NIU Journal of Humanities Copyright©2023
Nexus International University ISSN: *3007-1704*; 8 (1): 33-51

Human Evaluation of Yorùbá Noun Ontology

AKÍNDÉLÉ ÀKÀNJÍ ÀÌNÁ
Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria

ADEOLA MOYOSOLUWA AKINWALE
Federal University of Agriculture, Abeokuta, Ogun State, Nigeria

Abstract. African Scholars have engaged in Ontological annotation to present explicitly the content of concepts in a domain by using formal machine-readable labels which must conform to an agreed standard in order to achieve interoperability of Natural Language Processing (NLP) systems. As good as this development is, there is the need to evaluate the contents of the concepts labelled for semantic web activities. This paper therefore presents an evaluation report carried out in building a hybrid semantic web engineering system for Yorùbá nouns. Human evaluation method through the statistic evaluation tools was employed. To achieve the desired objective of proper human evaluation, two different technical workshops were held on the 13th and 20th April 2018 at the University of Ibadan, bringing twenty (20) Linguistics Scholars together to examine and validate the extracted lexical entries before annotation. After thorough scrutiny, each of the scholars evaluated the entries with the use of questionnaire. The SPSS data analysis software was used to analyse the responses to different question statements based on the five (5) variables of: Accuracy, Completeness, Semantic error freeness, Orthographic error freeness and Ambiguity freeness rating. Eight (8) questions were also used to measure the experts' perception on the usefulness of the model. The t-test on the experts' perception of the five variables indicated that none of the respondents has poor perception of the software. The study therefore concludes that thorough evaluation procedure is essential to benchmark the internal lexical contents for building ontology models.

Keywords: Yorùbá noun, Ontological annotation, Evaluation, Natural language processing

1. Introduction

Ontological annotation model in Natural Language Processing (NLP) makes explicit the content of concepts in a domain, using a set of arbitrary tags or labels which must conform to an agreed standard. The modern-day quest for technological innovation triggered language Scientist and experts to expandt their horizons in designing and building knowledge-based systems and machines to aid human tasks. Power (2002), Marakas (1999), Ruskova (2002) Several other models are developed for enhancing performance of the Semantic Web in different domains, each being specified for the purposes which its being built for. (Hutchins 1994, 1995, 2007). Others are developed as expert systems capable of providing efficient human- machine linguistic performances and analysis, and for providing a theory-neutral backbone in the processing of scientific documents pertaining to the linguistics domain. Andre and Rist (1993). Automatic evaluation techniques are as well-developed alongside these models to monitor effective performance. Odoje (2016), Ezeani, Onyenwe and Mark (2020). Onyenwe, Uchechukwu, and Hepple (2014). Interestingly, in order to pave the way for further inter-operable Natural Language Processing for Yorùbá language data, Aina (2019a, 2019b, & 2020), Aina and Taiwo (2019) implemented semantic web annotations for different concepts in Yorùbá language scholarship. Efforts to ensure accuracy of the contents of the built model form the basic reason this study is conducted.

1.1 Statement of the problem

With the background developmental tasks noted earlier, hidden and unperceivable statistic and stochastic errors in automatic evaluation techniques, the subjective and heterogeneity nature of Yorùbá language data, brings the need to complement the in-built automatic evaluation phase in Protégé (the editing tool used for implementation of Yorùbá noun annotation) with human evaluation methods. Also, capturing the intuitive knowledge of Yorùbá noun into a model as done by Aina (2019a & 2020) are highly resourceful in executing the development of Yorùbá language especially with the modern information and technological trend. However, a problem arises if intelligible and performative knowledge of the professionals in the field is not endorsed at the background of the models. The language specialist needs to be given the avenue to scrutinize the modelled language data. Effort to achieve the solution to this peculiar problem becomes the focus of this paper.

1.2 Objectives

The objective of this paper therefore is to report the scholastic scrutiny of the language specialists, with the view of manually checking, scrutinizing and evaluating the Yorùbá noun language data modelled and assess its accuracy, completeness, semantic error freeness, orthography error freeness, ambiguity freeness, its relevance to the content quality and the experts' perception of the usefulness of the annotation.

1.3 Research question

The paper provides answers to the following questions:
- What are the standard tools and techniques in evaluating the language data modelled for Yorùbá Noun ontology annotation?
- How accurate, complete, semantical, orthographical and ambiguity error free are the language data modelled for the Yorùbá Noun Ontology?

1.4 Research Methodology

In order to ensure that the Yorùbá language data content in the developed model is highly accurate. Two different validation and evaluation workshops, with Twenty (20) participants in each of the sessions, were organised in respect of the Yorùbá nouns modelled. In the first of the workshop, all participants that were drawn sparsely from different dialectal regions of Yorùbá territory had to go over the extracted terms line by line considering the mutual intelligibility of the data, The necessary correction were made before we proceeded to model the data. The presentation and analysis of the results of the human evaluation during the first and second workshop is the basis of this paper. Each participant was presented with a questionnaire of three (3) sections with a total of thirty-two (32) questions. The on-screen display of the model data is projected for all the participants to view and answered the question in the survey based on their assessment. The results were analyse using SPSS software made for statistical analysis of data. The results of the analysis are presented subsequently.

1.4.1 Workshop Participants

Twenty language experts were gathered together in a workshop, as a language study group Under the Supervision of a Professor of Linguistics. The language study group is formed in the University of Ibadan, Nigeria. This university is located in the southwest of Nigeria; it is the Premier University in Nigeria, an offshoot of University College London founded in 1948. It is well known for its specialties that ranked first in returning highest number of postgraduate degrees among its counterparts. The language specialists for the validation workshop in the study came from five (5) south western states of Nigeria where Yorùbá language and its dialects are spoken. They are highly intelligible users of the language and nearly all the participants have carried out many research works on the language. All participants have at least a Master degree in Linguistic as revealed by the demographic information provided in questionnaire.

2. Yorùbá Noun Ontology (YORNO) in Brief

The work presented Ontological annotation of concepts of Yorùbá nouns in explicit description, converted in machine readable format to serve other web searching agents. The set of instructions in this label or tag otherwise known as annotation is useful for other Natural Language Processing (NLP) purposes. The work adopted Methontology and Generative Lexicon (GL) theories as framework. Interpretive design was used. Four Yorùbá texts (Awobuluyi's *Essentials of Yorùbá Grammar* and *Ẹ̀kọ́ Gírámà Èdè Yorùbá*, Bamgbose's *Fonọ́lọ́jì ati Gírámà Yorùbá,* and *Modern Yorùbá Dictionary*) were purposively selected for their semantic and syntactic classification of Yorùbá nouns. The informally perceived domain knowledge of Yorùbá nouns from the four texts were extracted using

intermediate representations based on tabular and entity-relation notations to design three models: Yorùbá nouns according to Awobuluyi (YORNOA), Yorùbá nouns according to Bamgbose (YORNOB), and the hybrid Yorùbá noun Ontology (YORNO). Protégé 4.5 (a semantic web editing tool) was used to implement the ontological annotation.

