such as Yoruba languages, the majority of existing word similarity measures have been developed and are utilized for English texts. Existing works on semantic similarity can be categorized into three viz: document similarity, sentence similarity and word similarity. If two documents describe related ideas and are semantically comparable, they are deemed similar. The similarity between two concepts can be evaluated based on the similarity of the embedding corresponding to their textual contents [5]. Sentence similarity measures the semantic relatedness between two sentences using quantitative metrics [6], and has wide applications in tasks such as text search, natural language understanding and machine translation. Word similarity consists of computing the semantic likeness between words and senses, and it has its application in the task of distinguishing the meaning of words such as word sense disambiguation [7].

Yoruba is spoken not just in Nigeria, where there are about 30 million native speakers, but also in Togo, the Republic of Benin, Ghana, Sudan, Sierra Leone, and Côte D'Ivoire. Numerous people who speak the language also reside outside of Africa in nations like Brazil, Cuba, including Trinidad & Tobago [8]. Despite its number of speakers, it is classified as a resource scarce language because it lacks major machine-readable resources essential for natural language processing. This could be due to the interaction of tone and syntax in the language which makes low-tone monosyllabic verbs change to mid low when used in certain contexts [9], the dearth of optical character recognition system, the lack of input devices for texts with diacritics etc. One of the tools not available in the language is a tool for measuring semantic similarity between sentences. This study therefore addressed this gap, and presents a tool that could aid many Yoruba language processing tasks such as Information Retrieval, word sense disambiguation, etc. The rest of the paper is organized as follows: Section 2 reviews some literature to see methods used in existing studies, section 3 describes the method used in this study. Section 4 discusses the results of the evaluation of the developed system, while the paper is concluded in section 5.

2. LITERATURE REVIEW

Yoruba has many dialects [10] [11]; however, there is a consensus form of the language which is understood by all speakers of the language and also used in newspapers. This is known as *Standard Yoruba.* This study is focused on standard Yoruba. It is a tonal language and therefore uses pitch patterns to distinguish individual words or grammatical forms of words [12]. There are three (3) tones in the language which are represented with grave (\), acute (/) and no symbol () respectively. Tone information is needed to aid the pronunciation of words [13]. A Yoruba word can be broken down into many units called syllables. Each of the syllables carries the tone mark. A syllable marked with an acute sign (/) is pronounced with high tone, low-toned syllables are marked with grave sign (\). Any syllable with no mark is pronounces with mid-tone. Yoruba words can be monosyllabic, disyllabic or polysyllabic [11].

A Yoruba sentence is made of a noun phrase and a verb phrase [14], and exhibits Subject-Verb-Object word order [12]. Yoruba sentences are made up of words which belong to different grammatical classes such as noun, verb, adjectives, preposition, conjunction etc. [15]. This study made extensive use of the nouns found in Yoruba sentences to measure the similarity between two sentences that contain the same verb. This is because many Yoruba verbs exhibit polysemy [16], and can therefore be used in different contexts to represent different meanings.

Words are the building blocks for sentences, and therefore similarities among words could be used as clues to estimating similarity between sentences containing them. This study therefore reviews some existing work in words similarity measurements.

2.1 Words Similarity Measurement

[17] classified word similarity measurement methods into four. These are Corpus based approach, knowledge based, structural based approach and deep-learning based approach. Corpus based approach utilizes relevant information extracted from large amount of data for words similarity measurements. It makes use of Latent Semantic Analysis (LSA), to statistically analyze big corpus through counting of words in the corpus. Words are represented in vectors using statistical computation. This is with the assumption that words used in the same contexts, have similar meanings [18]. Structure based approach is hinged on the assumption that similar sentences should have similar meaning [19]. Deep-learning based approach makes use of a language model trained to capture the semantic space of a language [6], the assumption is that words with similar measures should have close vector in semantic space. Knowledge based approaches make use external resources such as ontology and other lexical resources, and measures the similarity between two concepts using their closeness in an ontology or taxonomy [20]. [21] classified the methods of measurement used in knowledge-based approach into Path length based, Depth relative measure, Information content-based measure, hybrid measure and feature based measure. In path-based similarity, the measurement of similarity is based on the structure of an ontology or taxonomy, by counting the number of edges between two concepts in comparison [22], [23]. Depth relative measurement makes use of shortest path approach, but also takes into account the overall structural depth of the edges joining the concepts in an ontology [24]. Information content-based measure improves the knowledge existing in the ontology or taxonomy with knowledge extracted from a corpus [25]. Hybrid measure combines knowledge gathered from multiple sources of

information for improved similarity measurements [26]. Feature based approach considers features similar to both concepts, and also the distinguishing features of each concept to estimate the semantic similarity between concepts [27].

