KIU Journal of Humanities Copyright©2019
Kampala International University ISSN: *2415-0843*; 4(4): 229–246

A Hybrid Yoruba Noun Ontology

AKÍNDÉLÉ ÀKÀNJÍ ÀÌNÁ,
Olabisi Onabanjo University, Ago-Iwoye,Nigeria.

Abstract. The primary purpose of ontological annotation in Natural Language Processing (NLP) is to make explicit the content of concepts in a domain, using a set of arbitrary tags or labels which must conform to an agreed standard. This method has contributed to the development of many European languages, but it has not been so much employed in African languages. This paper therefore develops a model to prepare the implicit knowledge of Yorùbá noun as found in the literatures into their component features so as to make it both human and machine readable. The ontology development process proposed by (Schultz, 1997) which is based on the activities identified in the IEEE standard for software development is the method of ontological annotations adopted to design the hybrid model for Yoruba noun. Bamgbose (1990)'s "*Fonọ́lọ́jì àti Gírámà Yorùbá"* , Awobuluyi (1978)'s *Essentials of Yorùbá Grammar* and Awobuluyi (1978)'s *Ẹ̀kọ́ Gírámà Èdè Yorùbá* were purposively selected and the implicit knowledge of Yoruba noun in these literatures were extracted as the data for implementing the model. 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 the hybrid model for Yorùbá Nouns. The model at the final edge specifies the semantic load and properties of each of the nouns captured according to the Yoruba Language Scholars' perspectives and classifications. The definitions in the annotation serve as backbones for machine learning, web searching and artificial intelligence agents and other (NLP) systems. The model through the annotation schema and label serves as repositories of data and make the content available for machine manipulation and semantic web reasoning. This paper recommends that more on ontological annotations for Yorùbá grammar concepts are needed to foster Yorùbá language engineering.

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

1. Introduction

Technological advancement in the world has brought about changes in human job performance activities. Designing and building knowledge based systems and machines to aid human tasks are now the order of the day. Systems such as Decision Support System (DSS), Artificial Intelligent Systems (AIS), Knowledge Based Management Systems (KBMS), Natural Language Processing System (NLPS) alongside their various software are being developed to increase efficiency of professionals and to bring about higher performance in commerce, agriculture, health, construction and language use. Human speeches are being analysed by speech synthesizers and speech taggers. (Jurarfsky and Martin 2006, 2009). Text analysis no longer needs human but automating text generating machine (Hutchins 1994, 1995, 2007). Language uses in different specialised domains are now being supported by Expert System (ES) (Andre and Rist, 1993, Appelt 1985, Bateman 1990). It is not an over statement to say that Computers and software programs are employed to carry out linguistic roles and human language intelligent tasks in all human endeavours. As great as these developments are, not much of it has been done in African Languages including Yoruba language, in fact Yoruba language is still been known as a resource scarce language. Also, as developments abound, the advancement in Internet activity has fostered sharing of information and re-use of the developed models. This brings about the calls to start building more models to support African Languages. To achieve this, there is the consequent need to develop vocabularies which are explicitly and formally defined for the relationship between concepts and meaning to shore backgrounds for the developed and the developing machine to support of African Languages.

Interestingly, few individuals have responded to the clarion call of employing computational methods to analyse Yorùbá language. Research in this area includes: Automated Speech Recognition (ASR) for Yorùbá Tone Systems in order to build capacity in human language (Adegbola 2006, 2008, and 2009), Machine Translations (Hassan 2009, Odoje 2010, 2013, 2017), a preliminary investigative work on Yorùbá Ontology development by Aina (2017). These are efforts to make Yoruba language grow in tune with modern trend. However, as much as these needs and the efforts are in place, there are enormous challenges regarding the inherent knowledge of language user which these models are solely dependent.

In the first hand, those artificial intelligent and knowledge-based models which had to be developed for Yorùbá language must operate upon the core grammar of the language, but the knowledge of core grammar are still implicit on pages of literatures and it has to be brought up and prepared for machine use. On the other hand, the varying viewpoints and assumptions on the structural and syntactic distribution in the context of the grammar of Yorùbá language from different scholars create background problems. In resource developed languages, when linguistic annotated tools are developed for Natural Language Processing (NLP) activities, differences and challenges that prevent these annotation soft wares from interoperating always emanate from variations in the factors taken into consideration when designing and annotating them. Factors such as methods of annotation, interoperable factors (which makes it difficult for one model to operate in another because of underlying intricacies) and in most times linguistic/theoretical factors whereby a developer annotates according to one linguistic theory and leave others. These factors are predominantly common and therefore pose serious problems to NLP activities. For example, consider the theoretical differences in the analysis of this English sentence as observed by Pareja-lora (2012:14)

"John gave Mary an apple"

The phrase tagger analysis will present it as tabulated in Table 1.1 and 1.2 as follows:

Table 1. 1. Machinese Phrase Tagger analysis of the English sentence 'John gave Mary an apple'. (Pareja-Lora, 2012:14)

Text	**Base form**	**Phrase syntax and part-of-speech**
John	John	nominal head, proper noun, single-word noun phrase
Gave	Give	main verb, indicative past
Mary	Mary	nominal head, proper noun, single-word noun phrase
An	An	premodifier, determiner
Apple	Apple	nominal head, noun, single word noun phrase
.	.	sentence boundary

Table 1. 2. Machinese Syntax analysis of the English sentence 'John gave Mary an apple'. (Pareja-Lora, 2012:14)

Text	**Base form**	**Syntactic relation**	**Syntax and morphology**
John	john	subj:>2	@SUBJ %NH N NOM SG
Gave	Give	main:>0	@+FMAINV %VA V PAST
Mary	Mary	dat:>2	@I-OBJ %NH N NOM SG
An	An	det:>5	@DN> %>N DET SG
Apple	Apple	obj:>2	@OBJ %NH N NOM SG
<s>	<s>		

In tables (1.1) and (1.2) above, there are two annotational models: (1) The Connexor's machinese Phrase tagger and (2) The Connexor's machinese syntactic annotation. The first model turns out the part of speech (POS) label for the item 'John' as (proper noun) which functions as nominal head and a single word NP, whereas the second model returns the fact that the item 'John' has a subject syntactic dependency on the contiguous lexical item and it functions as nominal head and supplemented the information with morphological properties of the words in the sentence. Supposing an artificial intelligent system (AIS) that will need the knowledge of the grammatical status of this word 'John' would be built, which one should the developer adopt?

In the same light, the problem occurs in Yorùbá scholarship too when scholars use different theoretical background or dialectological perspectives to analyse Yorùbá nouns. For example, some scholars analysed that if a qualifier noun begins with a consonant, there is a lengthening of the contiguous vowel into a mid-tone vowel which reflect possessive as observed in:

ilée Kọ́lá
ajáa Dàda (Bamgbose 1990, Yusuff 2006:3)

However, Awobuluyi (2013:4) says …

dípo
Iwé eKọ́lá
Ìlú uKánò
tí a máa ń wí nínú èdè Yorùbá àjùmọ̀lò, ohun tí ọ̀pọ̀
èniyàn ní ilẹ̀ Èkìtì àti Mọbà yóò wí ni:
ìwé iKọ́lạ́
ùlú iKánò …

(instead of
ìwé e Kọ́lá
ìlú u Kánò
that we use to say in standard Yoruba, what many people in Èkìtì and Mọbà will say is:
ìwé iKọ́lạ́
ùlú iKánò)

From this dialectological perspectives, Awobuluyi disagrees that any noun qualifier will always begin with an 'i' vowel insertion to have the general features of nouns which always begin with a mid tone 'i' as we have it in 'ibàtà' (sandal) versus 'bàtà' (sandal), 'iyàrá' (room) vs yàrá (room) etc. Not any vowel lengthening as claimed by other scholars.

