ISSUES IN KNOWLEDGE REPRESENTATION IN YORÙBÁ NARRATIVES

Alade, S. M.; Ninan, O. D.; Odejobi, O. A.

**Department of Computer Science and Engineering,
Obafemi Awolowo University Ile-Ife, Osun State, Nigeria**

Corresponding Author: Alade, S. M.; samshola@gmail.com

ABSTRACT: *Developing a logically and consistent knowledge representation scheme for narratives across multiple cultural domains raises the question of the components and structure of a framework within which African narratives, particularly Yorùbá narratives can be formalized or standardized. It is a common fact that narratives are often influenced by culture, language and cognition. Thus, our experience working with the representation of narratives texts are challenged by the current state in the domain of knowledge engineering. The aim of this study is to propose a method for the formalizing and standardizing the representation of Yorùbá folktale narrative text in order to make the information content and structure contained in the folktale narrative explicit and shareable. Moreover, two computational approaches known as Ontological-based representation and text-based representation were put forward for the representation of text information. This research is very important with the purpose of analyzing the content and structure of Yorùbá folktale narrative text.*
KEYWORDS: *Knowledge representation, Narratives, formalization, Computation, Yorùbá, Folktale.*

1. INTRODUCTION

This research work addresses some of the issues relating to how narratives are represented, particularly Yorùbá folktale narrative texts. The major challenge encountered in this ongoing research work on the development of a workbench for African folktales is the accurate and precise representation of the events and concepts contained or embedded in the narratives. This often relates to how narrative text can be codified, formalized, standardized and represented in a computer form.
([TMS06]) described narratives as the primary way in which human communicates with one another. Hence, it is a simple basic and constant form of human expression irrespective of ethnic origin, primary language and enculturation ([Cha80]). ([FWR10]) also corroborated that linguistic representations are used to communicate, convince, explain encourage and entertain. In addition, ([***17]) defined narrative as an account of events, facts given in order with the establishment of connections between them. This is in consonance with ([Zar12]) who described as a generic term that denotes a logical and temporally ordered sequence of elementary events.
Moreso, narrative contains the content and structure of events, that is, consists of events which describe the specific behaviour (actions and process), experience (a situation, state), temporal and spatially constrained that characterizes some group of entities. The narrative is categorized into four (4) forms namely: Informal, Somewhat formal, and Traditional form ([FWR10]).
Folktale, like other genres or forms of narratives, is defined by different scholars. ([Abr05]) defined folktale as a short narrative of the unknown author but in actual sense it is communal. Also, the folktale is often referred to as household tales or fairy tales. In addition, he stated that folktales are sayings, verbal compositions and social rituals which was handed down from generation to generation by words of mouth. They were never written down but were only present in one's memory in the sense that it is embodied within an individual (implicit). They tell stories, whether real or imaginary to teach children about the intricacies of life. Thus, they are a representation of the aspect of the aspect of the human experience of an event or story.
Today, there exist several types of folktale based on their different culture and origin.
Amongst the several types of folktale include the western folktales (Russian, American, English, etc.); Asian (Japan, Korea, China); African (*Yorùbá*, Igbo, Ghana, etc.). Africa has a long history of oral tradition, especially folktales which are inevitable and has been an effective way of conveying information, traditions, knowledge and morals of culture. Folktale has been one of the oral activities used a long time ago in ancient society, which is an aspect of the culture and tradition practiced. Indeed, storytelling has played a fundamental role in the development and structure of human society as well as the behaviour of individual people in society with varying background around the world. This is because the culture is knowledge and the folktale contributes to culture as an artefact, which reflects and contains cultural knowledge ([Fin12]).

In the context of a narrative, a standard *Yorùbá* folktale narrative is one that is written in standard *Yorùbá* (SY) language. It is a folktale narrative whose texts are written by using diacritical marks, including tone marks above vowels and sub-pointss under certain letters. They serve as pedagogical tools and function as problem-solving and entertainment purposes. Besides, folktales add value to life and teach morals which help in decision making thus stimulating and raising the consciousness of the society to change behaviour by endangering specific socio-cultural attitudes ([Ola13]).

