Journal of West African Languages

Volume 48.1 (2021)

ONTOLOGY ANNOTATION FOR NATURAL LANGUAGE DEVELOPMENT: A YORÙBÁ NOUN PRELIMINARY MODEL

Akindele A. Aina
Olabisi Onabanjo University
larexacademie@yahoo.com
akindeleaina@oouagoiwoye.edu.ng

and

Oyè P. Táíwò
University of Ibadan
oyepaultaiwo@gmail.com

Abstract
Ontological annotation, a machine-readable and explicit description of shared concepts in a domain of discourse has always been employed in European languages to aid computer in Natural Language Processing (NLP). This method, however, has not been so much employed in African languages including the Yorùbá language. This paper, therefore, proposes ontological annotation as a way to prepare the implicit knowledge of Yorùbá core grammar for (NLP) activities by isolating and tagging some Yorùbá nouns into their component features, such that the model implemented is both human and machine readable. Relational Content Analysis (RCA) as a method of ontological annotations was adopted to design a conceptual sample model for selected Yorùbá nouns. The informally perceived domain knowledge of Yorùbá nouns were extracted randomly, using intermediate representations based on tabular and entity-relation notations. Protégé 4.5, a semantic web editing tool was used to implement a sample model for Yorùbá nouns according to Bamgbose (1990)'s "*Fonọ́lọ́jì àti Gírámà Yorùbá"*. The model named YORNOB (Yorùbá Noun Ontology according to Bamgbose), at the final edge, provides the semantic load and properties of each of the nouns in it. The definitions in the annotation serve as backbones for machine learning, web searching and artificial intelligence agents and other (NLP) systems. The model developed in this paper serve as repositories of data and shared terminology for Yorùbá language development and use. This paper recommends that more on ontological annotations for Yorùbá grammar concepts are needed to foster Yorùbá language engineering.

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

1. Introduction:

Research activities in Yorùbá studies are gradually shifting towards the employment of Information and Communication Technology (ICT) in processing natural language corpus. Many models which have been used for the developed European languages are being adopted for developing African languages including the Yorùbá language. The trend of using machine to support or replace humans has helped to increase efficiency of professionals and achieved higher performance in all human endeavours. As lucrative and interesting as this development is, Odoje (2010:6, 2017:2) establishes the fact that African languages are still resource scarce languages and are yet to grow in tune with the modern trend. It is a fact that indigenous languages are losing their communicative status, functions and roles in the society. Adégbọ́lá (2006:7) emphasises this fact when he submits that:

> . . . many African languages still remain unwritten and many of those that are written are not available to modern information technologies… The need to appropriate ICT's for the language needs for Africa is pressing.

From this position, it is clear that there is the need for increased efforts towards making African languages compliant with world scientific developments.

Applying ICT to African languages as stated above is constrained by so many inhibiting factors. On the first note, development in the scientific world brought about ICT. Suffice it to say that the development of different applications to process human language is a product of research and an improvement over the established ones. The history of computer as a machine for exploring human activities indicated that it was deployed in developmental stages and eras of first, second, third and fourth generations, with improved Internet activities. It follows then

that employing this tool to natural languages is a subject of development as well. Further researches in natural languages using computational tools, are still much needed to accommodate the inventions and influences of computers on natural language.

Secondly, the level of convergence and globalisation phenomenon improves with modernity. Convergence is a product of digitisation of human concept and natural realities and its representation within the various levels of digital codes. Globalisation involves technological changes that are influencing the way humans interact. This is the point which Adegbola (2006:3) expresses by stating that "Globalization involves the creation of new and the multiplication of existing social networks, and activities that increasingly overcome traditional, political, economic, cultural and geographical boundaries". Meeting up with the developmental trend, therefore, poses a challenge to our indigenous languages one way or another. Poor infrastructure, limited or no scientific background institutional training and attitude of government/individuals in terms of their response towards digitised applications are additional inhibiting constraints preventing the employment of ontological annotation in African languages in recent times.

2. Statement of the Problem

We have highlighted previously, the background realties that call for processing natural languages with computers as much as the few factors inhibiting the activities. There is an important issue which is based on the fact that the core grammar of natural languages is not structured in the format which makes it sharable and accessible to machines for Natural Language Processing (NLP) activities. The core job of ontological annotation addresses this need. The Yorùbá noun and its phrases have been thoroughly analysed by different Yorùbá scholars. Johnson (1921), Bamgbose (1966, 1990), Awóyalé (1974), Awobuluyi (1978, 2013), and Owolabi (1976). Different criteria, such as functions, dialectal rendition, syntactic structures, derivation strategies and so on, are of concerns in their analysis. But the gap is that; there has been no discussion from the point of semantic web engineering methods which favours and allows machine NLP activities. The differences in theoretical viewpoints from Yorùbá scholars as contained in extant literature are still in implicit forms, thereby bringing about difficulty in presenting the Yorùbá noun as a concept for coding in machine formal language. There is the need to analyse the conceptual rudiments of the items in a way that both man and machine can make use of the conceptual knowledge.

Considering this need, some of the questions which evolved in discussing this preliminary model for handling Yorùbá nouns are: What are some attributing features which Bamgbose (1990) and other scholars identified for Yorùbá nouns? How are these features annotated or defined in a way that makes it machine readable? What tools are available for the annotation task? Lastly, what other applications can utilise the ontological annotations?

3. Aim and Objectives of the Work

The background need to the study and the problem associated as reflected in the previous questions inform the aim that ontological annotation is a way to natural language development in African languages and this gives the impetus to the objectives of extracting knowledge of Yorùbá nouns according to Bamgbose (1990) as a preliminary model to explain and demonstrate how ontological annotations can foster machine learning activities in African languages. It is also our objective to show how the conceptual knowledge extracted from the prose narratives will be processed and annotated for machine readability which other NLP agents can access and used for effective performance of their applications.

4. Research Methodology

To achieve the aim and objectives of this research, the prose narratives of Bamgbose (1990) on Yorùbá nouns were rigorously examined, the prototype of the knowledge in the work was extracted. Using Methontology, the extraction was represented in the intermediate tabular representation, and designed into concepts relations and taxonomies. The resulting design were implemented further, using Protégé 4.0 - a semantic web editing tool which projects and makes

the formalisms accessible to the semantic web. A brief demonstration of how a hypothetic Machine Translation (MT) application can utilise the annotation is provided.

5. Ontological Annotation

Ontology is a term in Philosophy which refers to "the science of being *qua* being" Guarino, Oberle and Staab (2009:10). The effort carried out to study the entities existing in any domain means that Ontology is already being developed for that domain. The study of the natural existence of a thing as it relates to the universe and other entities makes that thing Ontology-based. Ontology is extended to semantics when some of its tasks involve explicitness of a concept as it relates to other elements in the structure of utterances. Annotation involves labelling short comments to a book or piece of writing in order to explain parts of it. In essence, annotation involves adding comments, notes or explanations (usually short and coded) to a lexical item, text or a book. According to Pareja-Lora (2012: 32-34), it dates back to the ancient Egyptian writings (the hieroglyphic periods), when hieroglyphs were attached to describe the sound of a word, or add comments or explanations about the coded word. When a word is polysemous i.e. have several meanings, the ancient Egyptian writing uses sequence of hieroglyphs to disambiguate it. Pareja-Lora (2012) expresses the basic terminology concepts involved in annotations and stated in essence that the primary purpose of annotation is to make explicit, the contents of concepts in a domain using a set of arbitrary tags or labels which must conform to an agreed standard. He outlined two approaches to annotation: Linguistic Approach and Computational Approach.

On the one hand, the Linguistic approach considers five main levels of analysing a text or corpus, namely the morphology, syntax, phonology, phonetics and semantics. Only three of these linguistic levels of analysis have been fully developed to meet with the current ICT methods of annotating text or corpus. These are morphosyntactic, syntactic and semantic annotations. Lemma tagging, annotated as a model known as Lemmatiser is an example of morphosyntactic annotation. Part-of-speech tagging is also done by part-of-speech taggers (POS taggers). The annotation done in McEnery &Wilson (2001) is a good example of syntactic annotation. Computational approach, on the other hand, concerns itself with the situation whereby the target user of an annotation is a machine instead of human. Before a machine can understand and process annotations made by human, the annotations must be encoded in the format, clear vocabularies, syntax, schemes and tag forms which will make it readable for the machine. Exchanging annotation information requires encoding the text in a common, uniform and standard format. Consequently, a collection of annotation vocabularies and schemes with formal properties that are needed and devised for use in encoding texts and documents for the World Wide Web are known as Annotation Languages or Markup languages. Some of the examples of these Markup languages include: Generalised Markup Languages (GML), Standard Generalised Markup Languages (SGML), eXtensible Markup Language (XML), HyperText Markup Language (HTML) and Resource Description Format (RDF).

Furthermore, various schemes of parsing are also being employed by different annotators. McEnergy & Wilson (2001) point out that the differences in these schemes are expressed in:

a. The number of constituent types they employ, typically the number of tags in part-of-speech (POS) tagset,
b. The way in which constituents are permitted to combine with one another,
c. The grammar followed to parse and annotate the text.

Some weaknesses in information processing and techniques such as browsing information without taking its meaning into account have recently appeared in Web Services. This creates the need for a new Web with more relevance to the user. Semantic Web as the new Web, is actually an extension of the current one in that it represents information more meaningfully for humans and computers alike. It enables the description of contents and services in machine-readable form, enables annotating, discovering, publishing, advertising and composing services. This shared description of documents is ontological annotation, and

it is considered as the backbone of the Semantic Web. In other words, the current Web is transformed from being machine-readable to machine-understandable. One function of the Web is to build a source of reference for information on several subjects, while the Semantic Web is to add specification which is both human and machine transferable.