Also, the item 'jẹ́ẹ́jẹ́' (gently) analysed as an adverbial modifier by Bamgbose (1990: 174) is regarded as a noun by Awobuluyi (2010 and 2013:5). The question arise again that, supposing an artificial intelligent system (AIS) that will need the knowledge of the grammatical status of these nouns would be built, which one should the developer adopt? These difficulties, in turn, lead to much wasted effort especially when there is need to reinvent the wheel. This problem is operationally solved by ontology development, whereby the variation are resolved by designing a hybrid, formal (machine readable) and explicit specification for the differing viewpoints. The consequential implication from these, however, is that the heterogeneous nature of language data brought the need for an ad-hoc study to extract the knowledge of these nouns and consequently begin to develop on the fly a model that can combine these different or overlapping viewpoints in an explicit (more human readable) and formal (machine readable) formats. The essences of such ontology-based annotations make Yorùbá nouns available to build or support automated reasoning activities in Artificial Intelligent System (AIS), Expert System (ES) and Natural Language Processing (NLP).

Based on the previously stated problems, the broad aim of this study is not just to develop an ontology based framework for annotating Yorùbá nouns but to implement a model that makes the corpus interoperable in semantic web and foster its reuse in automated applications. This work is a further exposition of our previous works on ontological annotation for Yoruba noun. It complements the earlier work based on Awobuluyi functional perspective of Yoruba noun. And, the specific objectives of this paper are to:

Present the ontology development processes involved in harmonising together the annotated designs of the knowledge in Bamgbose (1990) and Awobuluyi (1978 and 2013) of Yorùbá

noun, for the purpose of building a hybrid knowledge -based ontology.
Implement a hybrid semantic web ontology model called 'HyYORNO' (Hybrid Yoruba Noun Ontology) for the classes identified from these three knowledge resources.

This paper provides answers to the following questions:
- What are the standard techniques, principles and methods in ontology developments to harmonise together the annotated designs of Yoruba Scholars perspectives of Yoruba Noun and the extant issues relating to their extraction to build knowledge based model?
- How can we implement a hybrid model to solve the background inter-operability and shareability challenges which arise from the different viewpoints and perspectives found in Yoruba knowledge resources?

2. Review of Literature

2.1. Ontology/Ontological Annotations

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 explanation (usually short and coded) to a lexical item, text or a book. According to Pareja-Lora (2012: 32-34) citing Zauzich (1992:20), 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. Pareja-Lora (2012) summarises the basic terminology involved in annotations this way:
Each elements of the vocabulary used to annotate the contents of texts is usually referred to as **a tag**, **a label** or, simply an art of **annotation.** Accordingly, the whole vocabulary used to annotate somehow a set of texts referred to as **tagset** or its **metadata**, and the way in which the tags of a tagset can be combined in order to express content will be referred to as **annotation language**…

From this statement, it is clear 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. There are two basic methods to annotations in information technology: (i) the linguistic and (ii), the computational methods. The linguistic annotation on the first hand has morphosyntactic, syntactic and semantic in its component. Morphosyntactic annotation involves linguistic labelling that implements the morphosyntactic structure of a corpus, the syntactic annotations code structural relationships observable between phrases and phrasal items in a natural language be it dependency parsing or phrase structure parsing. and the semantic annotation labels the content of terms, objects and strings with stated properties that connote what we think of as its semantic content values. Terms, such as semantic annotations, semantic tagging, semantic mark up and semantic labelling have existed in literature to refer to extraction of meaning in corpus according to Ide, Romary and Ejavec (2006).

The focus of the computational annotation is the machine instead of human. Before machine can understand and process the annotations made, it must be encoded in the format, formal and clear vocabularies, syntax, schemes and tag form. Computational annotation languages such as Standardised Generalise Markup Language (SGML), Hypertext Markup Language (HTML) etc help to achieve these tasks. However,

The web or internet serves as a repository of huge information which contains documents which search engines like Google, Yahoo etc can use to access different information. As good and enormous these pieces of information are, the contents of this information become a hard task for machines to access. In other words, most of these information are human readable, and this is what brought the need to create machine readable contents. The aim to meet with this need is what Berners-Lee, Hendler, and Lassila (2001) pursued by making texts understandable for computers. This also gave birth to semantic web engineering, also known as ontology or ontological annotations. In the semantic web, not only humans will read the web's content, but computers' set of instructions will also manipulate web contents meaningfully. So, the Semantic Web is an extension of the existing one whereby information is well-defined semantically, better enabling human and computer to work together co-operatively and not another repository.

2.2. Ontology-Developed Applications

The availability of Ontology tools and reasoning systems has contributed increasingly to the widespread development of different applications which use or include Ontology in their developmental processes. Taye (2010) agrees that Ontology leads to the sharing of knowledge between systems and people. Ontology plays a key role in the semantic supporting information exchange across different distributed environments. It is an extension of the web in that data are represented in machine readable way. Beeners-Lee, Hendler and Lassila (2001) demonstrated that Ontology application focusing on e-business environment play a crucial role by locating the best match for a requester looking for merchandise or other information. In addition, it helps online travel customers to obtain a response. Mihonubilt (2000) ‘s Artificial Intelligence (AI) application is solely used in AI community where its main goal boils down on facilitating knowledge sharing and reuse between program, services, agents, across a domain. A good example is that developed by.

Fensel (2001) developed his Ontology to provide easy communications between agents and reduce misunderstandings between knowledge manager and electronic gadgets customers. SNOMED is a clinical terminology ontology database developed by the College of American Pathologist. It contains more than 400000 class names. The database contains standardized structured vocabularies which domain experts can use to share and annotate information. Gomes-Perez (1994, 1995) explored what is called Ontolingua. Ontolingua application was developed by the Stanford University as a Knowledge Sharing Effort (KSE). It is an Ontology server which was designed to enable a collaborative development of other top level ontologies. It is implemented in an on-line portable format for verification and evaluation of the knowledge sharing units. Gruber (1993) also had earlier developed ontolingua being a language of extensive library of ontologies which the definitions can be reused for other ontologies. For its axiomatic basis, the Ontolingua server could view, create, edit and publish ontologies while allowing forming network of protocol application program interface for other programs and applications utilization.

The Ontological Integration of Naïve Sources (Onions) was developed by Gangemi, Catenacci, Ciaramita, Lehmann (2006). The problem of integrating heterogeneous sources of information in knowledge acquisition and sharing led to the development of this application. The software fosters the development of a preliminary non-formal Ontology and a schematic account of the conceptualization of a domain. ONIONS implements knowledge across various naïve sources, using schematic diagram and tools which is a refinement of that knowledge. Baker, Brass, Bechhofer, Goble, Paton and Stevens (1998) developed Ontology which explores a wide scope of bio-informatics tasks and instruments, formally specifying them so as to enable shareability among the bio-medical community. It uses the A-Box and T-Box of the Description Logic (DL) to represent knowledge so as to make it flexible and make the classification of concepts consistent for the purpose of making inferences and logical reasoning: There are many more ontology developed applications like Ontobroker (Onbroker): Decker (1999). Transparent Access to Multiple Bio-informatics Information Source (TAMBIS): Stevens, Goble and Bechhofer (2001). The Ontology Inference Layer: (OIL) Ceccaroni (2001). IBROW: Fensel and Morita (1998), Generalized Ontology for Linguistic Description (GOLD), Farrar and Langedon (2003).

3. Materials

The materials employed to carry out the design of YORNO domain ontology in order to address the numbers two (2) of the research questions raised earlier.

Protégé

Protégé 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 customizable user interface one can create and edit ontologies in any compatible single workspace.