The component features of Yorùbá nouns isolated and tagged from YORNOA and YORNOB were concept terms, subclasses, groups, attributes and relations. YORNOA has 23 concept terms, 13 subclasses, 10 groups, 13 attributes and two symmetrical relations while YORNOB has 20 concept terms, 12 subclasses, 10 groups, 13 attributes and 12 relations (four contrastive and eight non-contrastive). The concept terms, sub-classes, groups, relations and attributes in YORNOA and YORNOB were nested together in YORNO as a hybrid model. The implementation of the three models at the final edge revealed that some items classified as nouns in YORNOA are inconsistent and/or incompatible with the qualia structure of GL. Some examples like manner nouns [*kíákíá* and *wéréwéré* (quickly)], and polymorphic nouns [*mo* (I), *mi (*me), *o* (you), *ó* (he/she), *ìrẹ (*you), *irẹ̀* (him/her), *wọn (*them)] cannot be modelled as nouns because they do not exhibit independently the qualia properties and attributes required of nouns, unless they are stored dependently and linked to the subject they replaced. They are modelled as nouns in YORNOA only. The three models render nouns less ambiguous in machine learning because the semantic meanings of the lexicon have been defined and annotated. However, some semantic properties like 'mass', 'discreet' 'size' and 'individual' nodes were queried and unfilled in the qualia structure defined for YORNOA.

The Ontology-driven models for Yorùbá grammar serve as repositories of data and shared terminology to support manual and automated web-based Natural language processing activities. Those models solve the problems of ambiguity for machine learning.

3. Review of Literature

Annotation activities involve adding short comments to a book or piece of writing in order to explain parts of it. In essence, annotation provides comments, notes or explanation (usually short and coded) to a lexical item, text or a book. Zauzich (1992:20) related that annotation efforts started during the Ancient Egyptian writings in which an equivalent determinative was to be tagged at the end of the serial order to make clear the ambiguities of the contextual meaning. In essence, annotations were added to text to identify and decode appropriately the meaning of ambiguous fact in it. There were two broad approaches to annotation namely: The linguistic annotation and the computational methods. Each of these methods has the sub-components. Aina (2020: 20). However, the Computational approaches developed through stages of Generalised Markup Languages, (GML), and Standard Generalised Markup Languuages (SGML) to semantic web annotations. Pareja-Lora (2012).

The definition of ontology as a concept is well discussed in Guarino, Oberle and Staab (2009:10), Gruber (1993:200), Borst (1997:20), Studer, Benjamins and Fensel (1998:4) and others. Despite the bulky of dichotomies in the focus of its definitions because it cuts across different field of studies and each point to the fact that the classification of any semantic web annotation is based on set of attribute and properties which must be defined on a set of rules and constraint known as axioms, and it must be generally accepted by the members of the participant in the community knowledge known as the Domain of Discourse (DOD). Adekoya (2010).

Further to the above, other works on evaluation of machine constructed models includes Odoje (2016). The work aimed that machine translation activities are not enough but its evaluation is needed so as to monitor its improvement, particularly in fluency, accuracy and efficiency. Using the Ibadan and Akungba Structured Sentence Paradigm, the paper evaluates two translation activities of the Google Translate and the human translation. The findings reveal that human translation fares better in terms of accuracy and fluency. It's being made perfect by the quality and the quantity of training data. The semantic web annotation described in Aina (2020) is a means to an end of achievements to other Natural language processing techniques including Machine translations, failure to evaluate properly the contents annotated for web activities will in turn affect other web resources developed from it. However, we share the opinions that thorough human evaluation increases the efficiency, accuracy and quality of models built for use on the fly.

Gangemi, Catenacci, Ciaramita, and Lehmann (2017) developed a meta-ontology called O^2 which identifies structural measures, functional measures, and usability-profiling measures, to evaluate annotation of ontology. The metaontology is then complemented with ontology validation called oQual, which provided ways to simulate the best set of criteria when choosing any Ontology over the other. Finally,

the paper concludes that parametric design of evaluation and validation (diagnostic) tasks are an integral part of building effective ontologies. The scope of our evaluation is strictly manual than the formalisms adopted in this paper but the point of emphasis is on evaluating the content of semantic web annotations.

4. Data Analysis

In order to build the three models briefly summarised above, an entry frame which contains the lexicon of the Yorùbá nouns according to Yorùbá Scholars were generated. The entry frame is structured according to the Qualia Structure (QS) of the Generative Lexicon. The frame contains 100 items and it serves as the backbone of the models developed. The items were evaluated in the project workshop as stated earlier and the results of the evaluation are discussed subsequently. Example of one of the items in Aina (2020) is repeated below as a glimpse of the entry frame.

(i) Àbá (n) attempt, proposal, suggestion
YORNO B90 - Ọ̀rọ̀ - orúkọ afòyemọ̀
YORNO A78 - Non-Human Nouns
YORNO A13 - Ọ̀rọ̀ - orúkọ ohun
WORDNET - Nouns cognition
CAT - POS noun
SEM - DISCRETE = true
SIZE = discrete
IND = x
QUALIA - object = false
FORMAL = cognition
CONST = part of (x, y)
TELIC = readjustment

The screenshot of the implementation page where the lexicon above occurs in Aina (2020) is shown below in Figure 1:

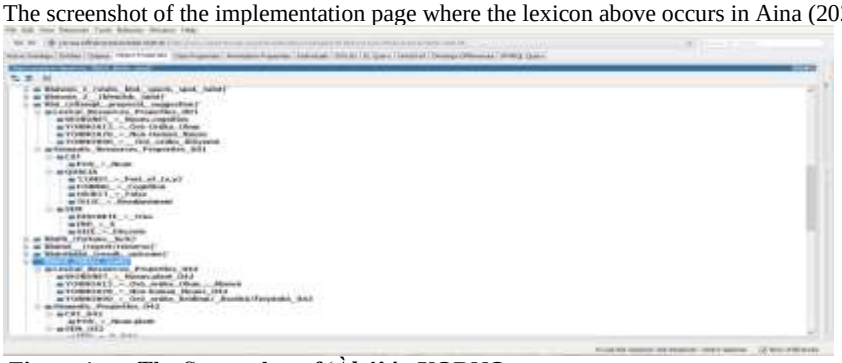


Figure 1. The Screenshot of 'Àbá' in YORNO

The source codes of the model were manually checked and evaluated for completeness, accuracy, orthographic error freeness, ambiguity freeness and semantic error freeness. The demographic information, content evaluation and perceptions of the professionals on the usefulness of the model were extracted through the use of questionnaires, and the results were statistically analysed. Twenty (20) questionnaires were distributed but Nineteen (19) were returned. The tables and charts below indicate the distribution pattern of the respondents in terms of age and educational qualifications:

Table 1: Age Distribution Table

	Frequency	Percent	Valid percent	Cumulative percent
26-30	4	21.1	21.1	21.1
31-35	5	26.3	26.3	47.4
36-40	5	26.3	26.3	73.7
41-45	1	5.3	5.3	78.9
46-50	2	10.5	10.5	89.5
51 and above	2	10.5	10.5	100.0
Total	19	100.0	100.0	