The text of narratives is an abstract representation of speech that can be in a written or digitized form ([Ode05]). Narrative text is definite in that it typically consists of words, a complete piece of writing or speech which follow the conventions and rules that determine the system of communication and structure used in particular text types ([CC01]). In addition, the people of Yorùbá have a unique culture and language and therefore a distinct view of the world. Therefore, the representation of Yorùbá narrative text from Yorùbá origin cannot be examined or discussed in separation from the cultural environment from which it emanates. The fact that texts are abstract entities indicates that they are a powerful tool in teaching and acquiring the knowledge about the concept embedded in the language of the narrative.

In other words, folktale narrative texts serve as an important instrument for entertainment, documentation, education and knowledge transfer ([Owo05]).

2. KNOWLEDGE

Knowledge is not an easy concept to describe. Although there have been several definitions, only a few describe the concept in sufficient details to capture it in practical terms ([CB12]). Likewise, ([Cha10]) opined that knowledge is a description of the world. According to ([NO13]), knowledge is concerned about the information that can be used or implemented, that is, it is information that has been expressed within a suitable context, meaningful and actionable in a particular domain since it evolves over time with experience which puts connection among situations and events in context.

2.1 Kinds of knowledge

There are several forms of knowledge ranging from indigenous knowledge, prior and posterior knowledge, procedural and declarative knowledge ([RB10)]. The declarative knowledge (concepts, facts, object) are the knowledge that describes what is known about a problem but the use which is not given, while procedural knowledge rules, procedures that describe how a problem is addressed in such a way that it uses the knowledge and information inherent and embedded in the narrative domain. Others are heuristics and meta-knowledge.

([DSS93)] observed that there are several categories of knowledge classified in terms of the way in which they are implemented in practice. These forms of knowledge are implicit and explicit knowledge. The implicit type of information with meaning is the knowledge contained in the experience of the individual. This has essential characteristics information relating to an individual or person making it difficult to formalize and convey its meaning, thought or exchange the inherent information ([NT95]).

Explicit knowledge is definite form of knowledge refers to the knowledge that is communicated formally and systematically using the abstract form of words, symbols, signs, gestures since it can be effectively transmitted and freely exchanged without any problem. It is independent of the individual's mind ([NT95]). In addition, ([GR05)] further clarified that explicit knowledge can also be categorised as structured (such as those that are well organised e.g .those in data repository), semi-structured (such as tags, XML, JSON, etc.) and Those unconventionally arranged or unstructured such folktale texts. However, the way to extract such information from a narrative text and its limitation are crucial to its representation.

2.2 Narrative representation

[***17] observed that representation helps people to organize the world and reality through the process of naming its elements. The narrative text representation has played a major role in the understanding of literary genres since the times of Aristotle, Plato, and other Greek philosophers. Literature has shown that there are several ways by which narrative texts can be represented. ([Cho59]) observed that grammar is a mechanism for describing the language has been used for representing the basic unit of the internal structure of folktale and the relationship among them. These include Rumelhart, Thorndyke, Mandler and Johnson story grammar ([Lee93]).

Furthermore, ([Nin15]) in her thesis explored the use of these methods but they were found to be inadequate for representing the content and structure of Yorùbá folktales. Similarly, Finite State Machine (FSM) was also explored in order to represent the Yorùbá folktale. As a result of the application of FSM in the modelling of discrete and concurrent systems, ([Nin15]) investigated how Finite-state Machine (FSM) could be used to represent the

narrative information in YFT because of the set of states involved and the transition. In the study, she observed that FSM is restricted in their capability since it was unable to handle the issue of simultaneous execution of processes or state involved in the manipulation of the Yorùbà folktale text. This representation expressed overcame the challenges posed by FSM because of the provision of dynamic and concurrency of the atomic proposition or states or process. Despite all the representational system presented, none of them has adequately represented the content and structure of Yorùbá narratives.

3. CHALLENGES OF KNOWLEDGE REPRESENTATION IN YORÙBÁ NARRATIVES

The Yorùbá folktale narrative texts (YFT) have incessantly become stories of language embarrassment to the readers because of its low and scarce resource available due to the involvement and support of the natives for unmindful practices which directly undermine the representation of this folktale narrative as cultural traditions.