Web Ontology Language (OWL)

The Web Ontology Language (OWL) is developed as a vocabulary extension of the Resource Description Format (RDF) and RDF Schema, a language which is derived as a combination of DARPA Agent Markup Language (DAML) and Ontology Inference Language (OIL). These two languages are used in expressing inferences through axioms and rules. So, OWL contains declarative contents with formal syntax and semantics that can be parsed and used unambiguously in computer models. OWL is enriched in JAVA Programming language which have constructs that supported transfer of knowledge in ontology elements like classes, object properties and types by automatic generation of source code.

4. Data for the Study

The prose narratives from Bamgbose (1990) and Awobuluyi (1978, 2013) is extracted as the data for this research. The prose narratives are reduced to minimal token to form the knowledge of Yorùbá nouns. The identified properties of each entity in the prose is described as a prototype of the knowledge, these properties are then modelled as the artificial representation of the entities combined to make whole of the domain. These token 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 according to Hassan, Odejobi, Ogunfolakan and Adejuwon (2013:4). The IR can be graphical or 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 graphical since the objective of our formalism is to enable both human and machine readability. The specific relations are defined before the ontology is implemented As it is, this work cannot cover exhaustively all the issues and knowledge about Yorùbá grammatical classes from all scholars because those theoretical issues are enormous and may be too large to be implemented at a go for a start-up ontology of this kind, the progression is always from the simple to complex, we therefore limit the coverage of knowledge extractable for Yorùbá nouns only to these two scholars. These selected texts form the major input data for the study. This is justified for the fact that the materials are straightforward description of Yorùbá language grammar, precise and detailed enough with simplified linguistics rudiments in a way that can assist learners who seek to understand the core grammar of Yorùbá language.

5. Methodology of Ontology Annotations

Methodology involves analysing the principles and procedures of inquiry in a systematic way. Methontology is a coined language to describe the standard techniques in ontological annotations. Our previous works have used this method to design an intermediate tabular representation for the two outstanding thoughts in Yoruba scholarship. For the purpose of this hybrid implementation the following development processes were followed to achieve our goal in this paper

5.1. Activities in YORNO Development Process

Succinctly, ontology development process is based on the activities identified in the IEEE standard for software development (Schultz, 1997). This guidelines and standards were strictly followed during the implementation of YORNO. The specific activities endeavoured during the design and implementation of YORNO includes:

- Software life cycle model process

- Project management processes which include planning, controlling and quality management.
- Development oriented-processes include which consists of carrying out feasibility studies on possibility and the need to develop artificial intelligence systems for Yorùbá language manipulations. This requirement analysis which involves gathering, identifying and classifying entities, activities, agents and other key concepts on Yorùbá nouns as a starting point for information modelling and retrieval..
- Post-Development processes which include hosting and installation, inter-operation support, maintenance and optimization.
- Integral processes which include evaluation, documentation, configuration management and training of other NLP developers.

YORNO management activities include scheduling, control and quality assurance which checks the quality of each methodology output (ontology, software and documentation).The scheduling activities for YORNO commenced January 16th 2016. Thorough manual checks and scrutiny was carried out on the different nouns implemented in YORNO from: (1) Abrahams (1965), Fakinlede (2005) and (3) Modern Yoruba Dictionary being the first-hand available Dictionaries. The process of activities involved when building ontologies are explained in figures (1) and (2) below:

Figure 1. Ontology Development Process (Corcho, 2005) Dictionaries.

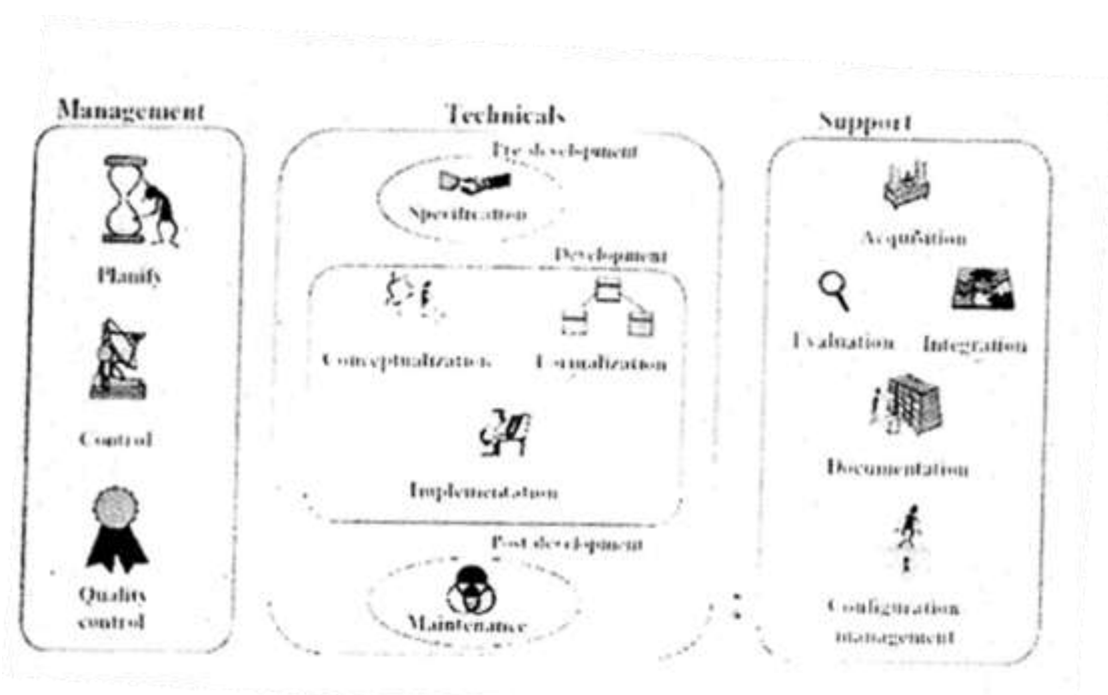
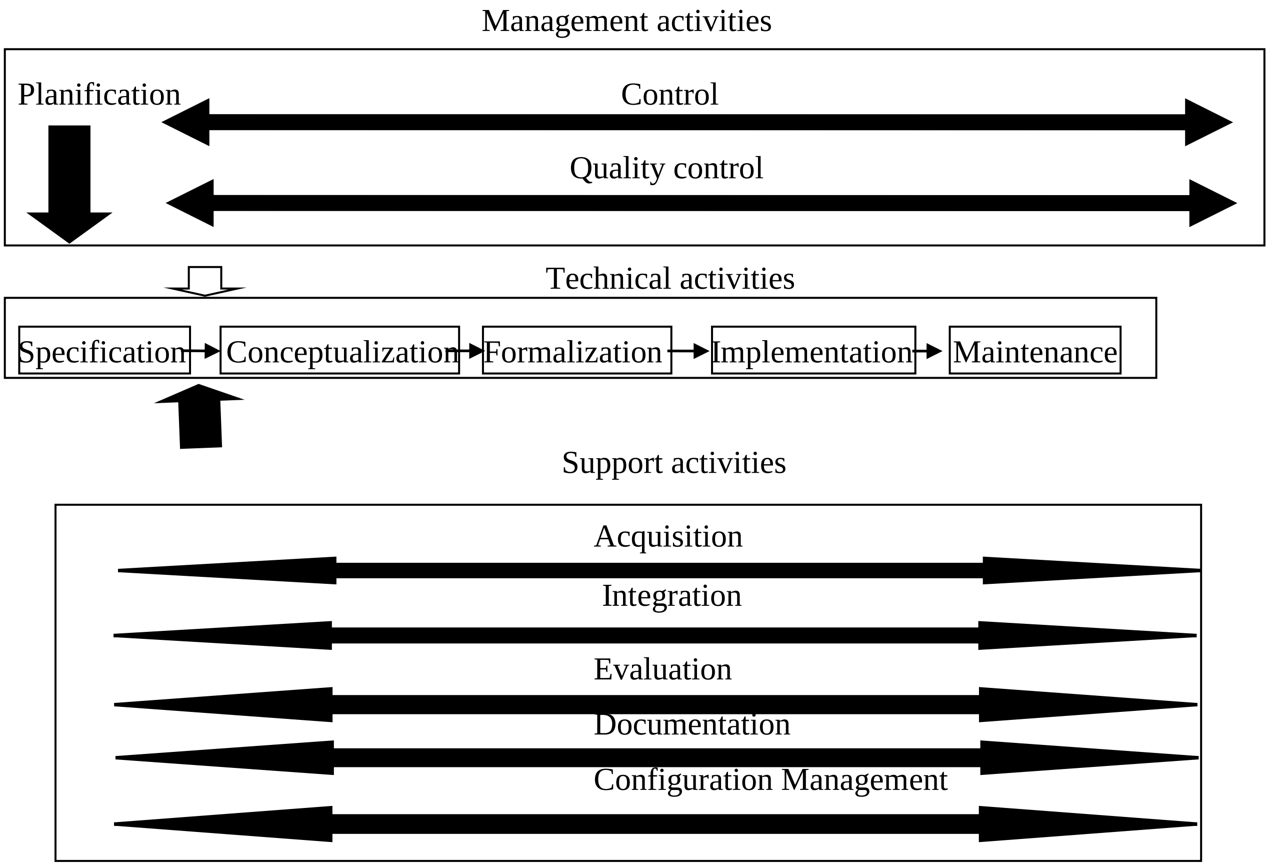