2.2 Sentence Similarity Measurements

Existing approaches of measuring similarity in sentences can be classified into three. These are Word based similarity, structure-based similarity and vector-based similarity. Word based similarity measures the similarity by considering a sentence as a set of words. Structure based similarity makes use of the information obtained from the structure of the sentences for the similarity measurement. Vector based approach uses statistical approach to generate a vector representation for each sentence, and subsequently estimates the similarity in the vectors.

This study used a word-based similarity approach to estimate the similarity between sentences by summing up the similarities in the composite words calculated over an ontology using [28] path-based measurement. The higher the value, the greater the similarity among sentences. Afterwards, the study made use of Mean Opinion Score for the evaluation of the system. Opinion Score is the score value, on a predefined scale, that a subject assigns to his opinion on a subject matter [29]. Mean Opinion Score is therefore the average of these scores across subjects. The 5-point popular predefined scale used for MOS are Excellent-1, Good-4, Fair-3, Poor-2, Bad-1. Despite the popularity of MOS for system evaluation, it is often used without sufficient consideration of how the data was obtained, and the inherent limitation imposed by the design of the subjective test. This study therefore carefully designed the subjective test to enhance the evaluation task.

3. METHODOLOGY

The method used in this study is made up of several processes. These include collection of Yoruba noun data, classification of the collected data into an ontology, digitization of the ontology and the implementation of the similarity measurement.

3.1 Data Collection

The total of four hundred and thirty-four nouns were collected from home domain and grouped into different classes. These include nouns that are used in everyday conversation. *Yoruba* words carry tone marks which include accent acute, accent grave and under dot. The collected noun data were therefore pre-processed with *Takada* text editor to facilitate the addition of the appropriate tone marks to the words. Sample of the collected nouns and their classes are shown in Table 1.

Table 1: Sample Noun Data Collection

Class	*Examples*
Abstract	*Iwà (character), ìmọ̀(knowledge)*
Wears	*fìlà(cap), bàtà(shoes)*
Properties	*Ile(houses),mọ́tò(car)*
Woods	*Aga (chairs), tábìlì(Table), ìbùsùn (Bed)*
Food	*Iyán (Pounded Yam), isu (Yam)*

3.2 Hierarchical Classification of Yoruba Nouns

The collected nouns were ordered in hierarchical form, starting from the general term till it reaches the specific terms. This is shown in Fig. 1. At the topmost of the hierarchy, is the class *Top* which represents the general class of all Yoruba nouns. The top class is further divided into *abstract* and *concrete* classes. The *abstract* class contains items which cannot be perceived with the five (5) senses. The *concrete* class contains items which are tangible and can be perceived with the human's sense organs. The *concrete* class is further sub divided into *living thing, non-living thing and location* sub-classes. The *living thing* subclass contains items which have life. The *non-living thing* subclass contains items which have no life. The *location* sub-class contains items which depict geographical locations. The *living* thing subclass is further divided into major classes *plant*, *human* and *animal*. The *plant* class contains items which are types of plants. The *Human* class contains items that are related to human beings. The *human* class is linked to subclasses *names*, *profession*, *pronoun* which are used in place of a person's name. The *body part* class contains items which are parts of human body. The *animal* major class contains items which are names of animals. The *non-living thing* subclass is divided into major classes solid, liquid and gas. The *solid* class is further subdivided into *wears, properties, woods, food* and *tools* minor classes. The *wears* class contains items which can be worn by human beings. The *properties* class contains items which can be regarded as valuables or human possessions. The w*ood* class contains items which have wooden origins. The *Food* class contains names of *Yoruba* foods. The *Tools* class contains items which are simple machines used in different places. The l*iquid* class contains items which are fluid in nature. The *Gas* class contains items which cannot be seen but can be felt. The *Location* subclass is further divided into *common* and *proper* major classes. The *proper* class contains name of specific village, town, city or country. The *common* class contains items which are name of places that are not specific to a particular location. Using graph theory notation, the ontology contains seventeen (17) leaf nodes which can be used to define the total number of classes that Yoruba nouns can be grouped into.