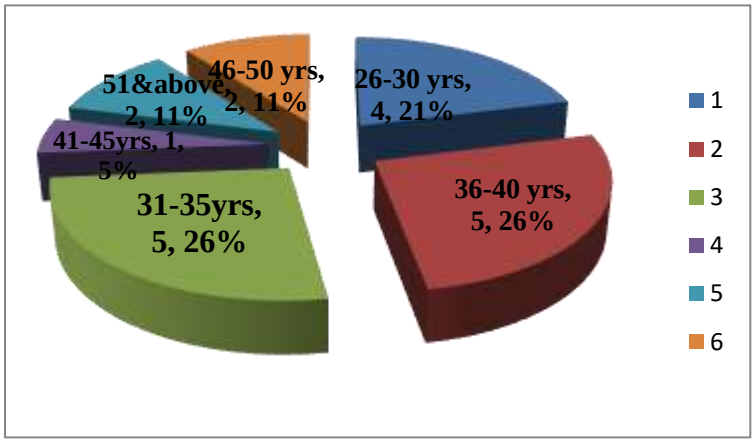


Figure 2. Age Distribution Chart

The age distribution of the evaluators reflects strength and effectiveness of the evaluation workshop held based on the following facts: (1) the larger percentages of the participants were between the ages of 31-40, namely 26.3% for ages between 31-35 and 26.3% were for the ages between 35-40 years. This is the age that rigorous academic prowess begins to grow. This reflects positively in the evaluation procedure. Also, the distributions of ages between age 46 and 50 were 10%, while 51 and above were also 10%. This also is a very good indicator because the wealthy experience of the participants, their rank in academics combined to make a good bear on the effectiveness of the evaluation.

5. Discussions

The different variables of content evaluation were extracted and discuss as follows:

5.1 Completeness

Checking and rating for completeness ensures that the entries are complete alongside with the English equivalent in terms of concepts they represent. The tables below reveal the plausibility of their responses to the number four questions under which stresses completeness and the ratings under completeness. The question statement is repeated here for emphasis: 'Can you say that the entries are complete in terms of concept they represent?' Their responses are analysed as follows:

Table 2: Question Statement (4) on Completeness

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	16	84.2	94.1	94.1
	No	1	5.3	5.9	100.0
	Total	17	89.5	100.0	
Missing	System	2	10.5		
Total		19	100.0		

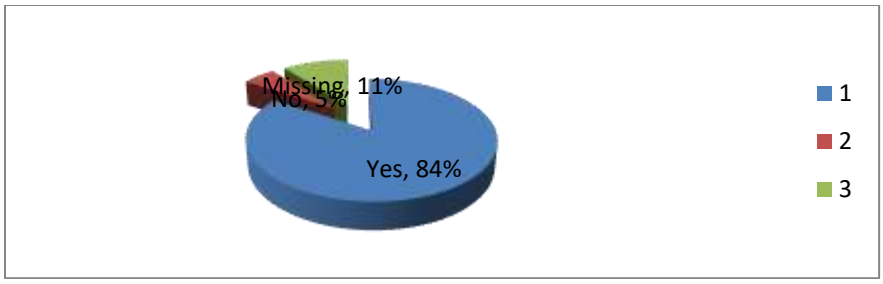


Figure 3. Chart Representing Question Statement (4) on Completeness

Table 3: Ratings of Question statement 4 on Completeness

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Good	6	31.6	31.6	31.6
	V. Good	13	68.4	68.4	100.0
	Total	19	100.0	100.0	

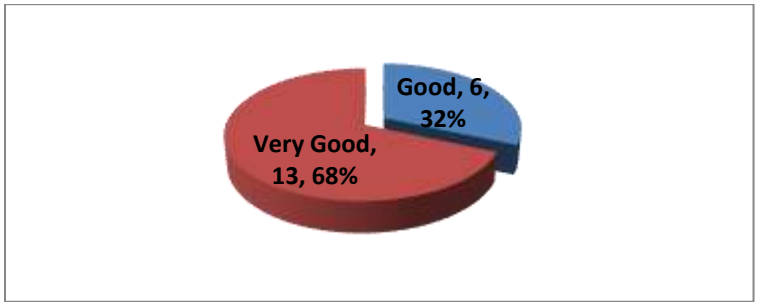


Figure 4. Chart Representing Rating of Question Statement 4 on completeness

Considering the responses, eighty four (84) which forms a highly larger percent agrees the entries are complete in terms of concept they represent. A good percentage also rated it very good compare with the whole statistics.

5.2 Accuracy

Checking and rating for accuracy ensures that the entries are semantically correct to a reasonable extent. The table below shows the responses to number four questions which stresses accuracy, and the questions for accuracy is repeated as follows: 'Can you confirm that the entries accurately represent the Yorùbá terms and concise too?'

Table 4: Question Statement (4) on Accuracy

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	15	78.9	93.8	93.8
	No	1	5.3	6.3	100.0
	Total	16	84.2	100.0	
Missing	System	3	15.8		
Total		19	100.0		

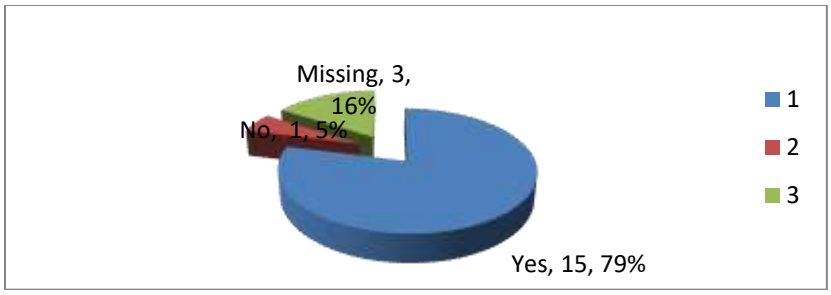


Figure 5. Chart Representing Question rating on Accuracy

Table 5: Ratings of Question statement 4 on Accuracy

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Good	6	31.6	33.3	33.3
	V good	12	63.2	66.7	100.0
	Total	18	94.7	100.0	
Missing	System	1	5.3		
Total		19	100.0		

Figure 6. Chart Representing Ratings of Question statement 4 on Accuracy

From the above, 15 (fifteen) responses which forms the larger percentage of 78.9 confirms that the entries are accurate and concise as to the terms they represent.

5.3 Semantic error freeness

Semantic error freeness checking and rating ensures that the items in the entry are semantically in tune with the Yorùbá semantic references and also close semantically to the English equivalents. It also measures the degree of its plausibility and conformity. The tables below show the responses to number four (4) question statement which confirms the semantic error freeness and ratings repeated here as 'Can you confirm that the entries are semantic error free to a good extent?'