Also, Yorùb*á* stories have continually poised a formidable representational problem to the modern reader and users. The narrative is without doubt driven with a linguistic perspective and cultural orientation which are completely an embarrassment to the cultural and moral sensitivities of the modern Yorùbá world. Due to the significance of YFT as a well revered cultural tradition and educational symbols, there is a pursuit to quickly examine these representational problems in the Yorùbá folktale narrative and in turn the annotation tool.

Representation has always being a problem in Artificial intelligence ([Cho59]) and in the narrative domain since the narrative system is aided by knowledge of the subject of the discourse. This is because it plays a major role in building an intelligent system. Fundamental to this is the representation of narrative text in the context of *Yorùbá* folktale narrative text is the reasoning. This is because it is not possible to represent a narrative without reasoning. hence, logic is fundamental to reasoning. The concern of representing narrative as it relates to problem-solving is the logic of discourse, which is the logic of intellectual engagement with different languages. The logic of discourse in the English language is different from the *Yorùbá* language. For instance, given two languages L_1 and L_2 (as English and *Yorùbá* language respectively). If a system for representing a language is developed around the logic of discourse of L_1, it will not be able to accept the language L_2. As a result of this inadequacies and inability of the system to represent the logic of discourse of other languages. Therefore, the need to address the problem and its solution processes in the context of *Yorùbá* world view emerges in the context of automated story analysis and synthesis.

Consequently, based on the language of discourse, another problem to be noted is that of the orthography. Orthography is a system of representing a language in a written form. It gives the standardized way of using a set of symbols to write a language. That is to say, it describes the symbol, graphene and diacritics used in a language which could cause loss of information or representational mistakes. Orthography can be transparent or shallow in the sense that the pronunciation can be predicted from the spelling (e.g. Italian, *Yorùbá*) or opaque For instance, suppose we want to represent narrative text *Ìjàpá àti Òjóla jo*. *ńsóre, o*. *n* and the existing representation systems do not show the tones in terms of its diacritics and hence is not described in the way it should.

A major characteristic of alphabetic writing systems is that letters can be mapped directly with phonemes except in the case of Yoruba where the same phoneme produces different graphemes: text o. n and *a* respectively. Again, writing is uncommon in that they are not written down. A significant number of languages ever spoken were deficient in writing systems. People create small, expedient collections of representation that can be used to express ideas and information each time there is a compelling need to identify distinct objects or quantities, and for thousands of years have been definitely been doing so. Hence, writing is a system of tranlating any and all expressions of a particular language requiring the use of signs, symbols, that represent the speech sounds instead of direct expression of objects or ideas.

Another issue encountered in the representation of *Yorùbá* narrative text is the grammatical structure of the *Yorùbá* language. The grammar here is the system or rule describing the structure of a language. It includes everything about the structure of the *Yorùbá* language such as phonology, morphology, syntax and semantics. ([Bam67]) stated that the first grammar of *Yorùbá* language was written by Ajayi Crowther in 1852, which was referred to traditional grammar. There are differences in the structures of languages where some are written in alphabetic letters (e.g. Armenia, Amharic, Georgian, Yorùbá, etc.), syllable (e.g. Devanagari or Nagari), or character (e.g. Chinese).

The characteristic feature of the grammar is the analysis of language in terms of the structure of another language, and the use of meaning. However, the use of other grammar in the representation of

Yorùbá narrative poses a serious challenge in the annotation of the language text. For instance, there are eight classes of speech (8 core tags) which are majorly defined for European languages but cannot be said of Yorùbá language. These classes of speech identified in Yorùbá narratives include Nominal, Verbal group, Adverbial, conjunction, and preposition ([Bam67]).

Besides, semantics is another issue encountered in the representation of narrative text. Semantics refers to the meaning of the *Yorùbá* word or text contained in the sentence. In Yorùbá language, there are some concepts or terms that make the Yorùbá narrative distinct. ([NO13)] asserted that these concepts and terms act as support for the understanding of the Yorùbá narratives through space and time. Such ideas are expressed in words, as shown in Yorùbá narratives. Examples of such concept and terms include days, months, seasons, number, time, relations and many others which are deeply rooted in the cosmology of the language ([App92]). Most times, the terms and words are often misunderstood, particularly when translating with English equivalent. Moreover, within the cosmology of Yorùbá the meaning of most words for communicating relationship in the narrative of Yorùbá is only rational, logical and permissible. For example, the term íyàwó (wife) is used to refer to a woman married to an ebí (family). Contrary to a woman's English definition of a woman who is legally married to a man. This is the reason why it is rare, and logically correct in terms of the condition in which the word is used another woman, as 'my wife'. Moreso, in a Yorùba context, where a woman refers to another woman as 'my wife'. So, the construction of an annotation tool for the representation of the structure and content of *Yorùbá* narratives would make it possible to express and adequately represent these concepts.