Figure 3.5. Ontology development process life cycle (Corcho, 2005)



5.2 Support Activities in YORNO Development Process

These activities are performed concurrently with the development-oriented activities. The objective of knowledge acquisition activity is to acquire knowledge form experts or by (semi) automatic ontology learning. Evaluation activity examines the developed ontologies, software and its documentation against a frame of reference. These important steps were carried out in three stages namely (1) Presentation of the extracted nouns to the Research supervisor, Professor Oye Taiwo of the Department of Linguistics, University of Ibadan. (2) Presentation of extracted knowledge as seminar at Ibadan Language Study Group (IBALSG), a research group of intellectuals and Professionals for thorough and intensive scrutiny of data. (3) A workshop on YORNO entry frame based on the lexico-semantic properties of the data implemented and the different classification of Yorùbá nouns according to the knowledge from the selected resources. (March 2018) Integration activity considers if other ontologies have been developed which can be possibly reused in conjunction with merging or alignment activities However, no ontology of this kind has ever been developed for Yoruba nouns. Merging produces a new ontology from the combination while alignment establishes mappings that preserve the original ontologies. Documentation details each completed stage and product and configuration management records ontologies, software and documentation versions in order to control changes.

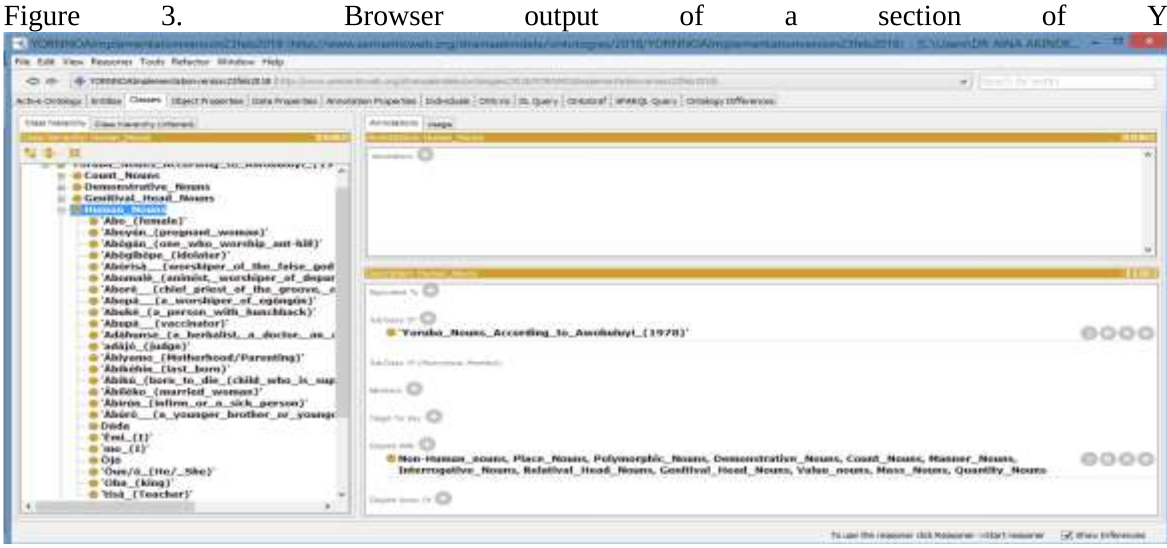
The activities in the management and support processes take place simultaneously with the development activities. The efforts applied to the support activities are not uniform along the life cycle. Knowledge acquisition, integration and evaluation are greater during ontology conceptualisation. The knowledge acquired at the beginning of the ontology development is integrated at the conceptual level before implementation and the conceptualisation are accurately evaluated in order to avoid propagation errors in line with Gruninger and Fox (1995)

6. Implementation of the Hybrid YORNO

Implementation is the process of turning the conceptual view of an idea into reality. It involves manipulation of research tools to build a system and also provides detailed description of the domain ontology. The implementation activity builds computable models using ontology implementation languages. This section is a major goal reported in this paper where we presents the implementation procedure of YORNO which combines the knowledge of Yorùbá nouns according to Awobulyi (1978 and 2013) and Bamgbose (1990). Protégé 4.5 was used to implement YORNO because of its expressiveness. We input the YORNO entry frame derived from the two workshops held for the ontology support process as reported earlier, into Protégé. This editor accepts the domain ontology and displays the hierarchical class structure from the parent class to all the classes and downward. Figure (3) shows how the various classes in the YORNOA relate to each other within the Protégé 4.5 alpha application. The protégé class hierarchy also shows the various sub-classes and relations that were added to the parent class in YORNO. For example in YORNO,

'Yorùbá Nouns according to Awobuluyi' is a Parent class, whereas
Common nouns, demonstrative nouns etc are the subclasses
Protégé visualizes the design in form of graph by showing all the internal structures such as root, ancestors, descendants, levels, siblings, cousins, branches etc. The graphical illustration of the part of the YORNO covers the sub-class "Common nouns" as shown in figure (3). Emuoyibofarhe, Lasisi and Oladosu while developing annotations for pharmaceutical concepts explains that semantic annotations improve web searches through access to web resources using their semantic description. (Emuoyibofarhe, Lasisi and Oladosu, 2007). In the same sense, YORNO was designed as a web application which can be integrated with other applications to provide access to information, its formal semantic flavour improves access to information stored on the web. The classification of the concept of Yorùbá nouns would facilitate effective manipulation of the data for natural language processors. An instance of the sub-class "Human nouns" was created with the slot as shown in Figure (3) while the output of Human nouns is displayed in the web browser as illustrated in Figure (4)

Figure 3. Browser output of a section of YORNO



The browser output is divided into three parts namely content, ontology and all resources. The content part shows the semantic link to the YORNO, the resources, classes, objects and datatype properties.