Table 6: Semantic Error Freeness Question Statement

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	17	89.5	94.4	94.4
	No	1	5.3	5.6	100.0
	Total	18	94.7	100.0	
Missing	System	1	5.3		
Total		19	100.0		

Figure 7. Chart Representing Semantic Error Freeness Question Statement

Table 7. Semantic Error Freeness ratings

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fair	1	5.3	5.3	5.3
	Good	4	21.1	21.1	26.3
	V good	13	68.4	68.4	94.7
	Excellent	1	5.3	5.3	100.0
	Total	19	100.0	100.0	

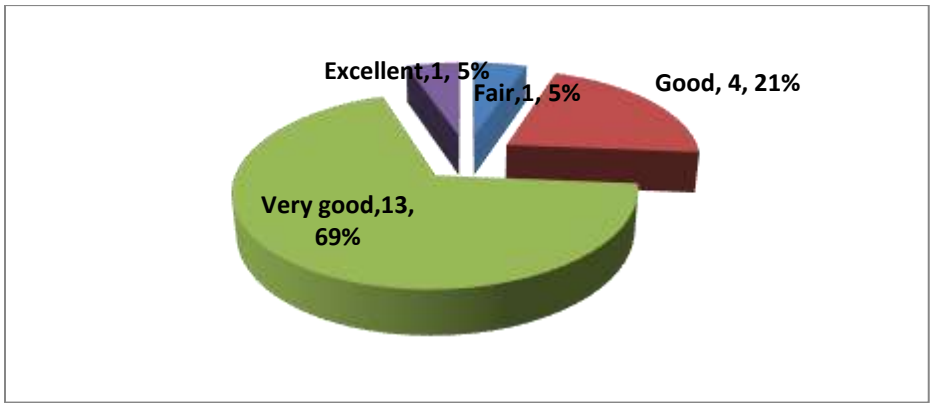


Figure 8. Chart Representing Semantic Error Freeness ratings

The statistic output above reflects that the entries are semantically in tune with the Yorùbá semantic references and are also perceived close semantically to the English equivalents. Out of the 19 responses, only one rated it excellent, while 13 which is 69% rated it very good, 4 which is 21% rated it good and only one rated the semantic error freeness fair

5.4 Orthography error freeness

Orthography error freeness checking and rating ensures that all the items in the entry are both English and Yorùbá orthographically correct. It also measures the degree of its plausibility. The table below shows the responses to number 4 question statements which emphasize the orthographical freeness and ratings of the model contents under orthography error freeness:

Table 8: Orthographic Error freeness Question statement

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	17	89.5	89.5	89.5
	No	2	10.5	10.5	100.0
	Total	19	100.0	100.0	

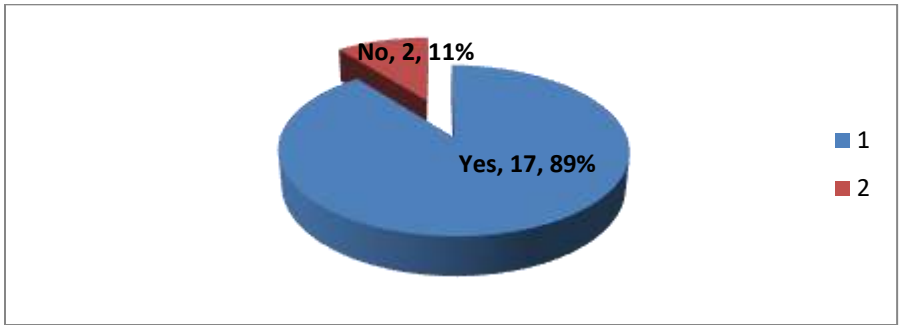


Figure 9. Chart Representing Orthographic Error freeness Question statement

Table 9. Orthographic Error freeness Ratings

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fair	1	5.3	5.3	5.3
	Good	5	26.3	26.3	31.6
	Vgood	13	68.4	68.4	100.0
	Total	19	100.0	100.0	

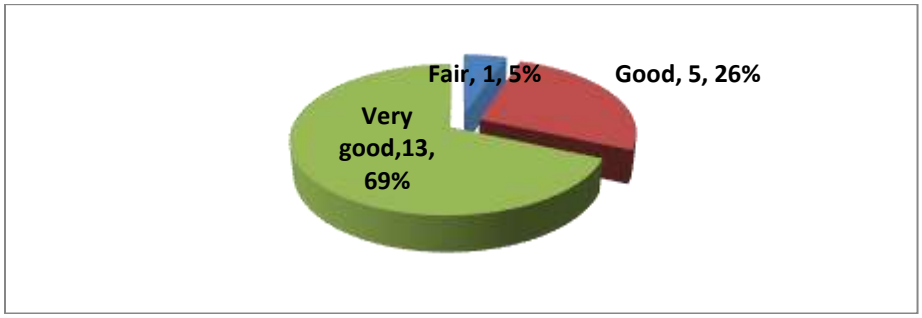


Figure 10. Chart Representing Orthographic Error freeness Ratings

Evaluating the orthographic error freeness, 13,5and 1were recorded for frequency, 68.4, 26.3 and 5.3 percent were recorded for very good, good and fair respectively. This shows that the rating were of pass mark.

5.5 Ambiguity freeness

Checking and rating for ambiguity freeness ensures that the entries are not ambiguous with the concepts they represent. The tables below reveal the plausibility of their responses to the number four questions which stresses ambiguity freeness and the ratings:

Table 10: Ambiguity freeness Question statement

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	18	94.7	94.7	94.7
	No	1	5.3	5.3	100.0
	Total	19	100.0	100.0	

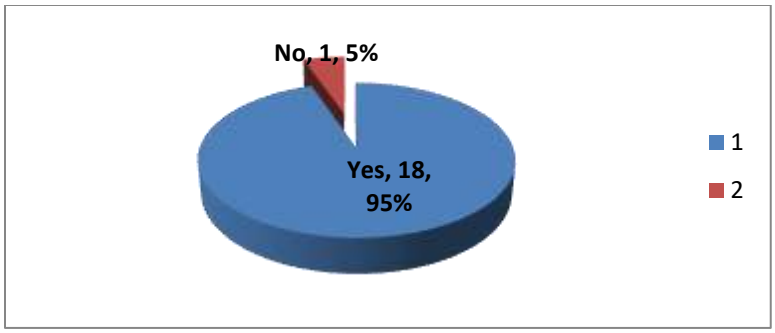


Figure 11. Chart Representing Ambiguity freeness Question statement

Table 11: Ambiguity Freeness Ratings

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Good	5	26.3	26.3	26.3
	V good	13	68.4	68.4	94.7
	Excellent	1	5.3	5.3	100.0
	Total	19	100.0	100.0	

Figure 12. Chart Representing Ambiguity Freeness ratings

Considering rating of orthographic error freeness, 13, 5 and 1 were recorded for frequency. This is of approximately 69, 26 and 5 percent recorded for very good, good and fair respectively. The majority of the rating confirms that the entries are not ambiguous withe concepts they represent.