Moreover, the inability to capture relevant knowledge needed for building a story workbench tool from a domain expert and text document is another challenge. This is due to the diverse kinds of information or knowledge such as implicit and explicit knowledge embedded in Yorùbá narratives. The acquisition of relevant and known information about the textual data are implicit and domain that is, they are embodied within people and artefacts ([RB10]). Again, the inability to capture or extract relevant knowledge may be as a result of lack of resources used in propagating, enhancing and representing the textual data in an electronic form.

Heterogeneity of textual data is another challenge faced during the representation of narrative text information. Here, it means when information sources and store correspond to different media. This may be as a result of technological differences in hardware or operating system. The heterogeneity may be syntactic (difference in the machine-readable aspect of data representation, and storage for digital media) or semantic heterogeneity where meaning is dependent on vocabulary and terms used in expressing information ([KS06]).

The integration of heterogeneous textual data from Yorùbá narrative in an attempt to provide relevant knowledge is another great challenge to the development of annotation tool for Yorùbá folktale narratives. Another major challenge to the creation of annotation tool for Yorùbá folktale narratives is the heterogeneous textual data from the Yorùbá narrative and its incorporation in order to provide relevant information. Nonetheless, in the field of textual data, the question of arranging information from more than one origin in an easily accessible structure is challenging for fields of research such as Database, AI and Information retrieval. This is because the data collected could be from known and unknown sources even when the sources of relevant textual data are individualized.

Furthermore, granularity is also a major challenge to representation. Granularity is described as the extent to which textual information or knowledge can be represented. It could be coarse or fine grained. With regards to computing, it describes the amount of computation in relation to communication. However in the context of representation of textual data, the level to which textual data are represented is determined based on the term – document or document to document similarity ([JL09]). Others include the challenge of extracting the knowledge from unstructured data particularly, the Yorùbá folktales which require techniques from several areas of computer science such as data mining, machine learning, natural language processing, analysis of information and knowledge management ([FS07]).

4. RELATION TO COMPUTATION AND ARTIFICIAL INTELLIGENCE

Having discussed the issues affecting the representation of Yorùbá narratives such as orthography, grammatical structure, heterogeneity of data and others. Orthography being described as the system of writing such that it is the sum total of all symbols used in language for writing as well as rules guiding word formation and spelling. It follows that the sequence of standardizing the symbols for its writing to represent linguistic utterance requires a representational system whose task involves the processing (encoding and decoding) of information contained in the narrative with important reference to language.

Language as an abstract collection of words, definitions and symbols for all aspects of culture including voice, writing, numeral, signs, gestures and non-verbal communications, can be regarded as computation because it involves creation and manipulation of symbols. This, therefore, focuses on computation since the processing of text will involve the creation and manipulation of symbols (numeric, literal and /or linguistic) relevant to the spatial and temporal description of narrative. The text symbols are then manipulated using a mechanism or device. So, an important task in the computational representation of a system is the description of its mechanism. This will facilitate the creation of necessary computing instrument for implementing the task of representing narrative texts.

This provides a general relationship as it coincides with all processes relevant to the study of texts in natural languages (natural language, which is a computer imitation of a human language) and their comprehension and integration with human language. Natural Language Processing is closely related to other computer linguistics fields, where computational techniques and other approaches are employed to describe the problems of language and phonetics, and in turn Artificial Intelligence (AI).

Since Yorùbá narrative serves a pedagogical and problem-solving purpose in the sense that it involves manipulation of symbols and symbolic structure which facilitates the creation of a device or tool to solve the problem by representing the text. Hence, this illustration gives proof of validity within a branch of science that involves the development and resolution of problems that are primariy concern ed with the implementation of intelligent human beh aviors such as information presentation, teaching, reading, communication, preparing, explaining, language acquisition and so on in machines in order to learn and solve the problem.