Ontological annotation, in essence, provides formal and explicit specification of concepts in a domain. The domain in our case is the Yorùbá noun. The job of ontological annotation involves using programmable editing tools that is machine readable (written in a language understood by computers) to organise the concepts of a domain into different relationships. Consider as an example, this well familiar and a kind of ontology known as WORDNET: its organisation and the display of a lexical item 'school' in the browser as shown in Figure 1 below:

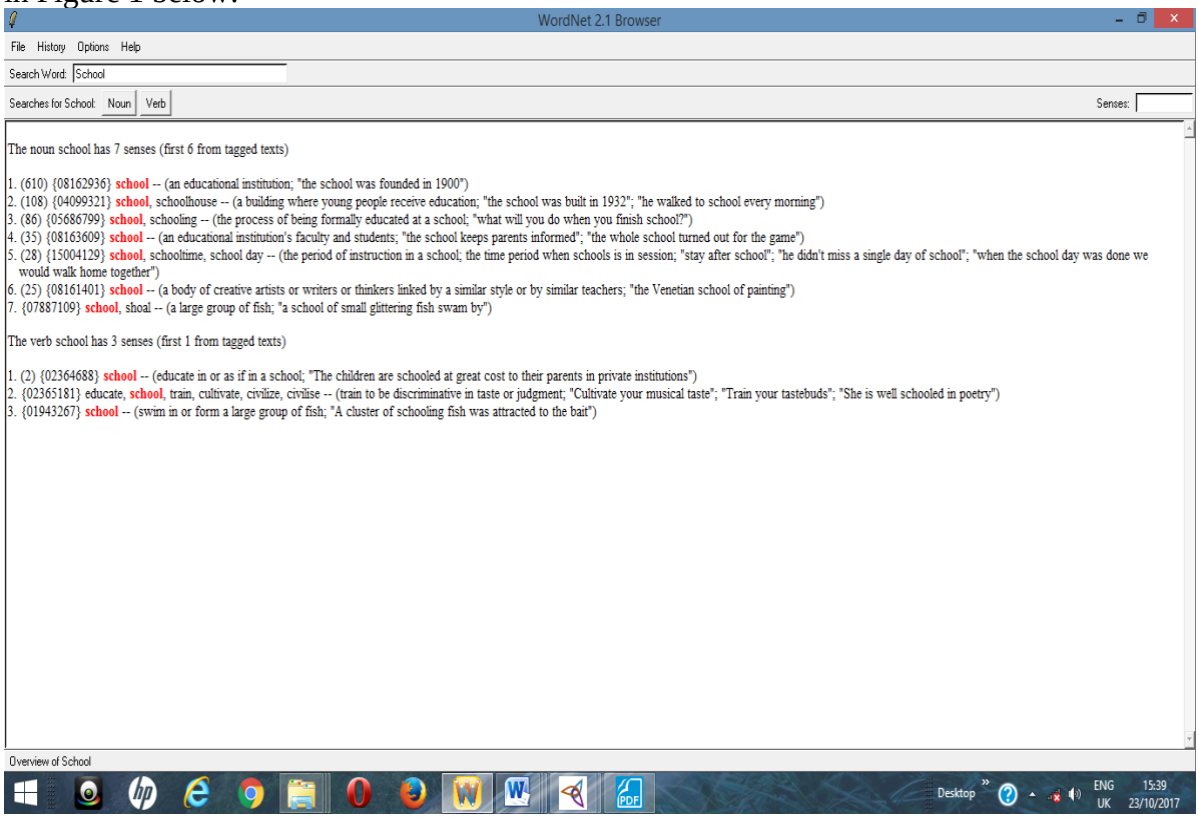

Figure 1: The Browser display for the item 'School' in WORDNET

This lexical item can still be further organized into different thematic relationship as exemplified in Figure 2 below:

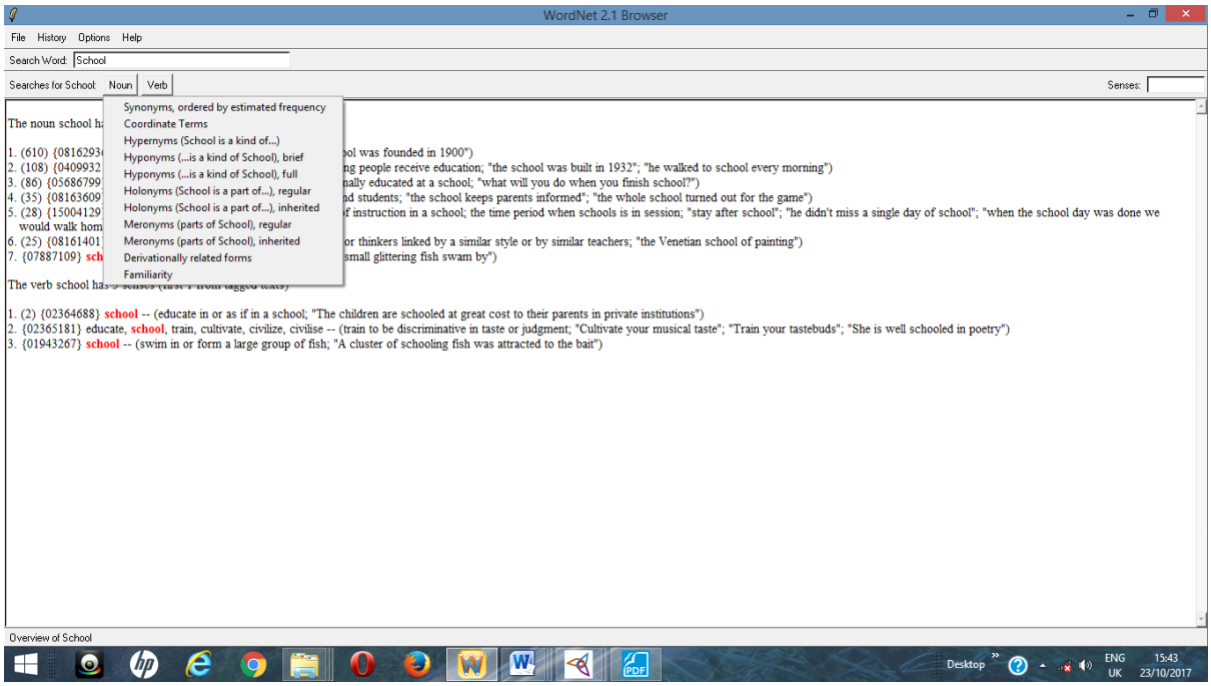

Figure 2: The Browser display for the item 'School' in WORDNET

The item can be represented as either a noun or a verb alongside the different meanings for each representation. It can also be processed and arranged into different taxonomic relations such as synonymy, hypernymy, hyponymy, holonymy and meronymy consisting of generalisation/specialisation hierarchy of concepts. WORDNET is an electronic thesaurus for human computer interaction. This is a layman explanation of activities involved in ontology. The ontological annotation designed and implemented in this paper goes further than organisation of nouns, but programs the Yorùbá noun taxonomy extracted from a resource text, using Protégé, a web-based editing tool expressed in RDF schema to implement the extracted concepts into semantic web making it machine-readable. The RDF schema is expressed in Web Ontology Language (OWL) on which Protégé is based. Through this medium, Yorùbá noun concepts implemented can be shared, reused and interoperated by web searching agents working on related NLP activities.

6. Review of Literature

6.1 Yorùbá Nouns

The focus of this section is to examine works that are related to Yorùbá nouns and Yorùbá NLP so that we can situate this study in its proper place among existing ones.

One of the earliest works that have endeavoured to describe Yorùbá grammar was Johnson (1921). The introductory part of this historical book contains the "Sketch of Yorùbá". The work noted the earlier efforts to produce the Yorùbá grammar from the immediate lines of English grammar which has made the earlier grammar more obscure. Regarding nouns, Johnson opines that simple nouns are generally formed by prefixing the vowels 'a', 'o', 'ọ' 'ẹ', 'ì' to a verbal root e.g. a+dé = *adé* (crown), ò+lù = *òlù*, (ginlet), ọ+lọ = *ọlọ* (grinder), ẹ+rù = *ẹrù*, (luggage), ì+mọ̀ = *ìmọ̀*, (knowledge).

This work admits that Yorùbá is non-inflective, hence it does not use terminal syllables to distinguish gender as done in English and other European languages. It only attaches the words: *akọ* (male) and *abo* (female) respectively to the common term or sometimes use "ọkùnrin" or "obìnrin". This is exemplified in words like *akọ ẹṣin* (a stallion), *abo ẹṣin* (a mare), *ọmọ ọkùnrin* (male child) and *ọmọ obìnrin* (female child). However, Johnson further explained that Yorùbá nouns inflects for number and case. This seems contradictory somehow, or possibly, what Johnson considers inflections about Yorùbá Nouns are not really inflections.

Awóbùlúyì (1978:7) defines nouns as any word functioning as the subject of a verb or the object of a verb or preposition in a grammatical sentence in the language, using this example:

> Òjó ra ìwé (Òjó bought books)
> Òjó is the subject while *ìwé* is the object. So, Òjó and *ìwé* are nouns.

Also, in the following structures:

Mo lọ sí oko.	(I went to the farm.)
Ó wà ní ìbẹ̀.	(He was there.)
Ó jáde ní kíákíá.	(He went out without delay.)
Fèrèsé náà wà ní ṣíṣí	(The window is opened.)

Oko (farm), *ibẹ̀* (there), *kíákí*á (quickly), *ṣíṣí* (opened) are nouns since they serve as the object of prepositions.

Awóbùlúyì did not contrast examples such as *ọjà* (market), *oko* (farm), *owó* (money), *èrò* (thought) of being nouns. However, the implications of his definition is that there are other words which earlier grammars did not analyse as nouns, but which are, in fact, nouns because they function as object of prepositions, including: *ibẹ̀* (there), *kíákíá* (quickly), *ṣíṣí* (opened). Awóbùlúyì (2013) is another important text on Yorùbá grammar that builds more on the earlier claims found in Awóbùlúyì (1978). For example, that *dà* and *ńkọ́* are verbs, has been re-explained in this new version. However, Awóbùlúyì (2013) expresses an improvement over the criteria for classifying word classes more than the previous edition. At least, instead of function criteria alone as earlier viewed, he mentioned forms, functions and syntactic behaviours as necessary criteria, and as a result, he pointed out five different features of Yorùbá nouns.

6.2 Works on Ontological Annotations

Some of the research works on ontological annotations in European languages include Gruber (1993:200), Borst (1997:20), Guarino, Oberle and Staab (2009:10), Studder, Benjamins and Fensel (1998). A common perspective in all of these works is that ontologies should be enshrined in the core aspects of computational structure; conceptualisation, specification of conceptualisations and shared ontological commitments. The classification aspect of ontologies depends on what they are used for. Concepts, attributes, class instances, relations and axioms are elements which are used to express these annotations. Roussey, Pinet, Ah kang and Corcho (2011), Stevenson and Ciravegna (2003) are similar in that they classified ontology into two: domain and linguistic ontologies. The first classification is based on the expressivity and formality of the languages used i.e. the natural language, formal language, or the type of Ontology that is applied to a domain with the specific viewpoint of how a group of users conceptualise and visualise some specific phenomenon within a domain. The domain Ontology is linkable to another specific application. For example, the plumbing network system could be linked to the whole architectural management system Kitamura and Mizoguchi (2004). In specific related example, the Ontology containing the Yorùbá grammatical classes such as proposed in this paper could be linked to an application processing natural language like machine translation, text analyser and so on.