5.6 Experts Perception on the Usefulness of the Model

The eight (8) questions discussed in this section measure the experts' perception on the usefulness of the model. Question statements 3, 4 and 7 were targeted to gauge specifically the respondent's perception on the usefulness of the model. The statistics of their responses are shown in the table below:

Table 12: Response to Perception Question Statement (3)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	10	52.6	52.6	52.6
	Strongly Agree	9	47.4	47.4	100.0
	Total	19	100.0	100.0	

Table 13: Response to Perception Question Statement (4)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	1	5.3	5.3	5.3
	Agree	8	42.1	42.1	47.4
	Strongly Agree	10	52.6	52.6	100.0
	Total	19	100.0	100.0	

Table 14: Response to Perception Question Statement (7)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Disagree	1	5.3	5.3	5.3
	Agree	7	36.8	36.8	42.1
	Strongly Agree	11	57.9	57.9	100.0
	Total	19	100.0	100.0	

5.7 Evaluation with the aid of Frequency Table

Frequency table is one of the common approaches in human evaluation. The frequency table is based on frequency, percentage, valid percentage and cumulative percentage. Five (5) questions were set for each of the five indicators i.e completeness, accuracy, orthographic error freeness, ambiguity freeness and semantic error freeness.
The highest score of respondents to each of the five (5) content indicators was 8. The respondents that score less than 4 were categorized as having poor perception in the content indicators i.e. Completeness, Accuracy, Semantic error freeness, Orthographic error freeness and Ambiguity freeness. The Respondents that scored between 4 and 5 were categorized as having average in all the five content indicators. The Respondents that scored between 6 and 8 were categorized as having high accuracy in all the five content indicators. The results were then used to form the frequency tables and the chart accordingly as shown below.

Table 15: The frequency table for Completeness

	Frequency	Percent
Average Completeness Score	4	21.1
High Completeness Score	15	78.9
Total	19	100.0

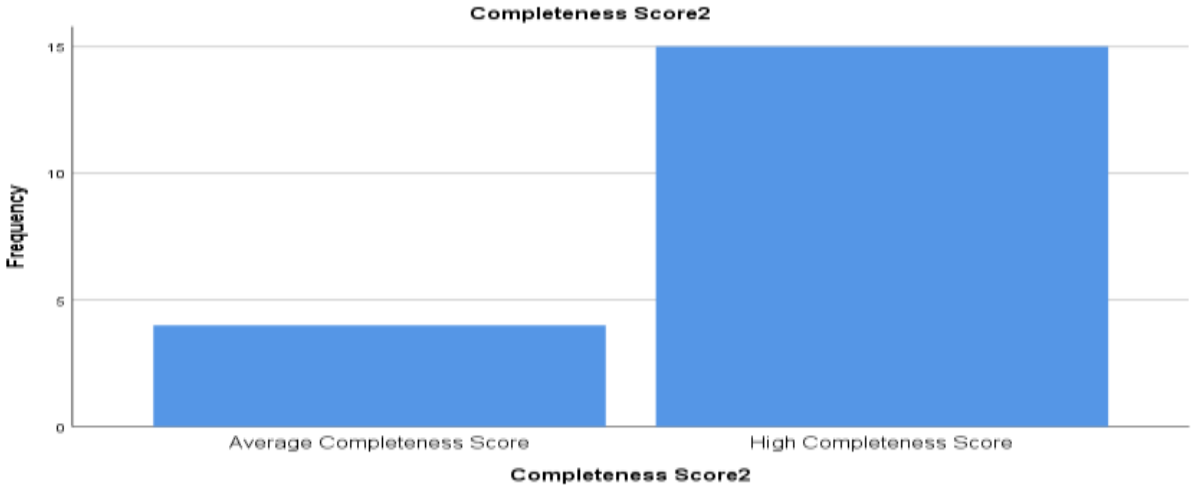


Figure 13. Chart Representing Experts perception as to Completeness

Table 16: The frequency table for Accuracy

	Frequency	Percent
5Average Accuracy Score	7	36.8
High Accuracy Score	12	63.2
Total	19	100.0

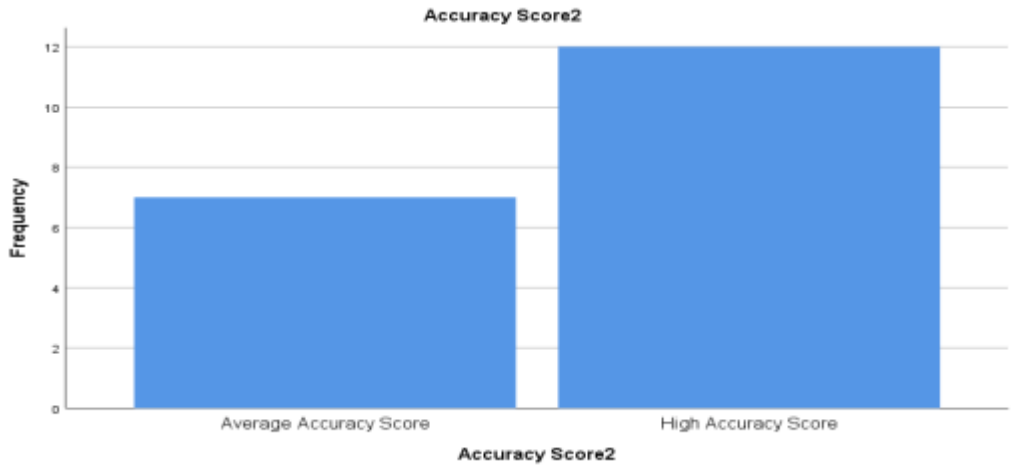


Figure 14. The frequency Chart for Accuracy Score

Table 17: The Frequency Table for Semantic error freeness

		Frequency	Percent
	Average Semantic error freeness Score	2	10.5
	High Semantic error freeness Score	17	89.5
	Total	19	100.0

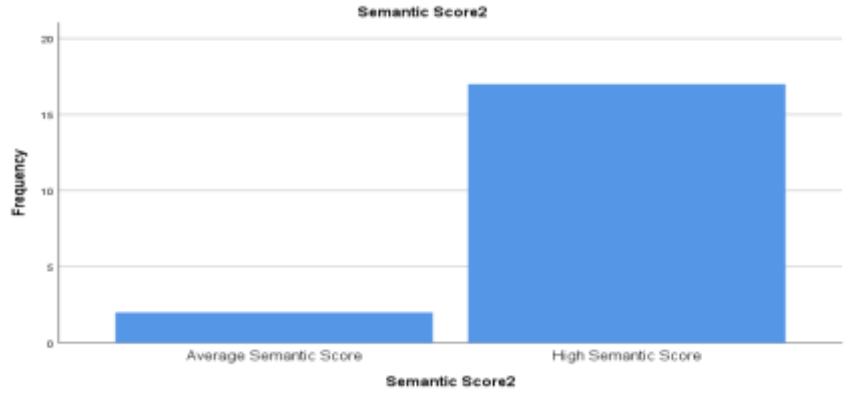


Figure 15. The frequency Chart for Semantic error freeness

Table 18: The frequency table for Orthographic error freeness

		Frequency	Percent
	Average Orthographic error freeness Score	2	10.5
	High Orthographic error freeness Score	17	89.5
	Total	19	100.0