5. POSSIBLE SOLUTION STRATEGY

A lot of work has been done in the area of computational modelling, Information Extraction for folktale narrative, as well as in the area of computational linguistics, and good automated results are obtained for computations. In an attempt to address the issue of representing Yorùbá narratives text in a way that is understandable by human and machine, ([FFF99)] opined that knowledge representation which employs symbolic language or a combination of approaches namely logic and ([Ami04]). Similarly, ([D+17]) provided a basis for formal representation of information embedded in the narrative which could be used as a means to construct a dictionary of concept contained in the narrative but could not give the additional information about the Yorùbá text.

Meanwhile, ([D+17]) proposed two approaches to the representation of text information known as Ontological based representation and text-based representation. The ontology base representations are detailed and comprehensive and can be removed from any direct application with a view of providing a formal representation of domain or sub-domain knowledge. Ontologies have in recent times become the main medium of representing information especially for several tasks in the domain of computer science such as bio-informatics, Agents and semantic web ([BO07]). It provides a basis for formal representation of information embedded in the narrative which could be used as a means to construct a dictionary of concept contained in the narrative but could not give the additional linguistic and interpretive information about the Yorùb*á* text such as part-of-speech, syntactic structure, event, time expression and so on. However, text-based representation is well suited since they are conceptualized as the process of enriching narrative text corpus inserted by human or machine or the combination ([PM17]).

6. CONCLUSIONS

This study contributes to the body of knowledge by establishing the fact that there are different forms of knowledge embedded in Yorùbá folktale narrative and that the narrative texts can be represented effectively both in content and structure using a computational approach known as annotation which will serve useful application in human language technology or engineering, development of annotation scheme for Yorùbá texts as well as digital processing of standard Yorùbá text.

Text annotation is the process of encoding and adding explicitly more interpretive information to the machine-readable text ([Lee93]). The use of annotation method makes it simple and faster to read and analyze the content and structure of the Yorùbá narrative corpus.

REFERENCES

[Abr05] **Abrams M. H.** – *A Glossary of Literary Terms*, Eight Edition, USA: Thomson Wadsworth, 2005.

[Ami04] **Amir M. S**. – *Knowledge Representation within Information Systems in Manufacturing Environments*. PhD Thesis of the Department of Information Systems, 2004.

[App92] **Appiah K. A.** – *In my Fathers House, Africa in the Philosophy of Culture*, New York, Oxford, 1992.

[Bam67] **Bamgbose A. A.** – *Short grammar of Yoruba Ibadan*. Heinemann Educational Books (Nig.) Ltd, pp. 1-63. 1967.

[Bos14] **Boshoff N.** – *Types of Knowledge in Science based practices*, Journal of Science and Science Communication 2014.

[BO07] **Brewster C., O'hara K.** – *Knowledge representation with Ontologies: Present Challenges - Future possibilities*. International Journal of Human-Computer studies Elsevier, 65:563-568, 2007.

[Cha10] **Chakraborty R.** – *Knowledge representation Issues, Predicate logic and rules*. A lecture note retrieved from www.myreaders.info/html/artificial-intelligence.html, 2010.

[Cha80] **Chafe W. L.** – *The Pear Stories: Cognitive Cultural, and Linguistic Aspect of Narratives reproduction* (ed). New Jersey: Ablex., 1980.

[Cho59] **Chomsky N.** – *On Certain Formal Properties of Grammars*. Information and Control 22:137- 167, 1959.

[Cio16] **Ciotti F.** – *Toward a formal ontology for narrative*. MATLIT: Materialidades da Literatura, 4(1):29-44, 2016.

[CB12] **Carlos R., Benjamin V.** – *A General Knowledge Representation Model of Concepts*, Advances in Knowledge Representation, Dr. Carlos Ramirez (Ed.), ISBN: 978-953- 51-0597-8, 2012. InTech, Retrieved from: http://www.intechopen.com/books/advances-in-knowledge-representation/a-generalknowledge-representation-model-of-concepts.

[CC01] **Cornbleet S., Carter R.** – *The language of speech and writing*. Routledge London, England, 2001.