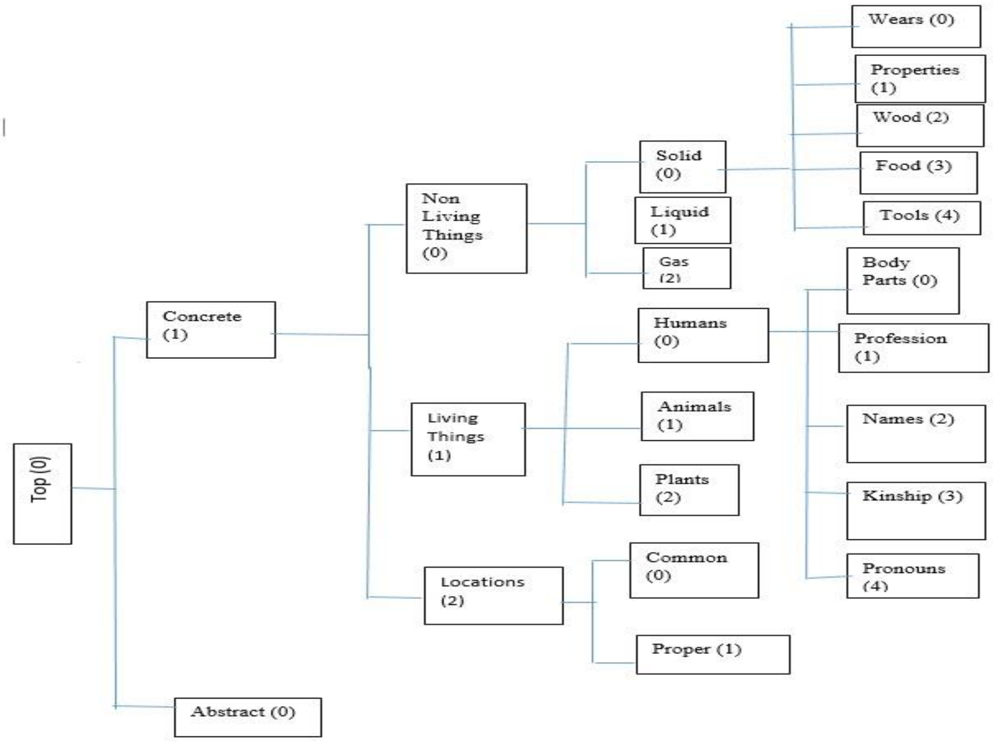
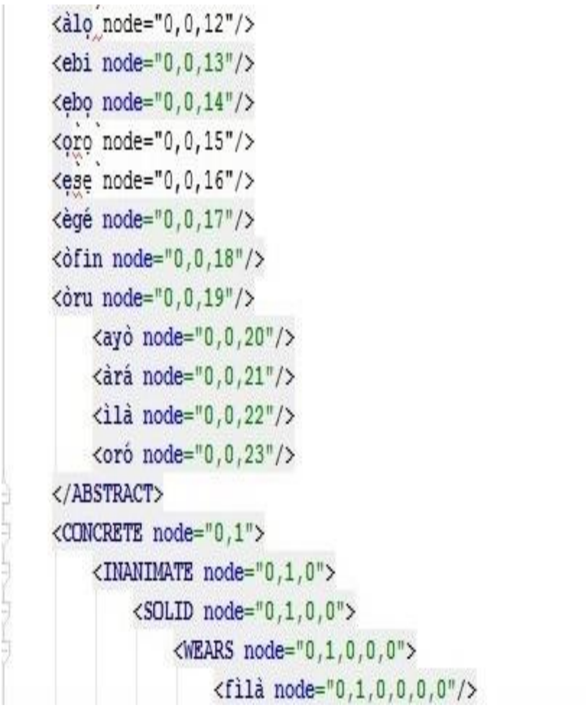


Figure 1: Hierarchical Classification of Yoruba Nouns.