Figure 16. Chart Representing frequency table for Orthographic error freeness

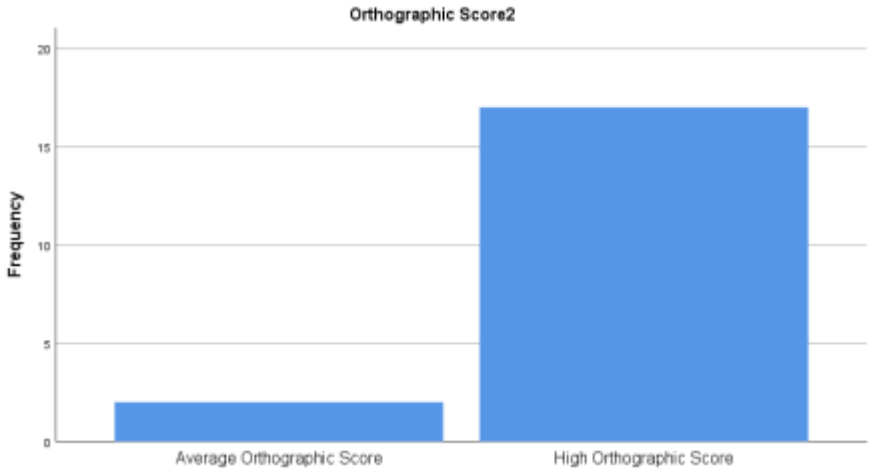


Table 19: The frequency table for Ambiguity freeness

		Frequency	Percent
	Average Ambiguity freeness Score	1	5.3
	High Ambiguity freeness Score	18	94.7
	Total	19	100.0

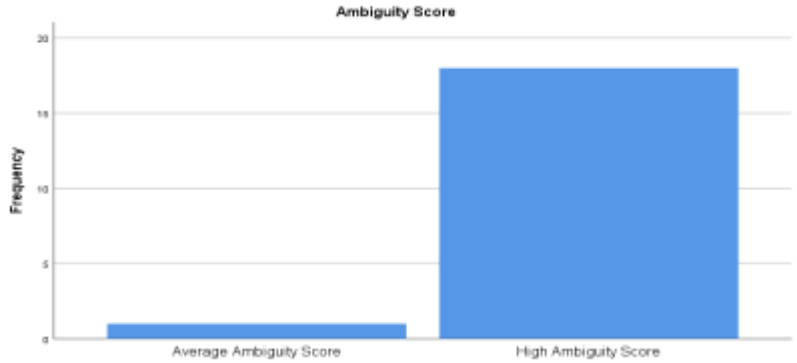


Figure 17. Chart Representing frequency table for Ambiguity error freeness

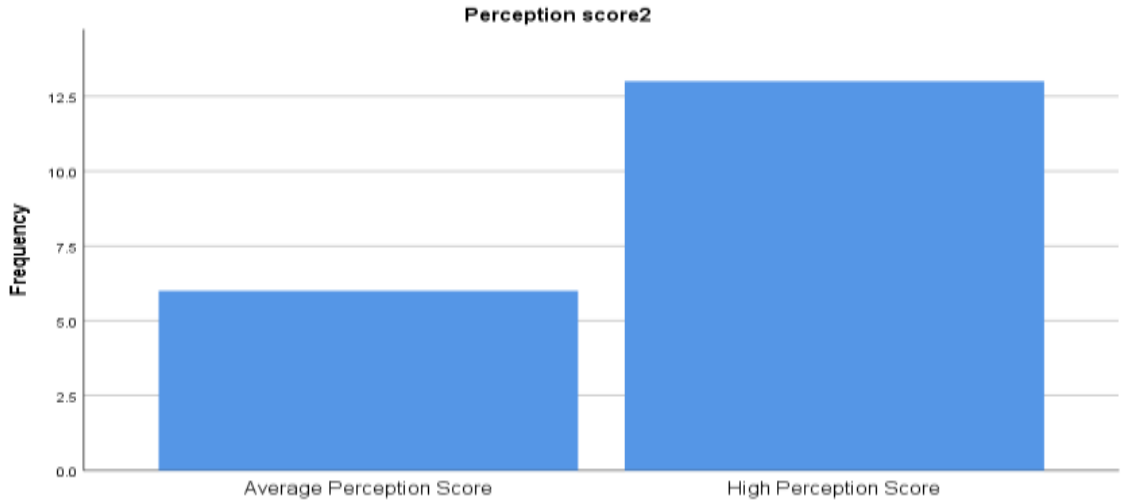
5.8 Experts' Perception of the Usefulness of Ontology Annotations

Eight questions were designed to determine the degree of the respondent's perception about the usefulness of ontology annotations to NLP and AI systems and the contributions it could make to Yorùbá language. Question statements 3, 4 and 7 were specifically targeted to gauge the experts' perception of the model. Using the T- test analysis, the highest perception score of the respondents was 32. Respondents that scored less than 16 were categorized as having poor perception of the software. Respondents that scored between 16 and 24 are categorized as having average perception while those who scored more than 24 were categorized as having high perception. Table 20 shows the statistics of their responses.

Table 20. Perception of the model

	Frequency	**Percent**
Average Perception Score	6	31.6
High Perception Score	13	68.4
Total	19	100.0

Figure 18. Chart Representing Perception of the model



5.9 T-test

Carrying out a T-test in data analysis involves the statistical examination of two population means. It examines whether two samples are different and commonly used when the variances of two normal distributions are unknown and when an experiment uses a small sample size. The T-test was used for the five content evaluation indicators and the experts' perception sections of the questionnaire and the following results were produced.

5.10 Completeness

One-sample t- test was carried out using 4 (half of 8) as test score. This was to test the following hypothesis;

H_0: There is no statistically significant difference between the mean completeness score and the test score, 4
H_1: There is a statistically significant difference between the mean completeness score and the test score, 4
The t-test gave a p value of <0.0001 (α=0.05). Since p is <0.05, the null hypothesis was rejected and the alternative hypothesis was accepted. The conclusion was that there was a statistically significant difference between the mean completeness score and the test score, 4. Table (15) above, shows that there is a significant difference, none of the respondents had poor evaluation of the concept 'completeness' as the content indicator of annotation made.

5.11 Accuracy

One-sample t- test was carried out using 4 (half of 8 which is the highest score for content evaluation question statements) as test score. This was to test the following hypothesis;
H_0: There is no statistically significant difference between the mean accuracy score and the test score, 4
H_1: There is a statistically significant difference between the mean accuracy score and the test score, 4
The t-test gave a p value of <0.0001 (α=0.05). Since p is <0.05, the null hypothesis was rejected and the alternative hypothesis was accepted. The conclusion was that there was a statistically significant difference between the mean accuracy score and the test score, 4. Table (16) above, shows that there is a significant difference, none of the respondents had poor evaluation of the concept 'accuracy' as the content indicator of annotation made.