The second classification based its scope on extraction and processing of human-language terms, without any direct reference to the real world. The main constructs of terminological ontologies are **synsets** (synonym sets) and the hyponymy/homonymy hierarchies. Example of such is WORDNET by Miller, Beckwith, Fellbaum, Gross and Miller (1990). Linguistic/Terminological ontologies exist in form of glossaries, dictionaries, controlled vocabularies, taxonomies, folksonomies Thesauri and lexical databases. The terms and relationship which the concepts hold with one another are expressed in their attributes and instances. Interestingly various terms could be stated to refer to one concept. For example, in

the Yorùbá land (South west Nigeria), the concept of femininity accepts the expressions "*obìnrin"*, "*abo"*, "*obí", "iyá"*, "*ìyámọ̀pó"* in referring to a woman, depending on formality and stylistic purposes. In fact, "*ìyámọ̀pó"* is a euphemism of womanhood. The roles of linguistic ontologies in these situations are twofold. The first is to provide the formal and mutual definition of the terms or vocabulary used. Secondly, it is necessary to provide the terminology agreement between a user's community and the mode or tenor of the discourse within which each of the items could be selected. More examples of this kind of ontological annotation are Farrar and Langeden (2003), Bateman, Henschel and Rinaldi (1995).

Some of these ontological annotations employed in European languages are also purpose-based. Examples are Mizoguchi (2005), Hama, Hori and Nakamura (1993), and Wielinga, Shcreiber and Breaker (1992). Other applications in this category that are domain-specific developed ontological annotations are Beeners-Lee, Hendler and Lassila (2001), focusing on e-business environment. Fensel (2001) developed his annotations to provide easy communications between agents and reduce misunderstandings between knowledge manager and electronic gadgets customers. Gangemi, Catenacci, Ciaramita and Lehmann (2006) looked at the problem of integrating heterogeneous sources of information in knowledge acquisition and sharing to develop Ontological Integration of Naïve Sources (ONIONS). This ontological annotation implements knowledge across various naïve sources, using schematic diagram and tools which is a refinement of that knowledge. We can go on and on to review ontological annotations implemented for the modern languages but we are constrained by time and space.

6.3. Yorùbá Corpus Natural Language Processing

Few individuals have made concerted efforts in applying computational resources to process Yorùbá language corpus. The work of Hassan, Ọdẹjọbi, Ogunfọlakan and Adejuwọn (2013) is the first and only attempt (as far as we know) to apply ontology into a Yorùbá concept. The work used formal concept analysis to analyse the data in Yorùbá cultural domain. The terms of Yorùbá cultural domains such as *agogo*, (gong), *bàtà*, (footwear), *ẹdan ògbóni*, (symbol for ògbóni fraternity) were gathered. The combination of observation, consultation, documented materials and prototyping was used to elicit samples of antiquities of Yorùbá cultural artefact. The cultural heritage samples were designed into taxonomy and protégé was used for the implementation of the prototype. The ontology was validated by the domain and ontology expert, but the end product of it was a software model for the development of semantic application for cultural heritage. Unlike this cited work, this study is dissimilar in that our domain of interest border on Yorùbá nouns. Also, the grammatical intricacies in our domain of application spells that our annotation design differs too.

Eludiora (2012) adopts phrase structure grammar as the linguistic background to build English to Standard Yorùbá Machine Translation System. It is based on the type two of Chomsky's formal context-free grammar. This work used the rule-based approach of Machine Translation (MT) to implement the system with python programming language. The source language (English) was translated into Yorùbá language as the target language. However, our observation is that this work noted that there were challenges in developing a Rule-Based Machine Translation (RBMT) systems for Yorùbá language because of the language intricacies, but it did not point out specifically that most of these anomalies in the web-based MT boils down to the fact that these online products do not have domain-specific ontologies at their background. If any Artificial Intelligent (AI) systems like the MT must be web-based, there is the need for formal, explicit specification of the resources and concepts in that particular domain which should serve as the backbone to the MT. That beams the search-light to ontology in the right direction.

Odoje (2017) complements his RBMT (Odoje 2010), with the Statistical Machine Translation (SMT) - a paradigm where translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora. This wok, focuses on whether statistical machine learning can perform exactly as human translation. It explains how difficult it is for an SMT to capture some intricacies such as cultural nuances, contextual meanings and semantic extension. Odoje opines that the Yorùbá verb *pa* can have different meanings, depending on contexts and usage patterns which the available corpora may not accommodate (Odoje 2017:70). It is our submission that a good ontological annotation

defined with all the semantic nuances of Yorùbá grammatical classes formally specified for machine readability and shareability will solve some, if not all of these teething semantic problems.

7. Methodology of Annotation

There are so many methods of ontology annotation as found in literature. Succinctly, ontology development process is based on the activities identified in the IEEE standard for software development. Some of the activities highlighted in Adekoya (2010) include:

1. Software life cycle model process
2. Project management processes which include planning, controlling and quality management
3. Development oriented-processes which consists of carrying out feasibility studies on possibility, requirement analysis (gathering, identifying and classifying entities, activities, agents and other key concepts as a starting point for information modelling and retrieval), system design (establishing framework for building the identified entities, activities, agents and other key concepts and how to handle the ambiguity and heterogeneity of the data and system implementation. Post-development processes which include hosting and installation, inter-operation support, maintenance and optimisation
4. Integral processes which include evaluation, documentation, configuration management and training of other NLP developers.

Other methods are outlined in Gruninger and Fox (1995), and Uschold and Gruninger (1996). The nature of data to be annotated determines the method to be adopted. This paper adopts Methontology as proposed by Gómez-Pérez, Fernández-López, and Corcho, O. (2004). The Yorùbá noun, according to Bamgbose's analysis which serves as our primary data, possesses more of lexical semantic attributes. Defining conceptual terms and attributes of the domain will make its abstraction easier than any of the outlined method. The highlight of this method is provided subsequently in the design section of the paper.

8.0 Materials

The materials discussed in the following sections were employed to carry out the implementation of Yorùbá Noun Ontology according to Bamgbose (YORNOB) domain ontology.

8.1 Protégé

Protégé was developed in collaborative research between the University of Manchester and The Centre for Biomedical Informatics Research, Stanford University. It is an open source, feature-rich ontology development environment for the Web which makes creation, uploading, modifying, and sharing of ontologies for collaborative works easily achievable. It has a full support for the Web Ontology Language (OWL), and direct in-memory connections to description logic reasoners. With its customisable user interface, one can create and edit ontologies in any compatible single workspace. This is made possible through visualisation tools which allow interactive navigation of ontology relationships. It is rich in advanced explanation support which aids tracking down inconsistencies. It exhibits refactor operations which allows ontology merging, moving axioms between ontologies, renaming multiple entities, which involves more knowledge representation techniques. The highly configurable user interface creates the perfect environment for ontology beginners and professionals alike.

Protégé 4.5 Desktop can be installed on various platforms like XMaCOSX, Windows, Linux, Unix Platform, Solaris, HPUX, AIX and other JAVA-enabled environment. It is distributed in the form of a ZIP file from the main Protege website, and includes the 64-bit Java Runtime Environment (JRE). So, it is not necessary to have Java pre-installed on the computer to run Protégé. Once the browser is opened, one can go to the Protégé 4.5 web site at http://protege.cim3.net/download/old-releases/. Various forms of protégé alongside its runtime

environment will be displayed. One will then click on the download button and after the displayed, click 'run'. This launches the 'install wizard'. The next thing to do is to follow the on-screen instructions to install Protégé in the Program /Files folder on the C: drive by clicking 'Next' for pre-install screens and then 'Install' when prompted. Several videos for demos and instruction manuals are always available online for consultation. For the purpose of our research, we registered as protégé user and several protégé user digest are always available for us online on how to deal with several issues and intricacies experienced during the use of this tool across different disciplines and data. The use of the OWLViz plugin, which allows the asserted and inferred classification hierarchies to be visualised, and is available from the CO-ODE web site, or can be installed when Protege 4.5 is installed is also recommended (but not necessary). Installation steps are documented and available on the above stated website.

Figure 3**.** The Protégé 4.5. Command Prompt

Figure 4**.** The untitled User Interface of Protégé 4.5.

The display in figure 3 is the outlook of protégé.exe compiler which compiles and executes any input tokens. Figure (4) is the untitled shots of the editor where each of the designed elements in the model is inputted one after the other.

8.2. Web Ontology Language (OWL)

The Web Ontology Language (OWL) is an international standard and semantic language for encoding and exchanging ontologies on the World Wide Web. The emphasis on the Semantic Web is such that information could be uploaded for humans to search and make queries ordinarily, but beyond this, information should be given explicit meaning in some formal presentation so that machines can process it more intelligently. Instead of mere developing standard terms for concepts as is usually done in XML, the Semantic Web provides formal definitions for the standard terms developed. Machines can then use inference algorithms to reason about the terms. The syntax of any application language depends on its performance. Because of specificity, which is required in constructing ontology, OWL is expressed in RDF graph and triples which occur in various syntactic forms. For example, the OWL syntax for the assertion 'Olú Àìná knows Dúpẹ́ Àbẹ̀kẹ́' is:

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/2007/02/22-rdf-syntax-ns#"
     xmlns:p="http://example.org/pers-schema#">
    <rdf:Description rdf:about="http://example.org/~dúpẹ́#àbẹ̀kẹ́">
        <p:knows rdf:resource="http://example.org/~olú#àìná" />
         <p:name>dúpẹ́ àbẹ̀kẹ́</p:name>
         <rdf:type
                rdf:resource="http://example.org/pers-schema#Person"/>
     </rdf:Description>
</rdf:RDF>
```

As with every other theoretical model, the weakness of one model leads to either the creation of a new model or the extension of the previous model to another one; the fact that RDF schema lacks strong semantic primitives brings about the development of OWL. In effect, each of the important RDF schema terms are either included directly in OWL or they are superseded by new OWL terms.

9. 0 Data for This Research

The prose narratives from Bamgbose (1990) is extracted and reduced to minimal token to form the conceptual knowledge of Yorùbá noun and its classification according to

Bamgbose. This forms the primary data for the design and implementation of the model. The coverage of knowledge extractable for Yorùbá noun is limited to Bamgbose (1990) because the theoretical issues in extant literature are enormous and may be too large to be implemented at a go for start-up ontology of this kind, the progression is always from the simple to the complex. However, the selection of this text is justified for its being straightforward, descriptive, precisely detailed enough with simplified linguistic rudiments to train machine and can also assist learners of Yorùbá language grammar. The identified properties of each entity in the prose are described as a prototype of the knowledge. These properties are arranged as tokens to form the artificial representation of entities combined to make a whole domain. The tokens are arranged to design an Intermediate Representation (IR) which carries all the semantic representation of the knowledge from the texts and serve as resource knowledge base. The IR can be a graphical or a tabular representation but the intermediate representation used for this design is tabular. This is because the refined grain of this procedure is closer to human readability than the graphical representation is, since the objective of our formalism is to enable both human and machine readability. The specific relations are defined before the ontology is implemented.

9.1 The Implicit Knowledge in Bamgbose (1990)

Bámgbóṣé's definition of Yorùbá nouns and the features with which to identify them are captured in the IR[1] designed ahead. It asserts that the behaviour of the noun in a structure could be the feature that will be used to mark its word class. The text identifies ten (10) contrastive noun classes. The prose narratives and the examples are translated in human readable form below:

1. Àrídìmú yàtọ̀ sí Àfòyemọ̀ (Concrete Vs Abstract) - Attributing Feature: Based on forms. (Visual/ non-visual). These nouns are identified with the form, whether visual, solid or not. It is a binary-feature of (+) or (-) of the feature value.
2. Àṣéékà yàtò sí Aláìséékà, (Countable Vs Uncountable) - Attributing feature: Based on number.

However, we add here that, some uncountable nouns can be quantified by qualifying those nouns with another group/kind describing nouns. For example, i*yọ̀* (salt) can be quantified by *àpò* (bag) to give:

Ó	bu	àpò	ìyò	mẹ́rin	
He/she draws	bag	salt	four		He/she takes four bags of salts

This shows that the same uncountable nouns can still be quantified when they are modified by other groups of nouns.

Aṣeékà (countable)	Aláìṣeékà (uncountable)
Ìwé	omi
Ajá	iyànrìn
Obìnrin	irú
Ewé	ẹ̀fọ́

Our further explanation then means that, nouns in the second group can be quantified to become:

Aláìṣéékà (uncountable)	quantified
Omi (water)	ìgò omi mẹ́ta (three bottles of water)

[1] The reader is implored to read Bamgbose (1990) for more details, but the narratives is captured and reduced in the model representation

Iyanrìn (sand)	ọkọ̀ iyanrìn méjì' (two loads of sand)
Ẹ̀fọ́ (vegetable)	ìdì ẹ̀fọ́ méjì (two packs of vegetables)

3. Ènìyàn yàtọ̀ sí Ènìyànkọ́ (Human Vs Non-Human) - Attributing Feature: Based on response to Question marker

   This class of nouns uses the question marker *ta* (who) for the human nouns while *kí* (what) is used for non-human nouns e.g.

Wọ́n rí olùkọ́	Ta ni wọ́n rí?
They saw teacher	Q-m for they saw (pst)
"They saw a teacher"	Who did they see?

4. Ọ̀rọ̀ orúkọ Ibìkan (Place Nouns) - Attributing Feature: Based on Preposing place question tag.

   This class also has a question marker attribute. This can be identified by preposing the place question marker *ibo* (where) to occupy the subject position. The answer to their proposed question gives this kind of noun:

Ó	wà	ní	Èkó	Níbo	ló	wà?	
He/she	is	in	Èkó	Q-m	that he/she		is?
He/she is	in	Lagos		Where is he/she?			

5. Ọ̀rọ̀ Orúkọ Ìgbà (Time/period Nouns) - Attributing Feature: Based on Preposing time/period question tag.

Ó lọ ní <u>àná</u>	Ìgbà wo ni ó lọ?
He/she go (pst) in yesterday	time which that he/she go (pst)
He/she went yesterday	What time/when did he/she go?

6. Òrọ̀ Orúkọ Òǹkà (Quantity Numeral Nouns) - Attributing Feature: Based on Quantity

   The attribute for this class is numerals. This class is marked by its numerical form. It is divided according to this work into three, namely:

Òǹkaye (cardinal)	ení, èjì, ẹ̀ta, ẹ̀rin etc. (One, two, three, four etc.)
Fígọ̀ (Figure)	oókan, èèjì, ẹ̀ẹ̀ta, ẹ̀ẹ̀rin (one, two, three, four etc.)
Òǹkàápò (ordinal)	èkíní, èkejì, ẹ̀kẹ̀ta, ẹ̀kẹ̀rin (first, second, third, fourth etc.)

7. Ọ̀rọ̀ Orúkọ Òye (Value Nouns) - Attributing Feature: Based on value question tag

8. Ọ̀rọ̀ Orúkọ Asọ̀ye (Quantifying nouns) – Feature Attribute: Based on quantity.
   Those which give exact numbers of items e.g. mẹ́wàà (ten), ẹgbẹ̀rún (thousand), the relative quantifiers which just quantify to some unspecified sets e.g. *díẹ̀*, (few/a few), *púpọ̀*, (many), *ogúnlọ́gọ̀* (several) and lastly, the universal quantifier which identifies the whole set without indicating exact numbers.