Various classes and attributes are displayed in all resources part. Ontology part displays how various classes are represented in the ontology. For instance, the sub-class “Human nouns” displayed in figure (4.1) is a sub-class of “Yorùbá nouns according to Awobuluyi (1978)” which in turn is a sub-class of entity “Thing”. The output reveals that the sub-class of “Non-Human nouns” which in turn is a sub-class of “Yorùbá nouns according to Awobuluyi (1978)” is a disjoint class to the sub-class displayed which implies that an individual cannot belong to these two sub-classes at the same time. Various instances (Individuals) of the sub-classes are also displayed. Each resource is semantically related to the other relevant resources. The designed classes, the properties, relations, instances as extracted from the prose narratives of the three texts are implemented to produce the model which a section of it as shown in figures (3). As mentioned, these activities enable a machine to recall any concept of Yorùbá nouns for processing based on the concept dictionary etc.

However, implementation activities for the designed model may not show all the incidence of mutual exclusivity and mutual inclusiveness at the background, but since our implementation is based on OWL syntax, all the background incidences are handled in OWL rdf syntax. Other complex representations that a manual curative method cannot logically be described have also been implemented simultaneously. Descriptive Logic (DL) forms the rules at which ontology languages like RDF, OWL are built upon, hence protégé which is logically describable within DL, is employed to implement the design fittingly. Some registered keywords below are examples of the complex, non-visual representation found in protégé:

- Equivalent with
- Members of
- Target of keys
- Disjoint with
- Sub class of
- Disjoint union with etc

The procedures we follow are outlined below:

- Start protégé. You will see the welcome to protégé dialing
- When the Welcome to Protege dialog box appears, press the `Create New OWL Ontology’ button.
- A `Create Ontology URI Wizard will appear'. Every ontology is named using a Unique Resource Identier (URI). Replace the default URI with http://www.yorno@awobuluyi and bamgbosei.com/ontologies/yorno.owl and press `Next'.
- I save the Ontology to a file on my PC, browsing to the hard disk to save the ontology to a new file, named `yorno.owl'. Once a file is chosen then one can press `Finish'.

From these procedure, the implementation of the Yorùbá noun ontology proceeds. For YORNOA, Thirteen classes were specified of which the data primer were 106 terms spread across the thirteen classes and the distribution is shown in table 2 below:

Table 2: Distribution of Classes in YORNOA

	Data Primer	**Dictionary entry**
Human Nouns	16	Many
Non-Human Nouns	12	Many
Value nouns	11	None
Quantity Nouns	8	8
Count Nouns	10	Many
Demonstrative Nouns	6	1
Genitival Head nouns	7	None
Interrogative Nouns	5	None

Manner nouns	18	1
Mass Nouns	9	Many
Place Nouns	2	Many
Polymorphic Nouns	6	None
Relatival head Nouns	5	Many

The following figures show some of the implementation of the elicited and designed concepts in YORNOA:

Figure 4 The class Hierarchy of (YORNOA)

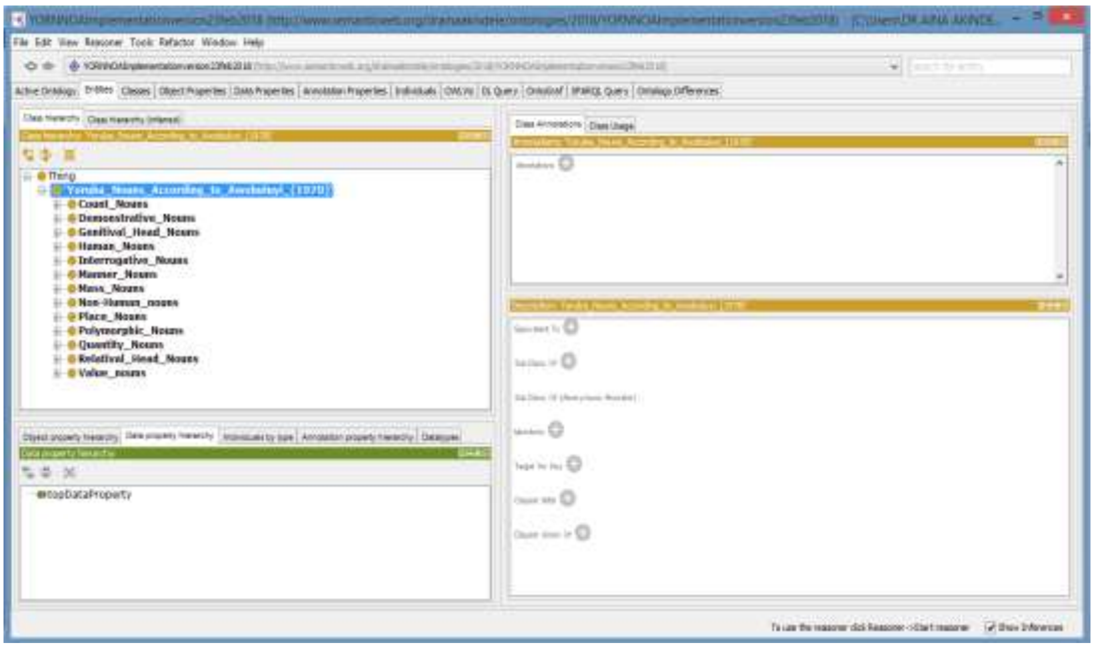
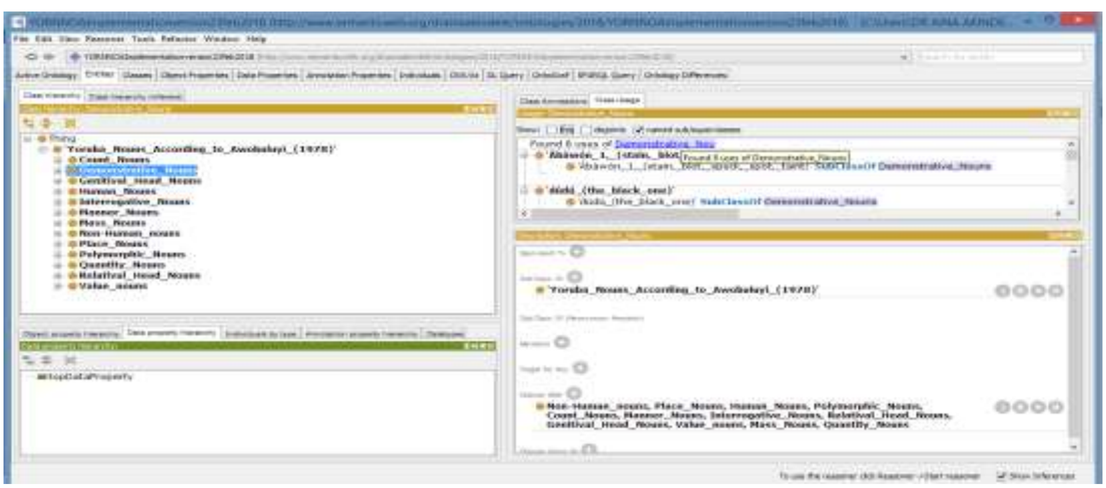


Figure 5 below shows the class hierarchy view, the annotation view and the description view of a subclass 'Demonstrative nouns' and its relationship with other classes:

Figure 5: The class Hierarchy, Annotation and the Description view of YORNOA



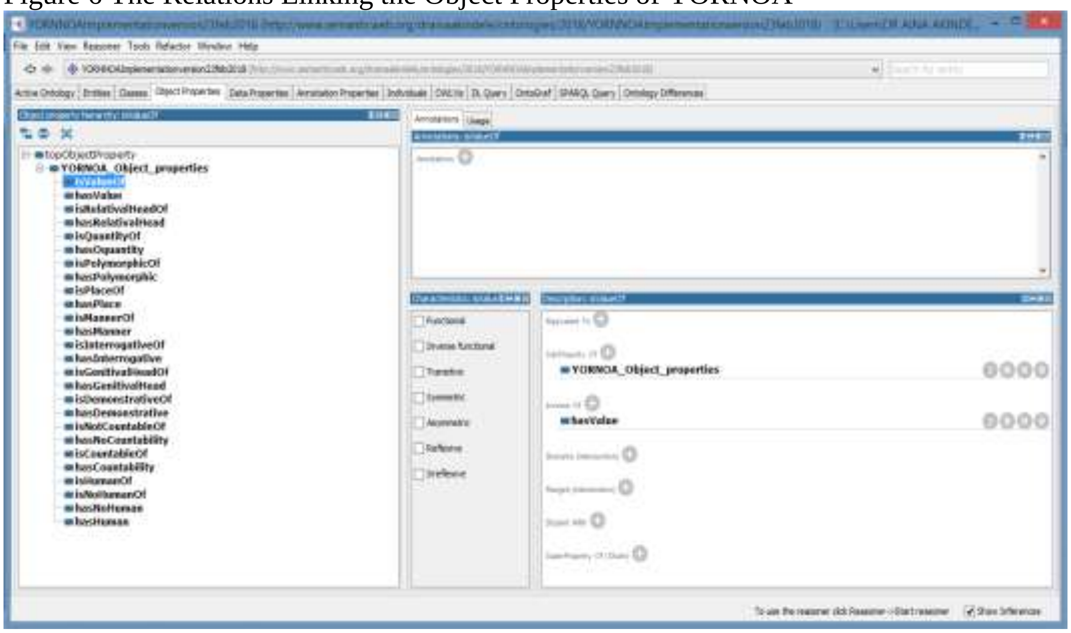
Apart from the hierarchy display at the hierarchy view pane, the annotation view pane shows all the instances of the subclass Demonstrative noun, while the Description view pane shows that Demonstrative noun is a subclass of Yorùbá nouns and is at disjoint with the other subclasses like Non-Human Nouns, Place Nouns and so on. That means an item cannot be a subclass of Demonstrative noun and be a subclass of other subclasses at the same time.

7. OWL Properties Implemented for YORNOA

OWL Properties represent relationships between classes. There are two main types of properties: Object properties and Datatype properties. Object properties are relationships between two individuals. Object properties link an individual to an individual. OWL also has a third type of property called Annotation properties. Annotation properties can be used to add information (metadata, that is data about data) to classes, individuals and object/datatype properties.

Properties may be created using the `Object Properties' tab in Protégé.  Figure 4.4. shows the buttons located in the top left hand corner of the `Object Properties' tab that are used for creating OWL properties. As can be seen from this illustration, there are buttons for creating Datatype properties, Object properties and Annotation properties. All properties created in this implementation are object properties. Figure 6 below shows examples of relations that links each type of property in the model.

Figure 6 The Relations Linking the Object Properties of YORNOA



8. Implementation of Yorùbá Noun Ontology According to Bamgbose (YORNOB)

The procedures as outlined in the previous section are followed in implementing YORNOB as well. From these procedures, the implementation of the Yorùbá noun ontology according to Bamgbose (1990) follows through. Twelve (12) classes were specified for YORNOB of which sixty one (61) terms were stored as the data primer which spread across the twelve classes and the distribution is shown in table 4.2 below:

Table 3: Distribution of Classes in YORNOB

	Data Primer	Dictionary entry
Ọ̀rọ̀ orúkọ Àrídìmu (Concrete Nouns)	04	Many
Ọ̀rọ̀ orúkọ Àfòyemọ̀ (Abstract Nouns)	04	Many
Ọ̀rọ̀ orúkọ Àṣéékà (Countable Nouns)	04	None
Ọ̀rọ̀ orúkọ Àlaiṣéékà (uncountable Nouns)	04	Many
Ọ̀rọ̀ orúkọ Ènìyàn (Human Nouns)	04	Many
Ọ̀rọ̀ orúkọ Ènìyànkọ́ (Non-Human Nouns)	04	Many
Ọ̀rọ̀ orúkọ ibìkan (Place Nouns)	05	Many
Ọ̀rọ̀ orúkọ ìgbà (Time/period Nouns)	05	None
Ọ̀rọ̀ orúkọ oǹkà (Quantity Numeral Nouns)	09	None
Ọ̀rọ̀ Orúkọ Òye (Value nouns)	03	1
Ọ̀rọ̀ Orúkọ Asọye (Quantifying nouns)	02	Many
Ọ̀rọ̀ orúkọ àṣafihàn (Demonstrative nouns)	05	Many
Ọ̀rọ̀ - orúkọ Aṣebèérè (Interrogative Nouns)	06	None

Figure 7 below show some of the implementation of the elicited and designed concepts in YORNOB:

Figure 7: The class Hierarchy of The Yorùbá Noun According to Bamgbose

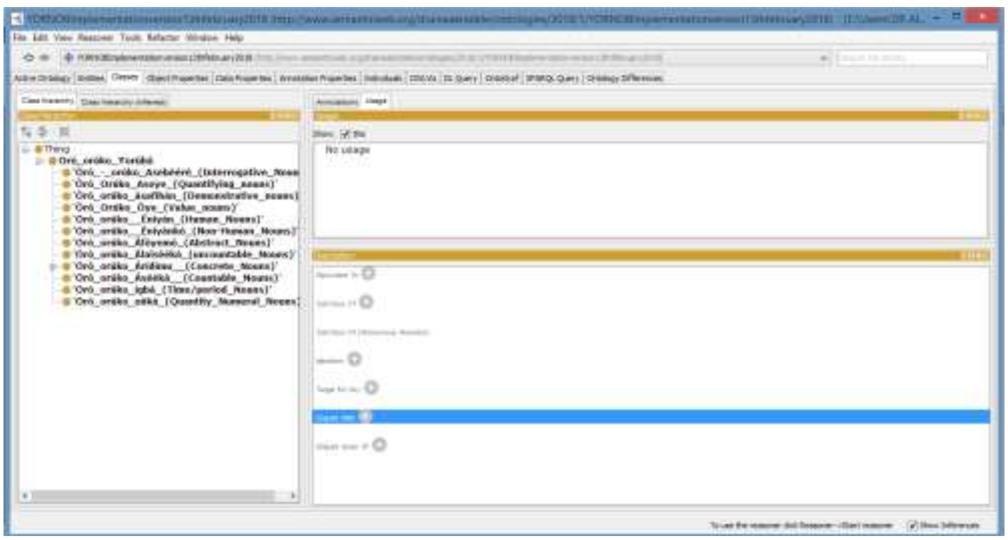


Figure 8 The Relations Linking the Object Properties of YORNOB

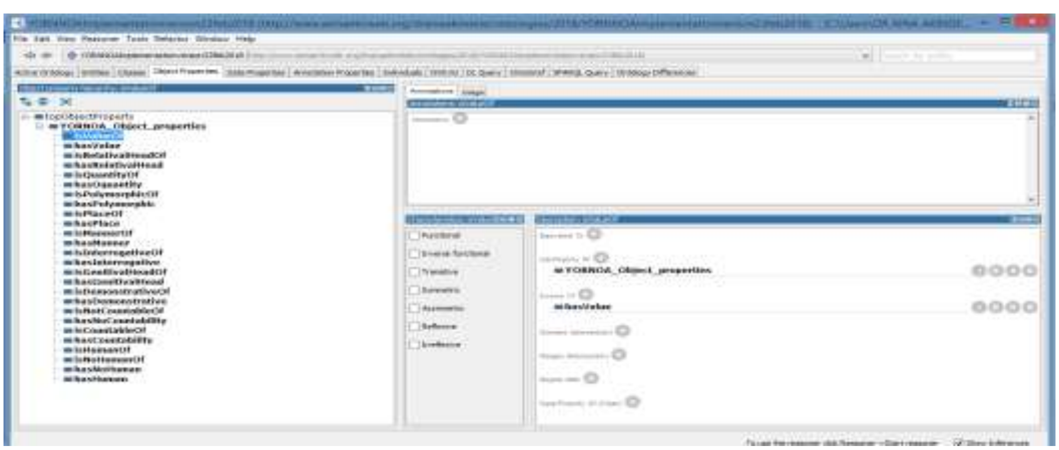


Figure 7: The view pane of the three hybridly nested model

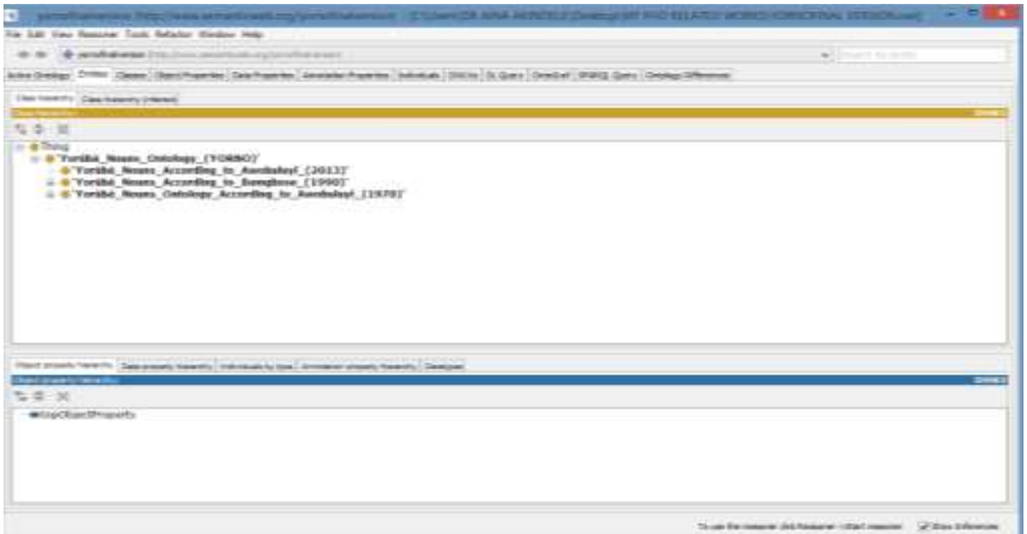
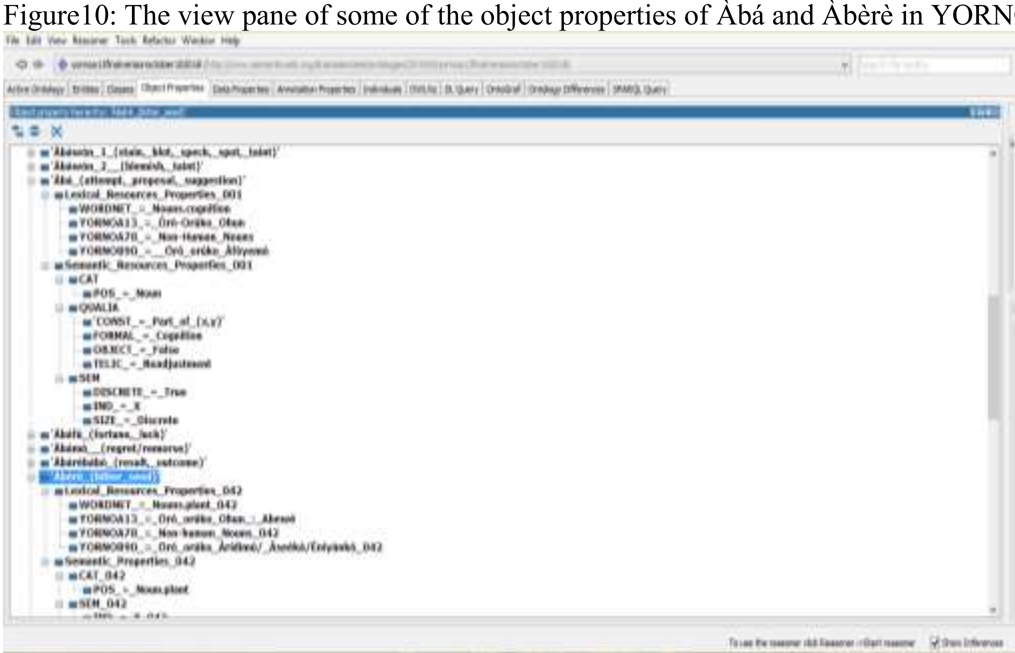


Figure10: The view pane of some of the object properties of Àbá and Àbèrè in YORNO



9. Launching of YORNO to Semantic Web

As a review, ontology annotations developed in this work serves these purposes amongst others. 1. Human readable documents that provides explanation about the concepts and their relationship in the domain of nouns. (2) A model which is machine readable which NLP systems and artificial intelligent systems can interoperate with and reuse. The editor we used for the implementation Protégé is web based, as much as ontology engineering is a semantic web based activities. Launching and publishing the ontology into the semantic web is an important complement.. This activity involves making the ontology an accessible resource in both human and machine readable format, with documentation and examples licensed and specified. Bernard (2013) highlighted the requirements for ontology documents as follows:

- The ontology must be available on the web with an open licence
- It must be machine-readable structured data (e.g. CycL instead of image scan of a table)

- It must have non-proprietary format (e.g., OBO instead of CycL)
- It must comply with the open standards from the W3C (RDF Schema and OWL)
- It must be reusable in other people's ontologies.

Apart from the requirements stated above, the following guidelines were provided in Bernard (2013) for an ontologist aiming at publishing on the web:

- The vocabulary must be published on the Web at a stable Unified Resource Identifier. (URI)
- Human-readable documentation and basic metadata such as creator, publisher, date of creation, last modification, version number must be provided.
- Labels and descriptions, if possible in several languages, to make vocabulary usable in multiple linguistic scopes must be provided.
- Make the vocabulary available via its namespace URI, both as a formal file and human readable documentation, using content negotiation.
- Link to other vocabularies by re-using elements rather than re-inventing.

In addition to the requirements and the guidelines stated above, Garijo (2013) furnished that the following procedure has to be strictly followed to load the ontology annotations into the semantic web:

i) Selection of the Name of the vocabulary/ontology. A unique name which the ontology will be reference to for its life time has to be carefully selected. I have named the product of this research, 'Yorùbá noun Ontology' (YORNO).

ii) Selection of The proper URI to Publish the Ontology. The next question that should be addressed after choosing the name is: Which URI do I choose? How do I ensure that it is not going to change? The URI one chooses for his ontology should be permanent and defined in a domain under one's control. This is necessary because of reusability. Imagine somebody is reusing the concepts defined in an ontology and the URI of that ontology changes from time to time The Ontologist reusing the model will no longer be accessible to the proper definitions and semantics of the reused term. This is why defining the URI of the vocabularies/ontologies in http://purl.org. is highly recommended. PURL means "Persistent Uniform Resource Locator". This command is widely used to give persistent URIs to resources. Once it is registered in the page, definition will be provided for a new domain, and it is necessary to wait for the approval from the admin of the domain host which will have to state the necessary condition for hosting the vocabulary before the operation is finally performed.

iii) Creation of the ontology in RDF/OWL. There are several editors to create vocabularies/ontologies and their properties according to the W3C standards. This research has chosen Protégé as the implementation tool. Since we have implemented the model with this editor, the next step is to change the base URI of the ontology (Ontology URI in Protégé) to the one registered as a PURL. Protégé will use a hash ("#") by default to identify the classes and properties declared in the ontology as well as pointing to the right spot in the documentation.