3.3 Digitization of the Ontology

The digitization process involves the conversion of the taxonomy into a machine readable format. Precisely, this involves representing the taxonomy in extensible mark-up language (XML). Each of the concepts in the taxonomy is represented as an XML tag with the attribute node. The value of an attribute is a sequence of numbers separated by commas. Each of the numbers depicts the relative distance or depth of a concept to the root node, Top. The leaf nodes at each level of the hierarchy are labelled from *0……n,* where n represents the number of sibling nodes occurring at a particular level. The attributes of all the concepts starts with 0, which is the root node Top. As a way of illustration, the attribute of the concept fila has value *0,1,0.0,0,0.* This shows that to access the concept *filà*, we have to transverse the graph from nodes *Top, Concrete, Non-living thing, Solid* and *wears*, where *Top* is the root node, and *wears* is the leaf node. The first five digits 0,1,0,0,0 in the attribute shows the index number of each of the nodes transverse before fila could be accessed, while the last digit 0 is the index number of *filà*: among all the items belonging to the class wear, Figure 2 shows a snapshot of the digitized ontology.



```
    <àlọ node="0,0,12"/>
    <ebi node="0,0,13"/>
    <ẹbọ node="0,0,14"/>
    <ọ̀rọ̀ node="0,0,15"/>
    <ẹ̀sẹ̀ node="0,0,16"/>
    <ègé node="0,0,17"/>
    <òfin node="0,0,18"/>
    <òru node="0,0,19"/>
        <ayò node="0,0,20"/>
        <àrá node="0,0,21"/>
        <ìlà node="0,0,22"/>
        <oró node="0,0,23"/>
    </ABSTRACT>
    <CONCRETE node="0,1">
        <INANIMATE node="0,1,0">
            <SOLID node="0,1,0,0">
                <WEARS node="0,1,0,0,0">
                    <fìlà node="0,1,0,0,0,0"/>
```

Figure 2: Digitization of the Ontology

3.4 Estimating Sentence Similarity

In this study, the semantic similarity among concept is estimated using [28] method. Here, the similarity among concept is quantified using conceptual distance. It is used to quantify the geometric distance between points representing concept. Conceptual distance between concepts is inversely proportional to the similarity among them, i.e. the lower the conceptual distance, the greater the similarity. The conceptual distance among concept is computed by counting the number of edges between the edges in the taxonomy.

Let C_1 and C_2 be two concepts in a hypernym network, the conceptual distance between C_1 and C_2 is given by

Distance (C_1, C_2) = minimum number of edges separating (C_1, C_2) (1)

Pathlength (C_1, C_2) = 1 + number of edges in the shortest path between C_1 and C_2 in the hypernym graph (2)

$$Simpath(C_{1.}\ C_2) = \frac{1}{pathlen(\mathrm{C1,C2})} \qquad (3)$$

Let S_t denotes the target sentence for which similarity is to be measured against, and Si represents other sentences from 1 to *n* with m arguments, and *j* is the counter numbering the arguments from 1 to m.

$$Simpath(S_{t,}\ S_i) = \sum_{j=1}^{m} \frac{1}{Pathlen(C_{tj,}C_{ij})} \qquad (4)$$

Equation 4 is a function used to measure the semantic distance between S_t and Sentences S_{1-n}, the sentence with the greatest semantic similarity is chosen as the most similar to sentence St. For example, given the test sentence *Òjó bọ́ sòkòtò*, for which we seek to calcualate the semantic similarity to the sentences, *O bọ́ sí ilé, Tolú bọ́ asọ rẹ̀, Adé bọ́ ọmọ rẹ̀* and *Òjó bọ́ sí yàrá*.

S_t= *Òjó bọ́ sòkòtò*

S_1= *O bọ́ sí ilé*

S_2= *Tolú bọ́ asọ rẹ̀*

S_3= *Adé bọ́ ọmọ rẹ̀*

S_4= *Òjó bọ́ sí yàrá*.