9. Ọ̀rọ̀ orúkọ Aṣàfihàn (Demonstrative nouns) It is attributed by the pointing to demonstrative qualifier. These nouns are used to form demonstrative qualifiers. Examples are *èyí, ìyẹn, ìwọ̀nyí, ìwọ̀nyẹ́n, iwọnni*. i.e. 'this', 'that', 'these', 'those', 'those of others'

10. Ọ̀rọ̀ orúkọ Aṣèbéèrè (Interrogative Nouns): It is attributed by the question markers.

Báḿgboṣé concludes his presentation on Yorùbá nouns by adding that Yorùbá nouns (i) Can be monosyllabic, disyllabic and polysyllabic; (ii) Can fall into more than one of these classes.

10. The Design of YORNOB

Methontology establishes the stages through which the ontology moves during its lifetime and the activities to be performed in each stage. It proposes how to organise the informally perceived view of a domain using a set of intermediate representations (IRs) based on tabular and graphical notations that can be understood by domain experts and ontology developers. In Methontology, ontological development lifespan relies on prototype evolved. Terms in each new version, that is the prototype is allowed to be added, changed, and removed. Pre-existing ontologies could be reused at any possible point in time in the evolved prototype life cycle, but it must be evaluated and adapted as the case may be. And if reuse is not feasible, then another one has to be created afresh. The elements of Methontology are concepts, relations and taxonomies. Elements or individuals in ontology are represented as Instances (Pareja-Lora, 2012). A prototype of the narrative knowledge on Yorùbá nouns as contained in Báḿgboṣé (1990) was extracted. The extraction was presented and validated for proper scrutiny by the domain Experts of Ibadan Language Study Group under the mentorship of Professor Oye Taiwo in the Faculty of Arts, University of Ibadan on the 28th of March 2017. The resulting extraction and design were presented in the following intermediate graphic representation in table (1) below:

Table 1: KEO (Knowledge Extraction Ontology), Extracted from Báḿgboṣé (1990)

Ọ̀rọ̀ orúkọ (Noun)	θ	Ọ̀rọ̀kọ́rọ̀ tí ó lè dá dúró tàbí kí ó gba ẹ̀yán nínu àpólà orúkọ. (Any word that can stand independently or take qualifier in a noun phrase)		Definition in relation to noun phrase i.e. content of N.P.	Non classical		Concept
Ọ̀nà tí a fi ń ya ọ̀rọ̀-orúko sọ́tọ̀ (criteria for identifying Noun)	θ						
Lílò bí ẹ̀yán ajọ́rukọ (Use as a qualifier)	θ	Ọ̀rọ̀ tí a bá lè fi yán ọ̀rọ̀ míràn gẹ́gẹ́ bí ẹ̀yán ajọ́rukọ. (Any word that can qualify another word as a noun qualifier). e.g. ilé ọba 'The King's house', Ọkùnrin olówó,'The rich man'.	-Qualify it with a noun qualifier. -If the item begins with a consonant, then the vowel should be lengthened	Only if the qualifier noun begins with a vowel	Stand in relation to other words		Concept
Gbígba ọ̀rọ̀ arọ́pò orúkọ gẹ́gẹ́ bí ẹ̀yán.(Taking	θ	Ọ̀rọ̀ tí ó bá lè gba ọ̀rọ̀ –arópò-ọ́rúkọ gẹ́gẹ́ bí ẹ̀yán gbọ́dọ̀ jé ọ̀rọ̀ orúkọ. (Any word that can take a pronoun as qualifier must be a	-Qualify it with a pronoun qualifier.	--	Stand in relation to other words		Concept

a pronoun as qualifier)		noun.) E.g. ilé e rè 'Your house' owó o wa 'our money					
Ìsọdorúkọ pẹ̀lú **oní** tàbí **ki**.(Nominalization by prefix **oní** or **kí**	θ	Ọ̀rọ̀ tí a bá lè ṣe ìsodorúkọ yìí fún gbọ́do jé ọ̀rọ̀-orúkọ. (If the prefixes can be added to a word to form a noun then that word must be a noun) e.g. ilé – onílé or ilékílé house of 'house owner'/ or 'any house'	Nominalize with oní/kí		Derivational with morphemes		Concept
Fífi wúnrẹ́n ìbéèrè dípò nínu gbólóhùn ìbéèrè. **(**Use of Question tag 'ta' and 'kí' to replace in interrogative sentences	θ	Nígbà tí a bá ṣe ìbéèrè, ó lóju wúnrẹ́n tí a gbọ́dọ̀ lò fún ìsòrí ọ̀rọ̀ orúkọ. A gbọ́dọ̀ lo 'kí' fún ọ̀rọ̀ – orúkọ tí kì í ṣe ènìyàn, 'ta' fún ẹ̀dá tí ó jẹ ènìyàn (Any word that can replace these questions tags in a structure must be a noun. 'ta' is used for human nouns and 'kí' is used for non-human nouns eg: Kí ni mo rí? 'What did I see?' Mo rí ilé 'I saw a house"	Prepose with question tag in a wh-structure to form a noun reply. Use 'kí' for –human and 'ta' for +human	This is restricted to only nouns.	Derivational with question tags.		Concept

Ọ̀rọ̀ orúkọ Àrídìmu (Concrete Nouns)	θ	"…ni èyí tí a lè rí tàbí tọ́ka sí…" (These nouns are identified with the form of visual, solid or not). e,g,.Ẹja 'fish', Àpótí 'stool' Ìwé 'book', Ọmọ 'child'.	Use the splitting verb fì … kan (to touch) to determine the kind of nouns involved		Its forms is physical/Concrete.	+	Concept
Ọ̀rọ̀ orúkọ Àfòyemọ̀ (Abstract Nouns)	θ	"…kò ṣeé fojú rí…" (These nouns are identified with the form of not being visual or solid.) eg ìfé 'love', alááfíà 'peace', ìmọ̀ 'knowledge' ikú 'death',	Use the splitting verb fì … kan (to touch) to determine the kind of nouns involved.		Its forms is discrete/Abstract	-	Concept
Ọ̀rọ̀ orúkọ Àṣéékà (Countable Nouns)	θ	"…ní èyí tí a lè lò pẹ̀lú ọ̀rọ̀ òǹkà nítorí pé irú ọ̀rọ̀ orúkọ bẹ́ẹ̀ sẹ́e kà…" (These nouns are identified by its being countable. The numeral nouns are indicators of this class of nouns.) eg: Ìwé 'book', Ajá 'dog', Obìnrin 'woman', Aṣọ 'cloth'	Place the item before the numeral noun to determine if it is semantically meaningful		Number dependent	+	Concept
Ọ̀rọ̀ orúkọ Àlaiṣéékà (uncountable Nouns)	θ	"… kò ṣéé lò pèlú ọ̀rọ̀ òǹkà. (These nouns are identified by its being non-countable. The numeral nouns are indicators of this class of nouns.) e.g.: Omí 'water', Iyànrìn 'sand', Ẹ̀fọ́ 'vegetable', iyọ̀ 'salts'	Place the item before the numeral noun to see if it is not semantically meaningful	These nouns can be quantified by qualifying those nouns with another group/ kind describing nouns eg àpò iyọ̀ mẹ́rin, 'three bags of salts', ìgò omi meta 'three bottles of water'.	Number dependent	-	Concept
Ọ̀rọ̀ orúkọ Ènìyàn(Human Nouns)	θ	"… ní èyí tí a lè fi **ta** ṣe ìbéèrè fún…" (The question marker **ta** is used for the human nouns) eg.	Prepose the item with the wh question 'ta'		Response to wh- marker.	+	Concept

		olùkọ́ 'teacher', akọ̀wé 'secretary', ìyàwó 'wife', sọ́jà					
Ọ̀rọ̀ orúkọ Èníyànkọ́ (Non-Human Nouns)	θ		Prepose the item with the wh question 'kí'		Response to wh- marker.	_	Comcept