5.12 Semantic Error Freeness

One-sample t- test was carried out using 4 (half of 8) as test score. This was to test the following hypothesis;
H_0: There is no statistically significant difference between the mean semantic error freeness score and the test score, 4
H_1: There is a statistically significant difference between the mean semantic error freeness score and the test score, 4

The t-test gave a p value of <0.0001 (α=0.05). Since p is <0.05, the null hypothesis was rejected and the alternative hypothesis was accepted. The conclusion was that there was a statistically significant difference between the mean semantic error freeness score and the test score, 4. Table (17) above, shows that there is a significant difference, none of the respondents had poor evaluation of the concept

'semantic error freeness' as the content indicator of annotation mad

5.13 Orthography Error Freeness

One-sample t- test was carried out using 4 (half of 8) as test score. This was to test the following hypothesis:

H_0: There is no statistically significant difference between the mean orthography error freeness score and the test score, 4
H_1: There is a statistically significant difference between the mean orthography error freeness score and the test score, 4

The t-test gave a p value of <0.0001 (α=0.05). Since p is <0.05, the null hypothesis was rejected and the alternative hypothesis was accepted. The conclusion was that there was a statistically significant difference between the mean orthography error freeness score and the test score, 4. Table (18) above, shows that there is a significant difference, none of the respondents had poor evaluation of the concept 'orthography error freeness' as the content indicator of annotation made.

5.14 Ambiguity freeness

One-sample t- test was carried out using 4 (half of 8) as test score. This was to test the following hypothesis:

H_0: There is no statistically significant difference between the mean ambiguity freeness score and the test score, 4
H_1: There is a statistically significant difference between the mean ambiguity freeness score and the test score, 4
The t-test gave a p value of <0.0001 (α=0.05). Since p is <0.05, the null hypothesis was rejected and the alternative hypothesis was accepted. The conclusion was that there was a statistically significant difference between the mean ambiguity freeness score and the test score, 4. Table (19) above, shows that there is a significant difference, none of the respondents had poor evaluation of the concept 'ambiguity freeness' as the content indicator of annotation made.

5.15 Perception

One-sample t- test was carried out using 16 (half of 32) as test score. This was to test the following hypothesis;

H_0: There is no statistically significant difference between the mean perception score and the test score, 16
H_1: There is a statistically significant difference between the mean perception score and the test score, 16

The t-test gave a p value of <0.0001 (α=0.05). Since p is <0.05, the null hypothesis was rejected and the alternative hypothesis was accepted. The conclusion was that there is a significant difference between the mean perception score and the test score, 16. As shown in Tables 15 -19 above. The significant difference in the mean score occurs because none of the respondents has poor perception of the software.

6. Conclusion

The clarion calls by the earlier Scholars to shift the focus of language research to the deployment of modern innovative technology for human language analysis is a good development. The emphasis is not only enough and complete without thinking of evaluation readings of the developed systems and models. Most of the editors available for semantic web annotations already have their in-built automatic evaluation tools which are not able to capture human intuitive language knowledge. This reason should force human evaluation activities to complementing automatic in-built assessment technology. The hypotheses examined in the T-test data analysis used for the five content evaluation indicators and experts' perception of contents of the annotations done for Yoruba noun Ontology gave out findings that reflects a necessity. The results show that there was a statistically significant difference between the mean score and the test score for the five variables of completeness, accuracy, semantic error freeness, orthographic error freeness and ambiguity freeness. This indicates that none of the experts had poor perception and evaluation of the ontology. Different human evaluation techniques are therefore recommended as a necessity when thinking of semantic web annotation.

References

Aina, A. A. 2019a. A Hybrid Yorùbá Noun Ontology. *KIU Journal of Humanities*. 4.4: 229-246. College of Humanities and Social Sciences, Kampala International University, Uganda.

…….. 2019b. Development and Preservation of Yorùbá Cultural Identity: An Ontology–Based Approach. *Ọ̀pánbàtà- Jọ́nà ìmọ̀ Áfíríka: LASU Journal of African Studies,*

7:19-29. Department of African Languages, Literatures and Communication Arts, Lagos State University, Ọjọ́, Lagos.

…….. 2020. An Ontology – Driven Annotation of Aspect of Yorùbá Noun. PhD Thesis submitted to the Department of Linguistics and African Languages, University of Ibadan, Ibadan, Nigeria.

Aina, A.A. and Taiwo P. O. 2019. Yorùbá Noun Ontology from Functional Perspectives. *Journal of the Linguistics Association of Nigeria*, 22.1: 189-21

Linguistic Association of Nigeria. www.jolan.org.ng/access/downloader.php. PDF.

Adekoya, F.A. 2010. An evaluation framework for benchmarking the internal graph structure of Ontology. PhD.Thesis. Dept. of Computer Science. Federal University of Agriculture, Abeokuta, Ogun State, Nigeria. xvi+171pp.

Andre, E and Rist, T. 1993. The design of illustrated documents as a Planning Task in M. T. *Intelligent Multimedia Interfaces*. Maybury ed. The MIT Press, Menlo Park (CA) Cambridge London (England), 94-116.

Awobuluyi, O.1978. *Essentials of Yorùbá Grammar.* University Press Ibadan.

---------- 2008. *Èkọ́ Ìṣẹ́dá Ọ̀rọ̀ Yorùbá.* Akure: Montem Paperbacks.

---------- 2013. *Èkọ́ Gírámà Yorùbá.* Òṣogbo. Atman Ltd.

Bamgbose, A. 1990. *Fonoloji ati Girama Yorùbá*: Ibadan: UPL

Borst, W. N. 1997. Construction of Engineering Ontologies. PhD thesis. University of Twente. Enschede. xvi+266

Gangemi, A., Catenacci, C., Ciaramita, M. and Lehmann, J. 2006 Modelling Ontology Evaluation and Validation. Proceedings of the Third European Semantic Web Conference. Sure Y. eds. *Springer*. London, U.K.

Gruber, T. R. 1993. A Translation Approach to portable Ontology specification. *Knowledge Acquisition* 5.2: 199-220.

Guarino, N., Oberle, D. and Staab, S. 2009. What is Ontology? *Formal Ontology in Information Systems* Guarino N. ed. Proceedings of FOIS'98, Trento, Italy, June 6-8, 1998, pages 3–15. IOS Press, Amsterdam.

Hutchins. 1994. Research Methods and System Designs in Machine Translation: A Ten-Year Review, 1984-1994. A Paper presented at an International Conference Machine Translation: Ten Year on. Cranfield University, England, 12-14 November 1994. *Conference Proceedings.* Clarke, D. and Vella A. Cranfield University Press, 1998. Paper No.4. Retrieved online 12 July 2016 from http://hutchinsweb.me.uk/Cranfield-1994.pdf

_____. 1995. Machine Translation: A Brief History. In *Concise History of the Language Sciences: From the Sumerians to the cognitivists*, Koerner, K. and Asher, E. eds. Oxford: Pergamon Press. 431-445.

_____ 2007. Machine Ttranslation: A Concise History. In *Computer Aided Translation: Theory and Practice*, C. S. Wai eds. Chinese University of Hong Kong

Marakas, G. M. 1999. *Decision support systems in the twenty-first century*. Upper Saddle River, N.J., Prentice Hall

Power, D. J. 2002. *Decision Support Systems: Concepts and Resources for Managers*. Westport, Conn., Quorum Books.

Odoje, C. O. 2016. Human Evaluation of Yorùbá-English Google Translation. *Journal of West African Languages*. 43.1: 79-92.