Using equation 4, the similarity between $S_{t.}$ and each of the sentences S1 to S4 is as shown below:

$Simpath(S_{t,}\ S_1)$ = 0.75

$Simpath(S_{t,}\ S_2)$ = 1.5

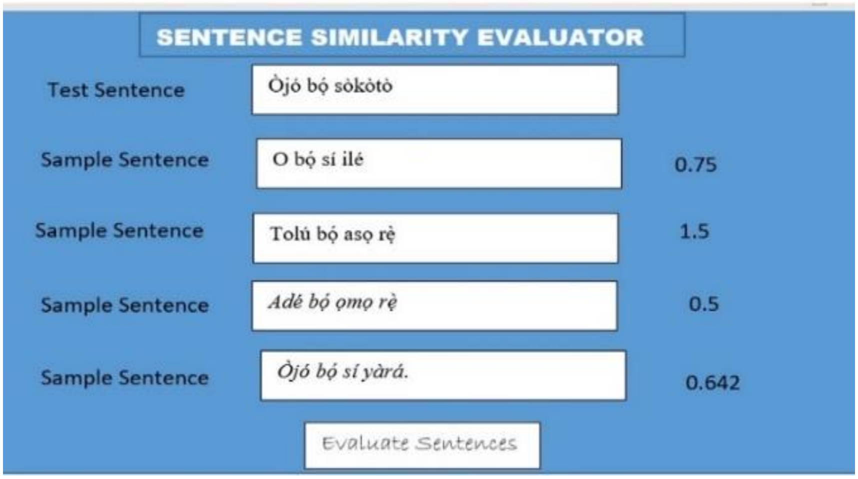
$Simpath(S_{t,}\ S_3)$ = 0.5

$Simpath(S_{t,}\ S_4)$ = 0.642

Sentence 2 has the greatest similarity with S_t, since it has the highest value using $Simpath$ function.

4. SYSTEM IMPLEMENTATION

The system was implemented using Natural Language Toolkit (NLTK) and Tkinter libraries of Python programming Language. This consists of many functions which are the building blocks of the developed software. Some of the major functions are:



Test Sentence: Òjó bọ́ sòkòtò

Sample Sentence: O bọ́ sí ilé 0.75

Sample Sentence: Tolú bọ́ asọ rẹ̀ 1.5

Sample Sentence: *Adé bọ́ ọmọ rẹ̀* 0.5

Sample Sentence: *Òjó bọ́ sí yàrá.* 0.642

Figure 3: Interface of the developed software

(i.) *CreateWidget()* was used for the creation of the Graphical User Interface.

(ii.) *VirtualKey()* used for the creation of virtual keyboard to make entries of diacritics found in Yoruba words possible.

(iii.) *getcaseNouns()* lists the nouns found in the entered Yoruba sentence.

(iv.) *getWordsDistance*() was used to calculate the number of nodes between two nouns using the ontology

(v.) *semanticSimilarity()* was used to estimate the similarity scores between two nouns by computing the inverse of the output of getWordsDistance() function

The system contains text boxes which allow users to input a test sentence, and the sample sentences for which similarity measurement is sought. It also contains a command button, which when clicked, displays the sentences ranked in order of similarity to the test sentence. This is shown in Figure 3.

5. SYSTEM EVALUATION

Although, many studies have been done in the creation of dataset for Yoruba language in different forms such as in reading comprehension [30], semantic corpus for corpus reviews [31], printed text images [32], Yoruba speech [33], Speech to image dataset [34], there is presently no report on the existence of similarity dataset for Yoruba sentences; the closest to this is the dataset report in [35], but it only estimates the similarity in English words found in the word-pairs. This study therefore opted for the use of Mean Opinion Score (MOS). The similarity test is said to be accurate if, according to human’s judgement, the similarity ranking of the sample sentences in relation to the test sentence is accurate. Ten test sentences, and the output of the similarity rank with sample sentences done by the developed system were given to ten (5) native speakers of the language to evaluate the developed system for accuracy. It was discovered that the accuracy of the system depreciates with increased number of words in the sentence. Higher accuracies were achieved with Subject-Verb-Object (SVO) sentences. Another observed limitation of the system is that the similarity measurement depreciates with difference in the number of words used for the test and sample sentences. That is, it expects the same number of words in both the test and sample sentence. Table 2 shows the rating of the accuracy by the users on the scale of 1-5 (1-poor, 5-excellent). The accuracy of 3.66 was achieved, which equates to 73.2%.