Ọ̀rọ̀ orúkọ ibìkan(Place Nouns)	θ	"… ni èyi tí a lè fi ibo ṣe ìbéèrè fún tí a sì ń lo aṣèránwọ́ ọ̀rọ̀ – ìṣe t**í/gbé** pẹ̀lú rẹ̀…" (This can be identified by preposing the place question marker "ibo" (where) to occupy the subject position. eg Oyo, ilé-ìwé 'school', orí 'head', ìdí 'bottom', ibí 'this'	Prepose the item with the wh question 'ibo'		Response to wh- marker.	+	Concept
Ọ̀rọ̀ orúkọ ìgbà (Time/period Nouns)	θ	"… ni èyí tí a lè fi **ìgbà wo** ṣe ìbéèrè fún…"(The noun is realized as an answer to the question marker: **Ìgbà wo** (when/what time) e.g odún 'year', oṣù 'month' alẹ́ 'night' ìrọ̀lẹ́ 'evening', ọ̀ṣẹ̀ 'week'	Prepose the item with the wh question ' Ìgbà wo '		Response to wh- marker	+	Concept
Ọ̀rọ̀ orúkọ oǹkà (Quantity Numeral Nouns)	θ	"…ni èyí tí a ́maá ń ṣẹ̀dá ẹ̀yán òǹka lára rè. Oríṣii wọn ni: Òǹkaye: ení. èjì, ẹ̀ta abbl, Figọ: óókan, èèjì, éèta abbl, Òǹkàpò: èkiní, ẹ̀kejì, ẹ̀ketà abbl (This class is marked by forming numerical qualifier from it. Its types are: Cardinal: one. Two, three etc, Figure: one, two, three			Reflects Quantity	+	Concept

		etc, Ordinal: First, second, third etc.)					
Ọ̀rọ̀ Orúkọ Òye (Value nouns)	θ	"… ni èyí tí a lè fi èló ṣe ìbéèrè fún.b.a. Náírà méta, àpò méjì (The question marker 'èló' (how much) is used to mention the specific value it possesses, eg three naira)	-Qualify the item with the wh question ' èló' -Use numeral qualifier concept if the numeral qualifier will be deleted eg Ó ku kọ́bọ̀ ?	Restricted to only the 'one' qualifying noun.	Response to wh- marker.	+	Concept
Ọ̀rọ̀ Orúkọ Asọye (Quantifying nouns)	θ	"… ni èyí tí o maá ń díwọ̀n ọ̀rọ̀ orúkọ tí ó tèle e.b.a. **ọ̀pọ̀** nínu 'ọ̀pọ̀ ènìyàn', **ṣàṣà** nínu 'ṣàṣà obìnrin' (This noun quantifies the following noun.eg. **many** in 'many people', **few** in 'few women'	Quantify the following item with it.	Restricted to environment only before other nouns.	Quantity based	+	Concept
Ọ̀rọ̀ orúkọ àṣafihàn (Demonstrative nouns)	θ	"… ni èyí tí a maá n ṣèdá èyán àsafìhàn láti ara rẹ̀. b.a. èyí, ìyẹn, ìwọ̀nyí, ìwọ̀nyẹ́n, iwọnni. (This nouns are used to form demonstrative qualifiers e.g. this, that, these, those, those of others)	Preceed it with another noun to form demonstrative qualifiers	Restricted to environment only after other nouns.	Pointing based	+	Concept
Ọ̀rọ̀ - orúkọ Aṣebèérè	θ	"… ni a fí máá ń rọ́pó orúkọ míìràn nínu gbólóhùn ìbéèrè. b.a. kí, ta, èwo, èló, ibo, èkelòó. (This noun is used to replace another noun in an interrogative sentence eg. what, who, which, where, how many/which number)	Replace it with another noun in an interrogative sentence.	Restricted to sentential context only.	Interrogative	+	Concept

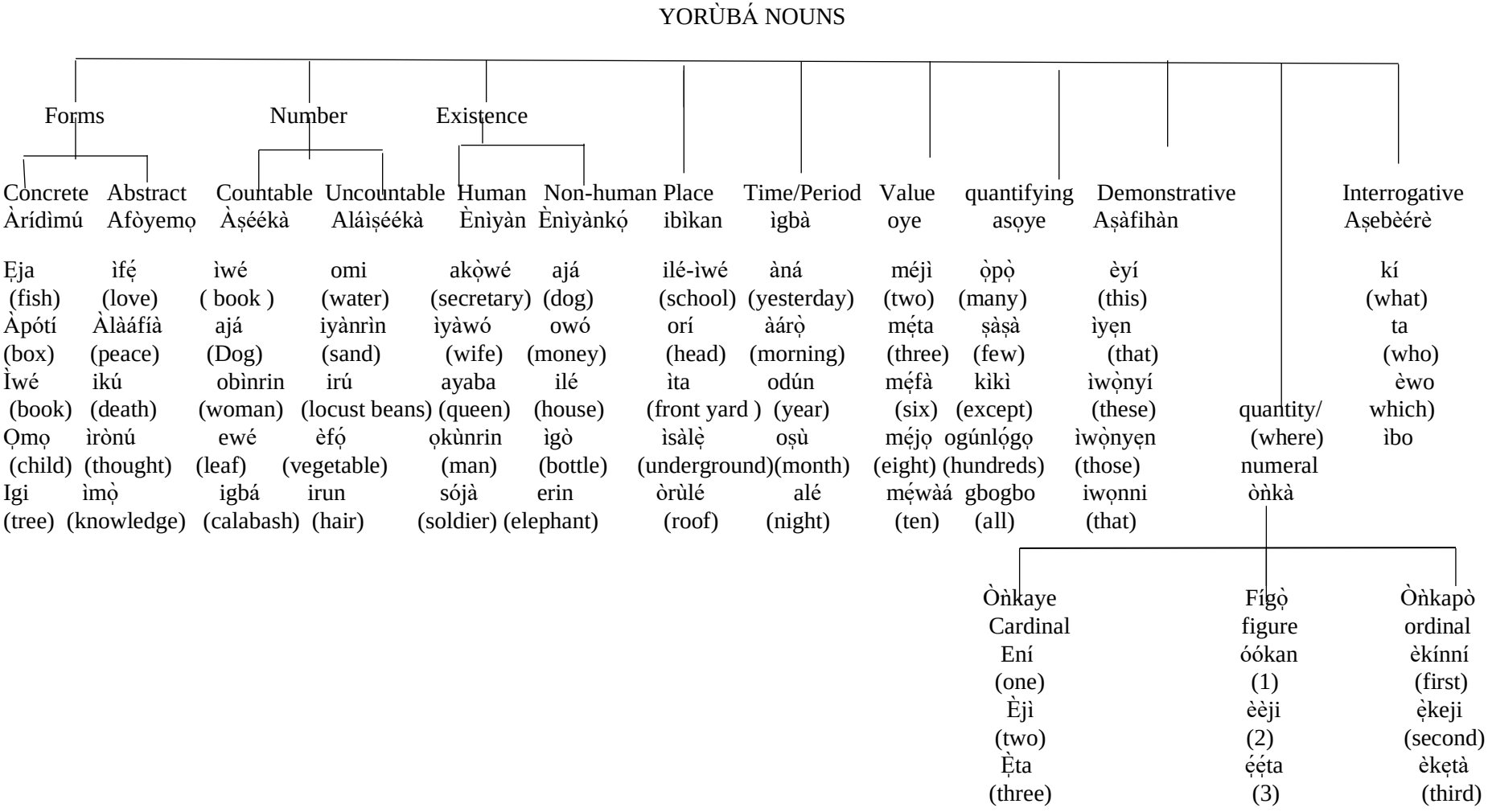

Table 2: Taxonomic Hierarchy of Yorùbá Nouns According to Attributing Feature (Bamgbose 1990)

11. Yorùbá Noun Identification Criteria Captured in RDF

The work provides us with four different criteria or features to identify a noun as follows:

a. Its use as a noun qualifier**:** Any word that can be used to qualify another word as a noun qualifier is a noun e.g. *ilé ọba* (the King's house), *okùnrin olówó*, (the rich man). *Ọba* (king) and *Olówó* (rich man) are noun qualifiers. Therefore, *ilé* (house) and *okùnrin* (man) too must be nouns alike. It added that if the qualifier noun should begin with a consonant, the vowel ending the first noun will be lengthened as in *ilé e bàbá* (Baba's house), *ajá a délé* '(Dele's dog).
b. Taking a pronoun as qualifier: Any word that can take a pronoun as qualifier must be a noun, e.g. ilé e rẹ '(your house), owó o wa (our money).
c. Nominalisation by the prefix *oní* or *kí*: If the prefixes can be added to a word to form a noun, then that word must be a noun, e.g.

ilé – onílé or ilékílé
house of 'house owner' or 'any house'

d. Use of Question tag *ta* and *kí*: Any word that can replace these question tags in a structure must be a noun. *ta* is used for human nouns and *kí* is used for non-human nouns as in:

Kí ni mo rí?	What did I see?	Mo rí ilé.	'I saw a house.
Ta ni mo rí?	Who did I see?	Mo rí olùkọ́.	'I saw a teacher'

Words are used to explain and define words in any human concept. In the same vein, words are employed to express and define Yorùbá nouns as done by Bámgbóṣé. In this same order, Ontology captures the words used to define a concept and redefines them in precise explicit language. Therefore, the four definition criteria by Bamgbose are captured and redefined in the formal expressions below.

Yorùbá Nouns
- QualifiesAnotherNoun
- TakesPronounAsQualifier
- BePrefixedByOtherNouns
- IsPreposedBy*Kí*
- IsPreposedByTa'

The RDF Schema for the above expression appears in the set of triples called an RDF graph as shown in figure 5 below:

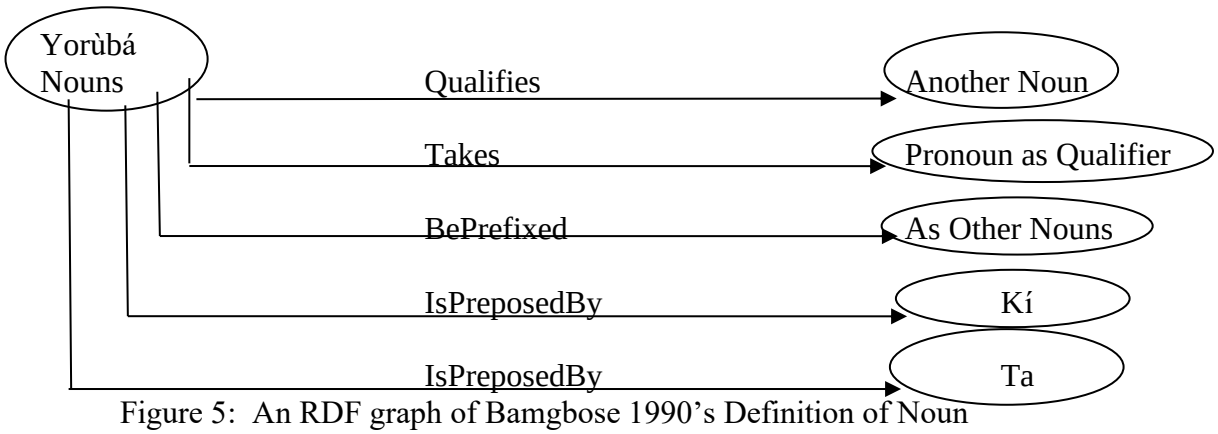

Figure 5: An RDF graph of Bamgbose 1990's Definition of Noun

The 'Yorùbá Nouns' in the ova are the subject, according to the RDF Schemata, The Linguistic Units - 'Qualifies', 'Takes', 'BePrefixed', 'IsPreposedBy' and 'IsPreposedBy' are

the Predicate i.e. the property is regarded as The Linguistic Attributes, while the other constituents in the remaining ovum are the object, i.e. The Linguistic Value (Pareja-Lora, 2012).

In order to facilitate the sharing and exchanging of graphs on the web, an XML serialisation would be defined. The RDF/XML serialisation for the above triples is written as follows:

<rdf : Description rdf: about = '#Yorùbá Nouns'>
< qualifies rdf resource = '# AnotherNoun'/>
< takes rdf resource = PronounAsQualifier'/>
< prefixed rdf resource = AsOtherNouns'/>
< preposedBy rdf resource = 'Ki'/>
< preposedBy rdf resource = 'Ta/>
</preposed by>
</rdf: : Description>

The screenshot of this RDF expression represented in the compiling notepad is shown in figure 6 below:

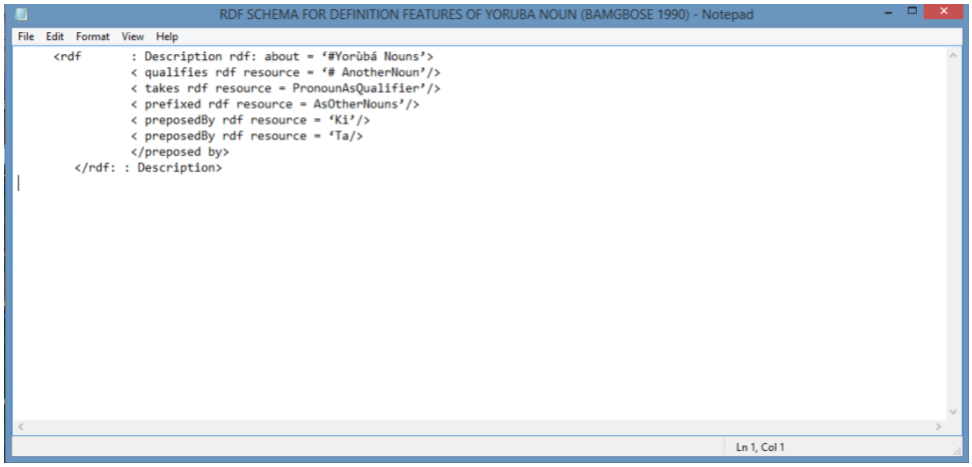

Figure 6. RDF Schema for Definition Features of Yorùbá Nouns (Bamgbose 1990)

12. Implementational Description of Yorùba Noun

The ontology design is not completed expressing the concepts in the domain with RCA alone. Some of the salient knowledge elicited from the narratives need to be expressed, showing the varying degrees of complexity of the classes and sub-classes of the items in the domain by representing them in various object property axioms as follows:

- Ọ̀rọ̀-orúkọ Àrídìmú *isIdentifiedBy* sense of touch
- Ọ̀rọ̀-orúkọ Àrídìmú *isAnInversedForm* of Àfòyemọ̀