[DSS93] **Davis R., Shrobe H., Szolovits P.** – *What is a Knowledge Representation?* Artificial Intelligence Magazine, 1993.

[D+17] **Deleger L., Campillos L., Ligozat A. L., Neveol A.** – *Design of an Extensive Information Representation Scheme for Clinical Narratives*. Journal of biomedical semantics, 8(1):37, 2017.

[Fin12] **Finlayson M. A.** – *Learning Narrative Structure from Annotated Folktales*. PhD Thesis of the Massachussetts Institute of Technology (MIT) USA, 2012.

[FS07] **Feldman R., Sanger J.** – *The Text mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, 2007.

[FFF99] **Frank G., Farquhar A., Fikes R.** – *Building a Large knowledge-base from a Structured Source*. IEEE Intelligent Systems and their Applications, 14(1): 47-54, 1999.

[FWR10] **Finlayson M., Winston P., Richards, W**. – *Computational Models of Narrative: Review of a Workshop*, AI Magazine, 31(2):97-100, 2010.

[GR05] **Gualtieri A., Ruffolo M.** – *An Ontology-based Framework for Representing Organisational Knowledge*, Proceedings of I-KNOW Conference pp. 71(78), 2005.

[HL10] **Hovy E., Lavid J.** – *Toward a Science of Corpus annotation: A new methodological challenge for corpus linguistics*. International Journal of translation 22(1):13-36, 2010.

[JL09] **Jing L., Lau R. Y**. – *Granular Computing for Text Mining: New Research challenges and Opportunities. In International Workshop on Rough sets, Fuzzy sets, Data Mining and Granular-Soft Computing*, Springer volume 5908: 478-485, 2005.

[KS06] **Kashyap V., Sheth A. P.** – *Information brokering across heterogeneous digital data: a metadata-based approach*,

volume 20. Springer Science and Business Media, 2006.

[Lee93] **Leech G.** – *Corpus Annotation Schemes*. Literary and linguistic computing, 8(4), 275-281, 1993.

[Mit95] **Mitchell W. T.** – *Picture in Critical Terms for Literary Study*. University of Chicago Press: USA, 1995.

[Nin15] **Ninan O. D.** – *A Formal Computational Specification and Modelling of African Folktale Narrative*. PhD Thesis of the Obafemi Awolowo University, Ile-Ife, Nigeria, 2015.

[NO13] **Ninan O. D., Odejobi O. A.** – *Towards a Digital Resources for African Folktales*. In Finlayson, M.A (ed), The Third Workshop on Computational Models of Narratives pp 77-82, 2013.

[NT95] **Nonaka I., Takeuchi H.** – *The Knowledge-Creating Company*. Oxford University Press, 1995.

[Ode05] **Odejobi O. A.** – *A computational model of prosody for Yoruba text-to-speech Synthesis*. PhD Thesis of the Aston University, Birmingham, UK, pp. 1-371, 2005.

[Ola13] **Olarinmoye A. W.** – *The Images of Women in Yoruba Folktales*. International Journal of Humanities and Social Science. 3(4): 1-11, 2013.

[Owo05] **Owomoyela O.** – *Yoruba Proverbs*. University of Nebraska Press: USA, 2005.

[PM17] **Poole D. L., Mackworth A. K.** – *Artificial Intelligence, Foundations of Computational Agent*, 2nd Edition. Cambridge University Press, 2017.

[RB10] **Rajiv S., Becerra F. I.** – *Knowledge Management Systems and Processes*, Journal New York: ME Sharpe, pp.13-31, 2010.

[Sav98] **Savage J. E.** – *Models of computation*, volume 136. Addison-Wesley Reading, 1998.

[TMS06] **Tuffield M. M., Millard D. E., Shadbolt N.** – *Ontological Approaches to Modelling Narratives*. Proceedings of 2nd AKT Doctoral Symposium, 2006.

[Zar12] **Zarri G. P.** – *Computational Models of Narratives as Structured Association of Formalised elementary event*. Proceedings of the workshop on Computational Model of Narratives CMN'12(ed) Mark Finlayson, Massachusset Institute of Technology, USA, PP.9-14, 2012.

[***17] **Oxford** – *Definition of Narrative by Oxford English Dictionary online*, 2017.