Multi-Relational Latent Syntax-Semantic “SYNTAXLSEM” Analysis Model for Extracting Qura’nic Concept a New Innovative for Sustainable Society

Asma Abdul Rahman1, Ahmad Abdul Rahman2, Mohd Fadhil Md Din3

1Department of Education in Linguistics, Faculty of Major Languages Studies, Islamic Science University of Malaysia, Nilai, Negeri Sembilan, Malaysia
2Center for Languages Studies and Generic Development, Universiti Malaysia Kelantan, Jeli, Kelantan, Malaysia
3Faculty of Civil Engineering, University Technology of Malaysia, Skudai, Johor, Malaysia

*Corresponding author E-mail: asma@usim.edu.my

Abstract

A flexible model that can represent Qur’anic concept is required for people to understand the content of the Quran. In this research, we propose a Multi-Relational Latent Syntax-Semantic Analysis Model (SYNTAXLSEM) based on a combination of Arabic Semantic and six multiple relations between words, which are synonym, antonym, hypernym, hyponym, holonym and meronym, to precisely extract Qur’anic concept. The existing literatures focus only on very limited relationships between words which could not extract the in-depth concept of Qur’anic without considering the importance Arabic Semantic. Therefore, the objectives of this research are: (1) to analyse and categorize Quranic words according to Arabic Semantic patterns, (2) to propose a new model for extracting Quranic concept using SYNTAXLSEM, (3) to investigate semantic relationships between Qura’nic words, and (4) to validate the proposed model with Arabic linguistic, and Qura’nic experts. This research will be conducted qualitatively through content analysis approach a new innovative technological technique. It is expected that the model will come out with a precise analysis for extracting Qur’anic concept. This will be very significant in enhancing the overall Quran’s understanding among the society in Malaysia and Muslim’s world for sustainable society.

Keywords: Multi-Relational; Latent; Syntax-Semantic; Model Extracting; Qura’ni, Sustainable Society.

1. Introduction

In linguistic study, a concept is referred as a mental structure, which totally depends on relationships between linguistic representatives amongst words and phrases. However, the interchangeable relationship could happen due to the inherent conciseness of natural language. Therefore, many researchers try to solve this problem using ontology approach and Latent Semantic Analysis (LSA) [6, 11], but it needs some considerations before it can be implemented on Arabic, especially in al-Quran, which represents a special classical Arabic language.

1.1. Problem Statement

A variety of problems have emerged in Malaysia community as a result of the rapid developments and the flow of humanitarian and scientific knowledge. Thus that people unable to adapt to the huge number of methods and means of modern education technology. The impact of foreign media and audio-visual and print and unsecured computer network bring about strange behaviour, not only in Malaysia social aspects, but also Arabic language education procedures. The reform of this problem calls for more attention to be paid on theories education teaching methods, so as to the quality and quantity the rules, principles, methods and curricula of teaching and evaluation. This may include putting emphasis moral, cultural and heritage education based on the Qur’an and Sunnah. The uniqueness of Arabic Semantic makes the combination with LSA and ontology more significant to extract Qura’nic concept. It is because Phonological analysis is a basic step in various applications including text mining, information retrieval (IR), machine translation, automatic summarization, and Arabic learning systems [8]. LSA on the other hand has limitation such as difficulty in differentiating fine-grained relations between lexical semantics, synonyms, antonyms and hypernyms. Due to that limitation, various studies have been carried out to overcome this weakness of LSA introducing Multi-Relational Latent Semantic Analysis using synonym, antonym and hypernym [6] but it has not yet been applied on Arabic Text such as al-Quran. Furthermore, Qura’nic ontology studies are lacking of accuracy of language because it depends on the translated version of al-Quran and uses aplastic noun [5, 9-10]. The existing literature shows that there are still gaps and problems in extracting Qur'anic concept as have been mentioned previously, due to inappropriate approaches that had been used ignoring the importance of Arabic Semantic and LSA in deriving the meaning of Qur’anic word. Therefore, in order to address the above mentions problem, we propose a new model in extracting Qura’nic concept based on Arabic Semantic and Multi-Relational Latent Semantic Analysis, using combination of six multiple relations between words, which are synonym, antonym, hypernym, hyponym, holonym and meronym.
1.2. Objective of the research

In order do the research, we need to analyze and categorize Qur'anic words according to Arabic Semantic patterns. Then to investigate semantic relationships between Qur'anic words. After that, a new model will be proposed for extracting Qur'anic concept using SYNTAXLSEM and validate the proposed model. Before that, we need to know how to analyze and categorize Qur'anic words according to Arabic Semantic patterns. We are also need to investigate the result semantic relationships between Qur'anic words. We are also having a problem such as difficulty in differentiating synonyms, antonyms & hypernyms. Due to that limitation, various studies have been carried out to overcome this weakness of LSA by introducing the notion of polarity. The recent LSA research combines multiple relations between words by constructing three relationships which are known as Multi-Relational Latent Semantic Analysis using synonym, antonym and hypernym [3].