Table 2: System Evaluation

Users	Accuracy rating
1	3.8
2	3.1
3	3.9
4	3.5
5	4.0
MOS	3.66

6. CONCLUSION

This study concludes that a computational tool for measuring similarity between two Yoruba sentences was developed using a knowledge-based approach that sums up similarity among words to connote the similarity in sentences. The system was developed without using an expensive similarity labeled dataset which is presently not available in the Language. Another by-product of this study is an ontology that represents Yoruba nouns in hierarchy using hypernym relationship. The development of this tool has contributed to Yoruba Language Engineering.

7. REFERENCES

[1] Hliaoutakis, A., Varelas, G., Voutsakis, E., Petrakis, E. G., & Milios, E. . “*Information Retrieval by Semantic Similarity. International journal on semantic”* Web and information systems (IJSWIS), *2*(3), 55-73. 2006

[2.] Dongsuk, O., Kwon, S., Kim, K., & Ko, Y. “*Word Sense Disambiguation based on Word Similarity Calculation using Word Vector Representation from a Knowledge-based Graph.”* In Proceedings of the 27th international conference on computational linguistics, 2704-2714, 2018.

[3] Kumar, S., & Bhatia, K. K. “*Semantic Similarity and Text Summarization based Novelty Detection”* SN Applied Sciences, *2*, 1-15, 2020

[4]. Alian, M., & Awajan, A. “*Arabic Semantic Similarity Approaches-Review.”* In 2018 International Arab Conference on Information Technology (ACIT), Lebanon, 1-6, 2018

[5.] Agarwala, S., Anagawadi, A., & Guddeti, R. M. R. “*Detecting Semantic Similarity of Documents using Natural Language Processing.”* Procedia Computer Science, 189, 128-135. 2021

[6.] Sun, X., Meng, Y., Ao, X., Wu, F., Zhang, T., Li, J., & Fan, C. “*Sentence Similarity based on Contexts.”* Transactions of the Association for Computational Linguistics, 10, 573-588, 2022

[7] Navigli, R., & Martelli, F. “*An Overview of Word and Sense Similarity”.* Natural Language Engineering, *25*(6), 693-714, 2019

[8] Balogun, T. A. “*An endangered Nigerian Indigenous Language: The case of Yoruba language.”* African Nebula, *6*, 70-82. 2013

[9] Salawu, A. “*A Linguistic Discourse on the interaction of Tones and Syntax in Yoruba.”* Unizik Journal of Art and Humanities, 44: 195-216, 2004

[10] Asaiah F. O. *“Development of a Standard Yoruba Digital Text Automatic Diacritic Restoration System”* PhD Thesis, Obafemi Awolowo University, Ile-Ife, Nigeria, 2014

[11] Fagbolu, O, Ojoawo. A., Ajibade. K., and Alese, B. *“Digital Yoruba Corpus*”. International Journal of Innovative Science, Engineering Technology. 2(8): 918-926, 2015

[12] Okanlawon J. *“An Analysis of the Yoruba Language with English”* Retrieved from., Retrieved from An analysis of the Yoruba language with english (northeastern.edu), accessed 21.01.2024, 2016

[13] Asahiah, F. O., Odejobi, O. A., & Adagunodo, E. R.. *“Restoring Tone-marks in Standard Yorùbá Electronic text: Improved model”* Computer Science, 18(3), 2017

[14] Yusuf O. *“Fundamentals of Syntax and the Studies of Nigerian languages”,* Shebiotimo Publications, ljebu-ode, Nigeria. 1998

[15] Awóbùlúyi O. (2008). *“Ẹ̀kọ́ Isẹ̀dá Ọ̀rọ̀ Orúkọ”*. Montem Paperbacks, Akure, Ondo state, Nigeria, 2008

[16] Adégòke-Elijah A., Qdéjobi O. and Saláwú A. “*Lexical Ambiguity Resolution for Standard Yorùbá verbs*.” American Journal of Engineering Research (AJER), 7:170-176, 2018

[17] Farouk, M. “*Measuring Sentences Similarity: a survey*.” *arXiv preprint arXiv:1910.03940*, 2019

[18] Islam, A., & Inkpen, D. “*Semantic Text Similarity using Corpus-based Word Similarity and String similarity”.* ACM Transactions on Knowledge Discovery from Data (TKDD), *2*(2), 1-25, 2008

[19] Lee, M. C., Chang, J. W., & Hsieh, T. C. “*A Grammar-based Semantic Similarity Algorithm for Natural Language Sentences.*” The Scientific World Journal, 2014.