Pareja – Lora, A. 2012. OntoTag: A Linguistic and Ontological Annotation Model Suitable for the Semantic Web. A Ph.D Thesis. Facultad De Informática Universidad Politécnica De Madrid. viii+365. Retrieved online on Aug. 28, 2017 from https:// www.researchgate.net.publication.

Pustejovsky, J. 1991. The Generative Lexicon. *Computational Linguistics*, 17. 4 Rector, A., Drummond N., Horridge, M., Rogers J., Knublauch, H., Stevens, R., Wang H., and Wroe, C. 2004. OWL Pizzas: Practical Experience of Teaching OWL-DL: Common Errors & Common Patterns. 14th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2004). 63-81.

Rayson, Uchechukwu and Hepple (2020). Igbo- English Machine Translation: An Evaluation Benchmark. Published as a conference paper at ICLR 2020 Studer, Benjamins and Fensel (1998:4)

Ikechukwu, Onyenwe, Chinedu Uchechukwu, and Mark Hepple (2014). Part-of-speech tagset and corpus development for Igbo, an African language. In Proceedings of LAW VIII-The 8th Linguistic Annotation Workshop, pp. 93–98, 2014.

Ruskova, N. A. 2002. Decision Support System for Human Resources Appraisal and Selection, *Intelligent Systems 2002*, Vol. 1 IEEE, Varna, Bulgaria. 354-357.

Zauzich, K. T. 1992. *Hieroglyphs Without Mystery: An Introduction to Ancient Egyptian Writing.* University of Texas Press.

APPENDIX

SECTION A DEMOGRAPHIC INFORMATION

(1) Sex: Male ( ) Female ( )

(2) Marital Status: Single ( ) Married ( ) Divorced ( ) Widowed ( )

(3) Age group: 21 – 25 ( ) 26 – 30 ( ) 31 – 35 ( ) 36 – 40 ( ) 41 – 45 ( )

46 – 50 ( ) 51 and above ( )

(4) Highest Educational Qualification:

(a) 'O' Level, GCE/WAEC ( )

(b) 'A' Level, OND/NCE ( )

(c) B.Sc. B.A, HND ( )

(d) M.Sc. M.Ed. ( )
(e) Ph.D ( )

(f) Areas of Specialization: Linguistics/Yorùbá language: Phonetics/Phonology ( )
Morphology ( ) Syntax ( ) Applied Linguistics
Literature: Prose ( ) Poetry ( ) Drama ( ) others: Please specify ( )

(e) Other professional qualifications ( ) Specify _____

(5) Working Experience: (in years) …….

SECTION B

CONTENT EVALUATION QUESTIONNAIRE

The following statements are designed to determine the degree at which you evaluate the contents of the model as to completeness, accuracy, semantic error freeness, orthographic error freeness, and ambiguity error freeness.
Please respond to the following items by ticking the response that most adequately fits in.

Completeness

No.	Statements	Yes	No	Remarks
1	Did you manually check the entries line by line?			
2	Did you observe any item that is not complete in any form? Please state the entry in the 'remarks' column			
3	Do you agree with the English equivalent of the entries? If your answer is no, state the entry and your suggestion in the 'remarks' column			
4	Can you say that the entries are complete in terms of concept they represent?			
5	How do you rate the completeness of the entry? Please enter your rating in the 'remarks' column (1) Poor (2) Fair (3) Good (4) Very Good (5) Excellent			

Accuracy

No.	Statements	Yes	No	Remarks
1	Are you highly competent in Yorùbá language?			
2	Did you observe any item that is not accurate in any form? Please state the entry in the 'remarks' column			
3	Do you agree that the English equivalent of the entries are accurate? If your answer is no, state the entry you are not comfortable with and your suggestions in the 'remarks' column			
4	Can you confirm that the entries accurately represent the Yorùbá terms and concise too?			
5	How do you rate the accuracy of the entry? Please enter your rating in the 'remarks' column (1) Poor (2) Fair (3) Good (4) Very Good (5) Excellent			

Semantic error freeness

No.	Statements	Yes	No	Remarks
1	Are you competent in Yorùbá language sematic references?			
2	Did you observe any item that contains semantic error in any form? Please state the entry in the 'remarks' column			
3	Do you agree that semantic references of the English equivalents of the entries are close to the Yorùbá language? If your answer is no, state the entry you are not comfortable with and your suggestions in the 'remarks' column			
4	Can you confirm that the entries are semantic error free to an extent?			
5	How do you rate the semantic error freeness of the entries? Please enter your rating in the 'remarks' column (1) Poor (2) Fair (3) Good (4) Very Good (5) Excellent			

Orthographic error freeness

No.	Statements	Yes	No	Remarks
1	Are you competent in current Yorùbá language orthography and spellings?			
2	Did you observe any item that does not conform to the modern Yorùbá orthography and spellings? Please state the entry in the 'remarks' column			
3	Do you agree that the English equivalents of the entries are also orthographically correct? If your answer is no, please state the entry you are not comfortable with and your suggestions in the 'remarks' column			
4	Can you confirm that the entries are orthographic error free to an extent?			
5	How do you rate the orthographic error freeness of the entries? Please enter your rating in the 'remarks' column			

(1) Poor (2) Fair (3) Good (4) Very Good (5) Excellent

Ambiguity freeness

No.	Statements	Yes	No	Remarks
1	Are you much familiar with ambiguity and ambiguous terms in Yorùbá language?			
2	Did you observe any item that is ambiguous in meaning? Please state the entry in the 'remarks' column			
3	Do you agree that the English equivalents of the entries and the Yorùbá equivalents are not ambiguous with the concepts they represent? If your answer is no, please state the entry you are not comfortable with and your suggestions in the 'remarks' column			
4	Can you confirm that the entries are free from ambiguity to a fairly large extent?			
5	How do you rate the ambiguity freeness of the entries? Please enter your rating in the 'remarks' column (1) Poor (2) Fair (3) Good (4) Very Good (5) Excellent			

SECTION C

EVALUATION OF EXPERTS PERCEPTION ABOUT THE USEFULNESS OF ONTOLOGY ANNOTATIONS

The following statements are designed to determine the degree of your perception about the usefulness of ontology annotations to natural language processing (NLP) and Artificial Intelligence (AI) Systems for Yorùbá language. The contributions it could make to the development of Yorùbá language as a whole.

Please respond to the following items by ticking the response that adequately fits your thoughts.

Strongly Agree (SA) = 4 points, Agree (A) = 3 points, Disagree (D) = 2 points, Strongly Disagree (SD) = 1

S/N	Statements	Strongly Disagree	Disagree	Agree	Strongly Agree
1	I have come to understand what ontology annotation does in the course of this workshop.				
2	I am aware it has been in existence in other languages of the world.				
3	Ontology models in Yorùbá will enable the existing NLP and AI systems.				
4	YORNO will foster developments of more applications that exploits Yorùbá grammar				
5	YORNO can serve as repository of data and can solve ambiguity problems in Yorùbá language.				
6	YORNO provides opportunity for software developers to explore the grammar of their choice since it is a hybrid model of two scholars while re-inventing the wheel.				
7	The model can contribute to the development of Yorùbá language.				
8	There is need for improvement and motivations for more research in this direction.				