- Ọ̀rọ̀-orúkọ Àfòyemọ̀ *isIdentifiedBy* sense of emotional thoughts
- Ọ̀rọ̀-orúkọ Àfòyemọ̀ *isAnInversedForm* of Àrídìmú
- Ọ̀rọ̀-orúkọ Àṣeeka *isIdentifiedBy* its countability
- Ọ̀rọ̀-orúkọ Àṣeeka *isAnInversedForm* of Àláìṣeékà
- Ọ̀rọ̀-orúkọ Àláìṣeékà *isIdentifiedBy* its non-countability
- Ọ̀rọ̀-orúkọ Àláìṣeékà *isAnInversedForm* of Àṣeeka
- Ọ̀rọ̀-orúkọ Ènìyàn *isDefinedBy* its reply to ‘ta’
- Ọ̀rọ̀-orúkọ Ènìyàn *isAnInversedForm* of Ènìyànkọ́
- Ọ̀rọ̀-orúkọ Ènìyànkọ́ *isDefinedBy* its reply to ‘kí’
- Ọ̀rọ̀-orúkọ Ènìyànkọ́ *isAnInversedForm* of Ènìyàn
- Ọ̀rọ̀-orúkọ Oǹkà *isIdentifiedBy* its number qualifier
- Ọ̀rọ̀-orúkọ Oǹkà *isSuperordinateTo* Òǹkaye
- Ọ̀rọ̀-orúkọ Oǹkà *isSuperordinateTo* Fígọ̀
- Ọ̀rọ̀-orúkọ Oǹkà *isSuperordinateTo* Òǹkapò

The axioms above entail one-to-many relations and their instances subscribe to that of mutual exclusiveness.

- Ọ̀rọ̀-orúkọ ibìkan *isDefinedBy* its reply to ‘ibo’
- Ọ̀rọ̀-orúkọ ìgbà *isDefinedBy* its reply to ‘ìgbà wo’
- Ọ̀rọ̀-orúkọ oye *isDefinedBy* its reply to ‘èló’
- Ọ̀rọ̀-orúkọ Asọ̀ye *isIdentifiedBy* quantifying the following noun
- Ọ̀rọ̀-orúkọ Aṣàfihàn *isIdentifiedBy* its ability to point demonstrative qualifiers
- Ọ̀rọ̀-orúkọ Aṣèbeere *isIdentifiedBy* its ability to replace another noun

The axioms above entail one-to-many relations.

The top object properties of these axiomatic expressions described above were modelled as shown in figure 7 below. Figures 8 and 9 show the class hierarchy of Yorùbá Nouns and re-

expression in Onto-Graf while figure 10 shows the annotation property view of Yorùbá nouns according to Bamgbose (1990).

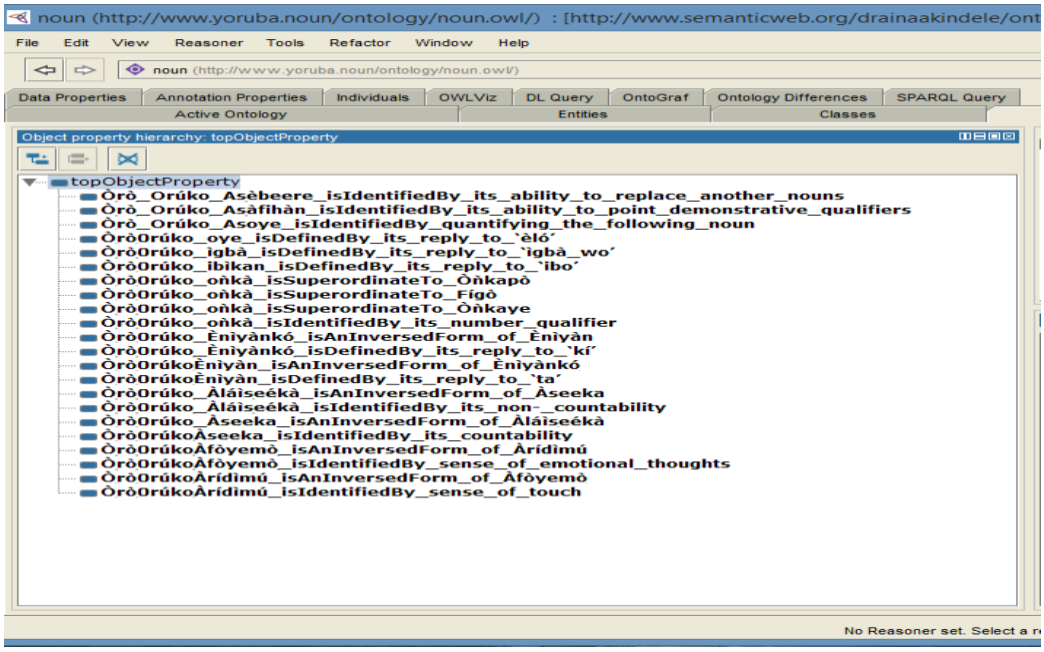

Figure.7: The Axiomatic Object Properties Hierarchy of Yorùbá Nouns Ontology

Properties represent relationships between two individuals. The implemented axiom properties discussed in section 12 is modeled in the top object property node of the editor in figure 7. What we see in the outward interface consist at the background, the RDF schema for each item in the model., Through the annotation schema, semantic web activities can be performed by web agents.

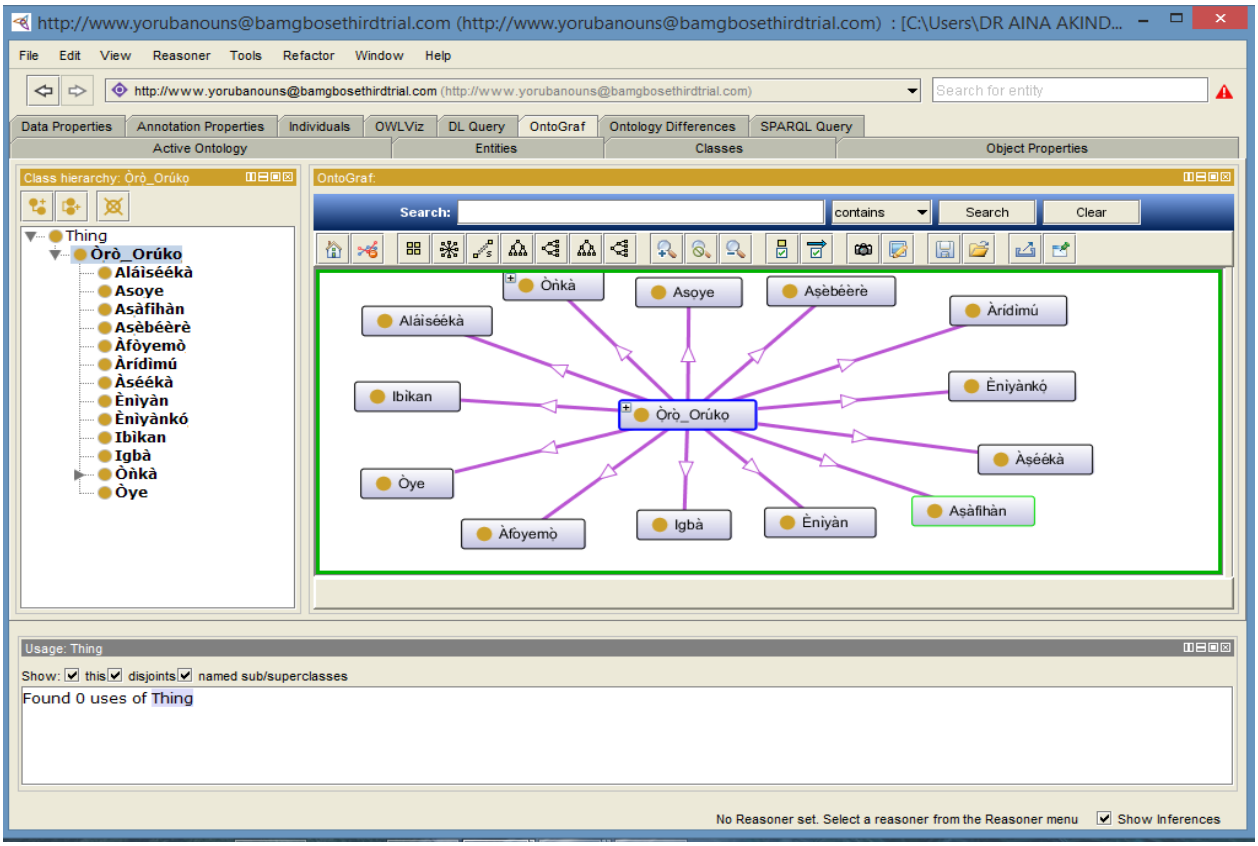

Figure 8: The Class Hierarchy of Yorùbá Nouns Re-expressed in OntoGraf

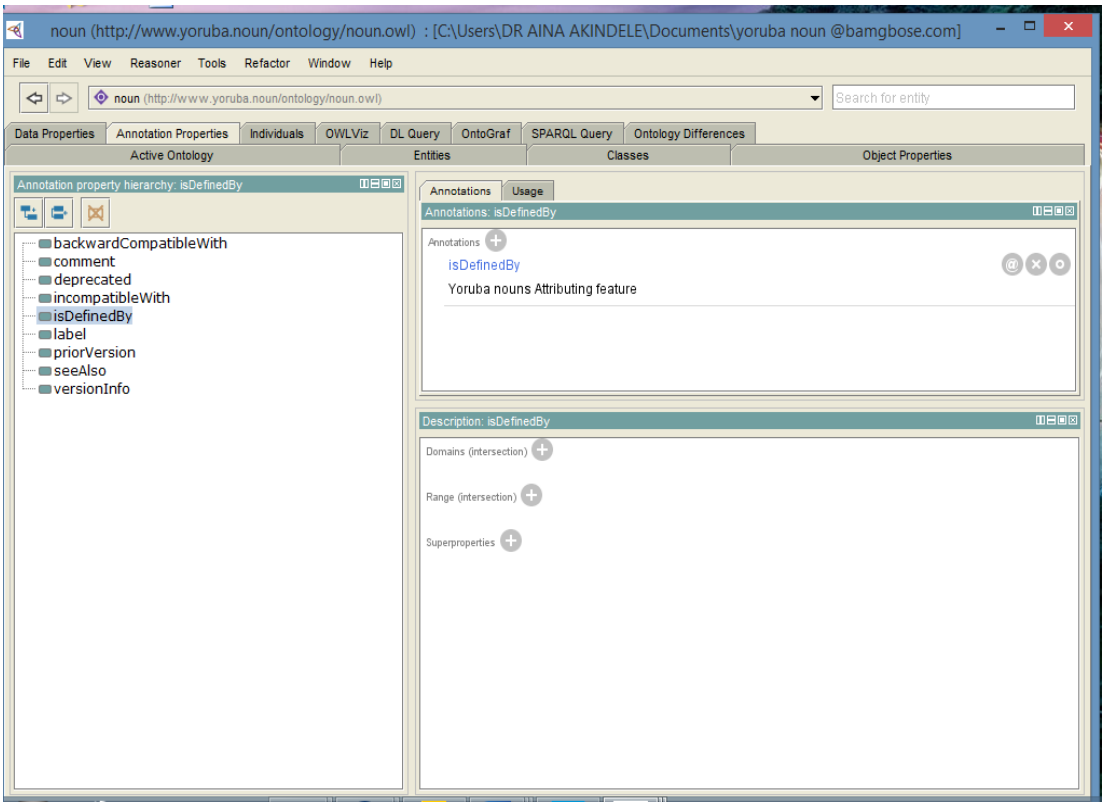

Figure 9: The Annotation Property View of Yorùbá Nouns of Bamgbose (1990)

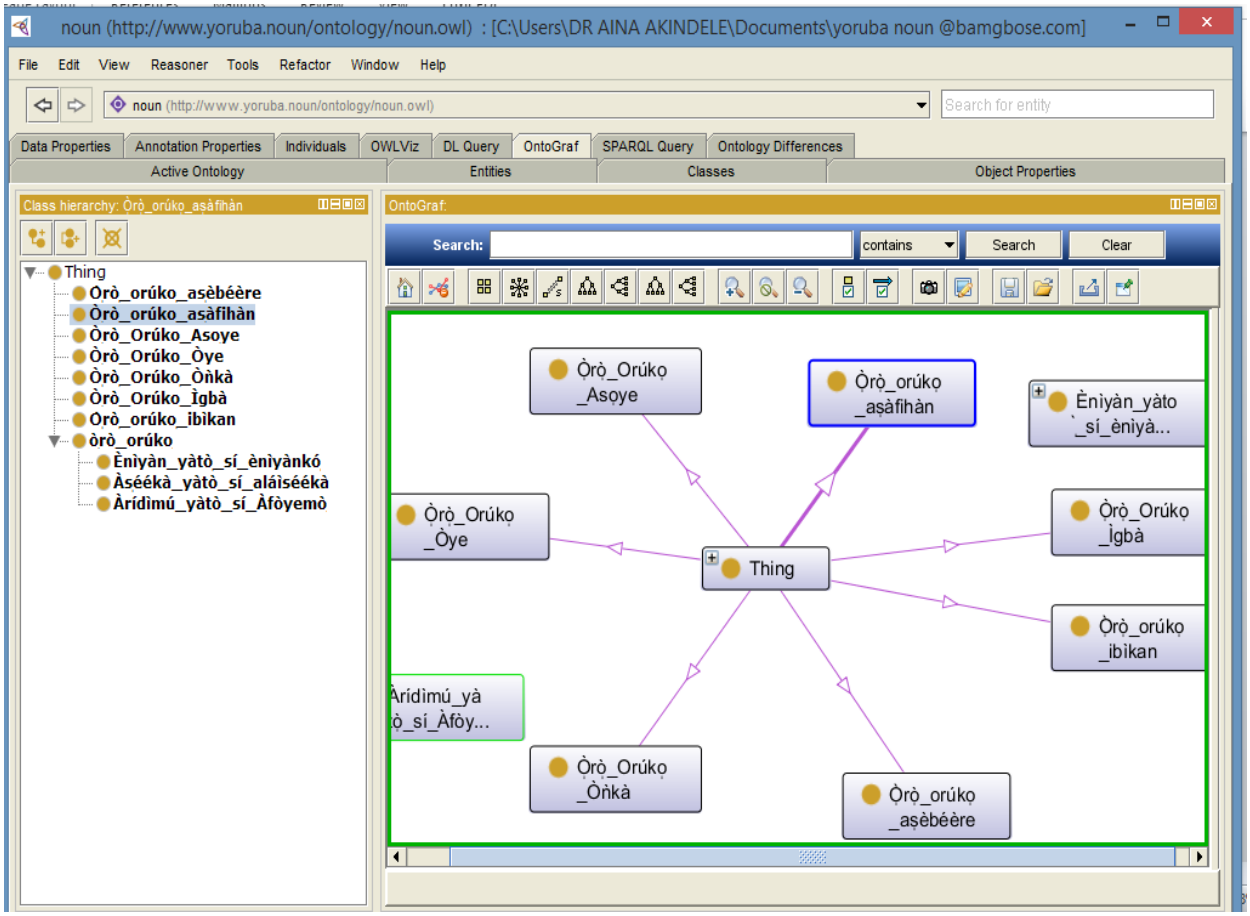

Figure 10: The class Hierarchy of Yorùbá Nouns Re-expressed in OntoGraf

Figures 8-10 are further annotation properties implemented in YORNOB. These are few displays of the model but behind the scenes are codes which are machine readable and needed for NLP activities.

13. Justification of some NLP Systems Utilising YORNOB

As explained previously, Ontology leads to the sharing of knowledge between systems and people. It plays a key role in the Semantic Web supporting information exchange across distributed environments. The Semantic Web represents data in such a way that it can be machine-processed. The following applications are some NLP applications that will need to utilise YORNO:

1. Artificial Intelligent systems
2. Decision Support Systems
3. Semantic Web Service Discovery
4. Multi agent Systems
5. Search engines
6. E- Commerce

However, the subject of how any of the above listed applications can utilise YORNOB is a subject of further and independent research. Depending on the focus of the implementation of the different applications, system development activities had to be generated to meet with the inter-operability needs. Furthermore, some NLP applications in Yorùbá language like Odoje (2010, 2017), Eludiora (2012), Aina (2018) may utilise YORNOB in their source codes. Importantly, the requirements for ontology utility as espoused in Berners-Lee, Hendler and Lassila (2001) must be followed. These are highlighted briefly as follows:

(1) The ontology must be available on the web with an open licence.
(2) It must be machine-readable structured data (e.g., CycL instead of image scan of a table).
(3) It must have non-proprietary format (e.g., OBO instead of CycL).
(4) It must comply with the open standards from the W3C (RDF Schema and OWL).
(5) It must be reusable in other people's ontologies.

The guidelines and procedures in Garijo (2013) must be strictly followed to load YORNOB into the utilising applications. Let us briefly demonstrate this hypothetically, that Eludiora (2012) will need to utilise YORNOB in its lexicon module which comprises lexical items representation of the Yorùbá parts of speech, let us say noun in more specific sense. Eludiora uses NLTK as its implementation tool, and NLTK is implemented in Python programming language. Python is an object-oriented programming language that contains corresponding standard library files convertible to OWL. The YORNOB annotations URL accessibility must be approved from the admin of its hosting site located in http://purl.org .[2] Having explained this, the next task for Eludiora (2012) to utilise YORNOB is to dereference ontological annotations following the outlined tasks in vocabulary dereferencing according to Berners-Lee, Hendler and Lassila (2001). Some of the steps include:

1. Set the purl redirection for Semantic Web resources (add a 303 redirection instead of 302) and add the target URL[3] of YORNOB into the utilising application loop. The particular loop this time is Eludiora (2012)). The redirection is looped into the lexicon module of its E-SY MT (English to South-west Yorùbá Machine Translation). One can only dereference a resource, only if one is in control of the server where the resources are going to be delivered.

2 Adhere to W3C documents standard by using the link 'http://www.w3.org/TR/swbp-vocab-pub/#recipe3',] which is a simple redirection for vocabularies with a hash namespace and create an htaccess file for the link. For YORNOB, the index.html file has the documentation of the ontology, while yorno-ontology1.1.owl contains the rdf/xml encoding. All the files are located in a folder called yornob.content. In order to avoid an infinite loop when dealing with the redirections of the vocabulary it must be tuned in this 'middle way extract' of the set of instruction and employed:

```
# Turn off MultiViews
Options -MultiViews
# Directive to ensure *.rdf files served as appropriate content type,
# if not present in main apache config
AddType application/rdf+xml .rdf
```

However, it has been said that the focus of each application implementation will determine how it will be utilised but the above procedure is a demonstration of YORNO utility, more so that approaches towards system development are many, depending on the tools and functions of the system.

Conclusion

This paper has introduced us to ontological annotations as a formal method of expressing relations in a specific domain. We have chosen the Yorùbá grammatical concept which is restricted to nouns only, and in accordance with Bamgbose's classification. We have presented the attributing properties for its sub classification, using FCA analysis table. The elicited knowledge has been defined for its object properties, axiomatic expression and protégé 4.5 which is expressed in OWL is used as the implementation tool. The justification of the

[2] The URI for all our YORNO models including the one developed in this paper is targeted at: http://purl.org/net/yorno.

[3] To link the URI, place the cursor on the ontology IRI of the ontology header, the code is displayed automatically and can be retrieved for linking the loop.

procedure for utilising our model into other NLP system is demonstrated, using the proposed lexicon module of Eludiora (2012). This attempt is a demonstration that ontology is a means to an end in natural language processing. The annotated nouns which were formally specified in this work are useful for shareability, interoperability and reuse in knowledge-based systems and artificial intelligent systems even though it may not be an end directly in itself but it serves as a means to an end. If researches are further geared towards producing modern electronic applications which make use of the machine-readable model developed in this work at the background, the Yorùbá language will be developed and modernised for nation's economic growth and development.

REFERENCES

Adegbola, T. 2006. Globalization: Colonizing the Space Flows? *Globalization and the Future of African Languages*. F.O.Egbokhare and C. Kolawole. Ibadan Cultural Studies Group, Ibadan. 2 – 11.

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.

Aina, A. A. 2018. The Development of Gege Akoto- A Yoruba orthography Editing Model. *Yorùbá: Journal of Yoruba Studies Association of Nigeria*. 8(1).

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.

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

Awoyale, Y. 1974. Studies in the syntax and semantics of Yorùbá nominalisation. PhD Thesis. University of Illinois.

Bamgbose, A 1966. *A Grammar of Yorùbá:* Great Britain: Cambridge University Press

----------1990. *Fonoloji ati Girama Yorùbá*: Ibadan: UPL

Bateman, J. A., Henschel, R., and Rinaldi, F. 1995. Generalized upper model 2.0: documentation. Technical report, GMD/Institut f¨ur Integrierte Publikations- und Information system, Darmstadt, Germany.

Berners-Lee, T., Hendler, J. and Lassila, O. 2001. The Semantic Web: Scientific American. May, 2001. Retrieved July 12 2014 from http://www.sciam.com/

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

Eludiora, S.I. 2012. Development of an English to Yorùbá machine translation. PhD Thesis Dept. of Computer Science and Engineering. O.A.U. Ife.

Farrar, S. and D. T. Langedon. 2003. A linguistic Ontology for the Sematic Web.GLOT International.

Fensel, D. 2001. Ontologies: A Silver Bullet for knowledge Management and Electronic Commerce, Springer-Verlag.

Garijo, D., 2013. How to (Properly) Publish a Vocabulary or Ontology in the Web. Ontology Engineering Group, Universidad Politécnica de Madrid.

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.

Gomez-Perez, A., Fernandez-Lopez, M. & Corcho, O. 2004. Ontological Engineering: examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. *Springer-Verlag*, London, U. K.

Guarino, N., Oberle, D. and Staab, S. 2009. What is an 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.

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

Grüninger, M. and Fox M. 1995. Methodology for the design and evaluation of ontologies. IJCAI'95 Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, Canada

Hama, T., Hori, M., & Nakamura, Y. 1993. Task-Specific Language Constructs for Describing Constraints in Job Assignment Problems. Technical Report, IBM Japan

Hassan, J.A., Odejobi, O.A., Ogunfolakan, B.A. and Adejuwon, A, 2013. Ontology Engineering in Yorùbá cultural Heritage Domain. *African Journal of Comp. & ICT*s IEEE. 6.5:181-198.

Johnson, S. 1921. *The history of Yorùbás.* Cambridge University press.

Kitamura Y., Mizoguchi R. 2004. Ontology-Based Functional-Knowledge Modeling Methodology and Its Deployment. In: Motta E., Shadbolt N.R., Stutt A., Gibbins N. (eds) *Engineering Knowledge in the Age of the Semantic Web*. EKAW 2004. Lecture Notes in Computer Science, vol 3257. Springer, Berlin, Heidelberg

Miller, G.A., Beckwith, R., Fellbaum, K., Gross D. and Miller R. 1993. Introduction to WordNet: An On-line Lexical Database. Retrieved 24/02/2015 from https:// www.researchgate.net

McEnery, A. M. and Wilson, A. 2001. *Corpus Linguistics: An Introduction*. Edinburgh University Press. Edinburgh.

Mizoguchi, R. 2005. Knowledge Acquisition and Ontology, Proceedings of the KB and KBS 2005, Tokyo, Japan. 121-128

Odoje, C. O. 2010. English-Yorùbá Rule Based Syntax in Machine Translation. M. A. Project. Department of Linguistics and African Languages. The University of Ibadan, Ibadan.

----------2017. Statistical machine translation. PhD. Thesis. Dept. of Linguistics and Africa Languages. University of Ibadan. xvi+218pp.

Owolabi, K. 1976. Noun-Noun Constructions in Yorùbá: A Syntactic and Semantic Analysis. Ph.D Thesis. Dept. of Linguistics and African Languages, University of Ibadan.

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.

Rousey, C., Pinet, F., Ah Kang, M. and Corcho, O. 2011. An Introduction to ontologies and ontology Engineering. *Ontologies in Urban Development Project; Advanced Information and Knowledge Processing* 1:2. Falquet G. ed. Springer-Verlag London Limited.

Stevenson, M. and Ciravegna, F. 2003. Information extraction as a Semantic Web technology: Requirements and promises. In: Workshop on Adaptive Text Extraction and Mining (ATEM03) held with ECML/PKDD 2003, Cavtat.

Studer, R., Benjamins, R. and Fensel, D. 1998. Knowledge Engineering: Principles and Methods. *Data and Knowledge Engineering (DKE)* 25.1-2:161–197.

Uschold, M. and Gruninger M. 1996. Ontologies: Principles, Methods and Application. Knowledge *Engineering Review* 11.2:105-129.

Wielinga, B. J., Schreiber, A. T., & Breaker, J. A. 1992. KADS: A modelling approach to knowledge engineering. *Knowledge Acquisition* 4,1: 5-53

Zauzich, K. T. (1992), *Hieroglyphs without mystery: an introduction to ancient Egyptian writing*. Japan: University of Texas Press.