4) Redirection of the Permanent URI to the Ontology File. After the above procedure, the owl file have to be hosted somewhere. Where it is hosted is not as important as its safety so long that it won't be deleted abruptly. As for YORNO we proposed that the system admin stores the owl file in http://vocab.linkeddata.es/yorno/ - ontology1.1.owl. Further we go back to the purl page and add the basic redirection to the target URL that have been set up on the page, entering the URI of our ontology, it will be redirected to the OWL file loaded in protégé, and the job is done.

9.1 Derefencing of the Ontology

Dereferencing of ontology involves the process of preparing the ontology that makes it easier for locating or referring a particular concept amongst many in a vocabulary. Entering PURL into the web browser redirects the user to the html documentation of the ontology. If the same user has protégé on the operating system and enters the same URL in Protégé, the created ontology file will be loaded in the system. By dereferencing, it makes it possible for a user to choose the specific items he wants to reuse. Depending on the type of request received by the server in the form of an RDF files. The interface returns the html pages only for those people looking for mere information about the ontology and loads the resource which is the RDF files for structured content to provide machine readability. In additions, when a certain concept is sought for, the browser takes the user to the exact part of the document defining it. For example, I want to know the exact definition for the concept “agbandan” in the YORNO, then you can paste its URI (http://purl.org/net/yorno#Agbandan) in the web browser. The user can detach and use that aspect simply because it has been dereferenced. .This makes life easier for users when reusing the created ontology.

The following outlined tasks should be taken to dereference the vocabulary according to Bernard (2013), Berners-Lee (2009). First, set the purl redirection as a redirection for Semantic Web resources (add a 303 redirection instead of 302, and add the target URL where you plan to do the redirection. A reminder is important at this point that one can only dereference a resource, only if one is in control of the server where the resources are going to be delivered. However our proposal recommends that the resultant System Admin stores the ontology in http://vocab.linkeddata.es/yorno. However, it is important to adhere more 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. Also, one will have to create an htaccess file for the link. For YORNO, 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 yorno.content. In order to avoid an infinite loop when dealing with the redirections of the vocabulary these set of instruction are employed

10. Summary and Conclusion .

People, organisations and individuals in Yorùbá society must communicate between and among them and make use of software systems to make their job easier. This situation directs people attentions to building artificial intelligent machines to complement human efforts. The core grammar of Yorùbá language which those artificial intelligent agents must be built upon then must be brought to the fore. This research therefore has addressed these problems by developing a conceptual model to address the conceptual and terminological confusion relating to Yorùbá nouns so that we can come to a shared understanding. Such understanding functions as a unifying framework for the different viewpoints and also serves as the basis for communication between people with different needs and viewpoints. human and machines, then machine to machine. The product of the research at the end serves as a feeder to other NLP activities.

This work is also a step towards bridging the gap of development between Yorùbá language and other developed European languages. It is an exposition and demonstration of the fact that our grammar can be prepared for computers to process. The ontological annotations done for Yorùbá nouns in this work solve the problems of ambiguities for machine learning. The annotated nouns in this work is useful for shareabity, interoperability and reuse in knowledge based systems and artificial intelligent systems. Researches in this direction produces modern electronics applications which make use of the machine readable model developed in this work developed and modernized for national economic growth and development.

References

Abraham, R.C. 1965. *Modern Yorùbá Dictionary*. Oxford University Press. London.

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.

(2008). Indigenizing Human Language Technology for National Development. A Lecture delivered as First ARCIS Distinguished Guest Lecture, U.S.A.

(2009). Building Capacities in Human Language Technology for African Languages. EACL 2009 Workshop on Language Technologies for African Languages–AFLaT. 53–58. Athens: Association for Computational Linguistics.

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.

Aina, A. A. (2017), Modernization and Globalization of Yoruba Language through Ontology Development- A Preliminary Investigation in Current Studies in Yoruba Culture, Language and Literature. *A Festschrift for Olugboyega Alaba.* 816-834,

Andre, E and Rist, T. (1993), The design of illustrated documents as a Planning Task in M. T. Maybury ed., 'Intelligent Multimedia Interfaces', AAAI Press, The MIT Press, Menlo Park (CA) Cambridge London (England), pp. 94-116.

Awobuluyi, O. (1978). *Essentials of Yorùbá Grammar*: Ibadan: UPL.

Awobuluyi, O. 2008. *Èkọ́ Ìṣẹ́dá Ọ̀rọ̀ Yorùbá.* Akure: Montem Paperbacks.

Awobuluyi, O. 2011. Linguistics *and Nation Building.* Ibadan.

Awobuluyi, O. (2013). *Ẹ̀kọ́ Gírámà Yorùbá*: Òṣogbo: ATMAN Limited.

Bamgbose A. (1965). *A Grammar of Yorùbá*: Cambridge University Press

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

Corcho O. and A. Gómez-Pérez (2001), "Solving Integration Problems of E-Commerce Standards and Initiatives through Ontological Mappings" in Proceedings of IJCAI 2001 Workshop on Business & the Intelligent Web, Seattle, USA

Eludiora, S.I. (2012). Development of an English to Yorùbá machine translation. PhD Thesis Dept. of Computer Science and Engineering. O.A.U. Ife.xi+206pp.

Emuoyibofarhe, N., Adebayo, S., Lala, G.O., and Aremu, R.O., (2015), A Yoruba Cultural Tradition Repository Knowledge Based System. International *Journal of Emerging Trends in Science and Technology* Vol. 2 issues 7: 2830-2841.

Gómez-Pérez, A., Fernández-López, M., Corcho, O. (2004) Ontological Engineering: With Examples from the Areas of Knowledge Management, E-Commerce and the Semantic Web. London: Springer-Verlag.

Guarino, N., Oberle, D. and Staab, S. (2009). 'What is an Ontology? Formal Ontology in Information Systems'. *Proceedings of FOIS 1998*. Trento Italy.

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

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

Hassan, J.A., Odejobi, O.A., Ogunfolakan, B.A. and Adejuwon, A, (2013). Ontology Engineering in Yorùbá cultural Heritage Domain. *African Journal of Comp. & ICTs* IEEE. 6.5:181-198. Retrieved online Oct. 20, 2015 from http/www.ajocict.

Hutchins, W. (1999). The Historical Development of Machine Translation, Submission for the Degree of Doctor of Philosophy by Publication at the University of East Anglia url:http://www.hutchinsweb.me.uk/PhD-1999.pdf, (Accessed: 23/10/2011).

Iyanda, A. R. 2014. Design and Implementation of a Grapheme-to-Phoneme Conversion System for Yorùbá Text-to-Speech Synthesis. PhD Thesis. Dept. of Computer science. Obafemi Awolowo University. Ile Ife, Osun State, Nigeria. X+186pp.

Johnson, S (1921), *The History of Yorùbás.* Great Britain: Love & Brydone Printers Limited, Thetford Norfolk.

Lawson, T. (2004), A Conception of Ontology. mineo: Cambridge.

McEnery A. M., Wilson, A. (2001) Corpus Linguistics: An Introduction. Edinburgh: Edinburgh University Press.

Noy, F. I. & D. L.McGuiness, 2000. Ontology Development 101: A guide to Creating your First Ontology. Standard Knowledge System Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SM1 – 2001 – 0880.

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.

Odoje, C. O. (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 (2012). ONTOTAG- A linguistic and ontological Model Suitable for the Semantic Web. A PhD Dissertation, Facultad De Informatica, Universidad Politecnica De Madrid.

Taiwo, O. (2006a), *Mofoloji*. Ibadan: Layof Ventures Publishing Educational Service.

Taiwo, O. (2006b), The Indivisibility of {ai} in Standard Yorùbá. *The Nigerian Linguists Festschrift Series* 6:557-563.

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

Uschold, M., King, M. (1995). Towards a Methodology for Building Ontologies. IJCAI'95 Workshop on Basic Ontological Issues in Knowledge Sharing Ed. D. Skuce 6.1-6.10.

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