[20] Oussalah, M., & Mohamed, M. (2022). “*Knowledge-based Sentence Semantic Similarity: Algebraic properties.*” Progress in Artificial Intelligence, *11*(1), 43-63, 2022

[21] Elavarasi, S. A., Akilandeswari, J., & Menaga, K. *“A Survey on Semantic Similarity Measure.”* International Journal of Research in Advent Technology, *2*(3), 389-398, 2014.

[22] Cai, Y., Pan, S., Wang, X., Chen, H., Cai, X., & Zuo, M. “Measuring Distance-based Semantic Similarity using Meronymy and Hyponymy relations” Neural Computing and Applications, 32, 3521-3534, 2020

[23] Gan, M., Dou, X., & Jiang, R. *“From Ontology to Semantic Similarity: Calculation of Ontology-based Semantic Similarity.”* The Scientific World *Journal*, 2013.

[24] Taieb, M. A. H., Aouicha, M. B., & Hamadou, A. B. *“Ontology-based approach for Measuring Semantic Similarity”,* Engineering Applications of Artificial Intelligence, *36*, 238-261, 2014.

[25] Sánchez, D., & Batet, M. *“A Semantic Similarity Method based on Information Content Exploiting Multiple Ontologies”* Expert Systems with Applications, *40*(4),1393-1399, 2013

[26] Gao, J. B., Zhang, B. W., & Chen, X. H. *“A WordNet-based Semantic Similarity Measurement combining Edge-counting and Information content theory”.* Engineering Applications of Artificial Intelligence, 39, pp 80-88, 2014.

[27] Solé-Ribalta, A., Sánchez, D., Batet, M., & Serratosa, F. "*Towards the Estimation of Feature-based Semantic Similarity using Multiple Ontologies."* Knowledge-Based Systems, *55*, 101-113, 2014

[28] Rada, R., Mili, H., Bicknell, E., & Blettner, M. "*Development and Application of a Metric on Semantic nets.*" IEEE transactions on systems, man, and cybernetics, 19(1), 17-30. 1989

[29] Streijl, R. C., Winkler, S., & Hands, D. S. *"Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives*" Multimedia Systems, *22*(2), pp. 213-227, 2016

[30] Aremu, A., Alabi, J. O., & Adelani, D. I. (2023). *"YORC: Yoruba Reading Comprehension dataset."* arXiv preprint arXiv:2308.09768, 2023

[31] Shode, I., Adelani, D. I., & Feldman, A. *"YOSM: A new Yoruba Sentiment Corpus for Movie Rviews." arXiv preprint arXiv:2204.09711*., 2022

[32] Oni, O. J., & Asahiah, F. O. (2020). "*Computational modelling of an optical character recognition system for Yorùbá printed text images"* Scientific African, *9*, e00415.

[33] Guthin, A., Demirsahin, I., Kjartansson, O., Rivera, C. E., & Túbòsún, K. *"Developing an open-source corpus of Yoruba speech"* In Proceeding of International Speech Communication Association (ISCA), Shanghai, China, 404-408. 2020

[34] Olaleye, K., Oneaţă, D., & Kamper, H. "YFACC: "*A Yorùbá Speech–Image Dataset for Cross-Lingual Keyword Localization Through Visual Grounding"* In 2022 IEEE Spoken Language Technology Workshop (SLT) Doha, Qatar ,731-738, 2023

[35] Alabi, J., Amponsah-Kaakyire, K., Adelani, D., & Espana-Bonet, C. "*Massive vs. curated embeddings for low-resourced languages: the case of Yorùbá and* "Twi. In Proceedings of the Twelfth Language Resources and Evaluation Conference, France, 2754-2762, 2020