The above model encodes the raw data in a 3-way tensor to encode multiple word relations, which are synonym, antonym & Hypernym. Each slice captures a particular relation and is in the format of the document-term matrix in LSA. The tensor decomposition method was applied to generalize the representation and discover unseen relations between words. Although this recent work tries to explore concept-based information access via ontology approach and LSA on natural language [11], it has not yet been done in al-Quran, which represents the purest and most authentic form of the classical Arabic language [2-4]. Muslims believe that words of Holy Quran are divine and eternal. No alteration is whatsoever possible as Allah Almighty Himself has taken into His Hand.

Ontology on the other hand, is defined as a description of the concepts and relationships that can exist for an agent or a community of agents [11]. The use of ontology facilitates identification of concepts and their linguistic representatives, given a key concept. It is similar to a dictionary or glossary, but with greater detail and structure. The existing researches show two types of approaches in Qur'anic language computational models based on ontological approach. First, traditional approach which is based on models of Roman language done by [9] using English Translation of Holy Quran by [5] using Syammil Al-Quran Miracle the Reference. These two studies lack of accuracy of language because it depends on the translated version of al-Quran. Second, modern approach is based on an authoritative and rich source of Arabic language, i.e. the Holy Quran. This study uses lexicon ontology development based on the Unified Process for Ontology (UPON), an ontological engineering approach [10]. The study focuses mainly on nouns from the “time” semantic field, which is aplastic noun - having no forms except one.

<table>
<thead>
<tr>
<th>المفهوم</th>
<th>الرمز</th>
</tr>
</thead>
<tbody>
<tr>
<td>عذر</td>
<td>ج</td>
</tr>
<tr>
<td>حر</td>
<td>ع</td>
</tr>
<tr>
<td>ركذ</td>
<td>د</td>
</tr>
<tr>
<td>ذنب</td>
<td>ط</td>
</tr>
<tr>
<td>حرام</td>
<td>ج</td>
</tr>
<tr>
<td>محظور</td>
<td>ذ</td>
</tr>
<tr>
<td>عاصفة</td>
<td>ح</td>
</tr>
<tr>
<td>حرام</td>
<td>ط</td>
</tr>
<tr>
<td>عاصفة</td>
<td>ح</td>
</tr>
<tr>
<td>حرام</td>
<td>ط</td>
</tr>
<tr>
<td>عاصفة</td>
<td>ح</td>
</tr>
</tbody>
</table>

Fig. 1: Multi-Relational Latent SYNTAXLSEM formula Analysis [2-4]
In the above model, semantic dimensions need to be determined, and new words will be organized into a hierarchical classification with general concepts at the top, and specific at the bottom. Words in the hierarchy are associated with components via ontological relations. This classification structure of the ontology implies that the deeper word moves into the hierarchy, the more arguments the componential formula will have, and therefore the meaning narrows. In contrast, words at higher levels have fewer arguments in their componential formula, and therefore the meaning broadens. This model however did not include the root of words which is the appropriate to Arabic language. The existing literature shows that there are still gaps in extracting Qur’anic concept because the aspect of Arabic Semantic and LSA does not fully utilize. Therefore, in order to address the above problem, we propose a new model in extracting Qur’anic concept based on Arabic Semantic and Multi-Relational Latent Semantic Analysis, using combination of six multiple relations between words, which are synonym, antonym, hypernym, hyponym, holonym and meronym.

3. Methodology

This research is focusing on designing a multi-relational LSA model in extracting Qur’anic concept based on combination of Arabic Semantic and six multiple relations between words, which are synonym, antonym, hypernym, hyponym, holonym and meronym. The researcher will use text analysis method as follows:

1st Phase: Phonological Root Analysis
This phase will be conducted in following steps:
1. The researcher will identify digital document file of Qur’anic text. The text’s authentication will be verified by expert.
2. The digital document will be analyzed using corpus analysis using Wordsmith version 2.1.
3. Each word in text will be categorized into two families of verb and noun, excluding particles.
4. The researcher will identify root of each word in these two categories, then categorize the words belong to similar root in another sub-family.
5. To enhance the confidence of investigation, the result then will be triangulated with three Arabic referred dictionaries:
   i) Lisan al-Arab by Ibn Manzuur,
   ii) Mu’jam al Waseet by Majma’ Luqughah al-Arabiyyah in Cairo, and
   iii) al-Mu’jam al-Asabi by al-Munazzamah al-Arabiyyah lil Tarbiyyah wa Thaqafah wal ‘Ulm

2nd Phase: To Propose a New Model for Extracting Qur’anic Concept
Based on previous 1st phase result, the researcher will follow the next steps:
1. Create and propose a model using six multiple relations between words in Latent Semantic Analysis as follow:
   i) Synonym: A word or phrase that means exactly or nearly the same as another word or phrase in the same language.
   ii) Antonym: A word opposite in meaning to another
   iii) Hypernym: A word with a broad meaning constituting a category into which words with more specific meanings fall
   iv) Hyponym: A word of more specific meaning than a general or super ordinate term applicable to it
   v) Holonym: A concept of which this concept forms a part
   vi) Meronym: A term that denotes part of something
2. Related connections of word to conceptual meaning then will be identified.
3. Linking the related connections to key concept in tree-form drawing.
4. Describing the design of extracting Qur’anic concept based on ontological approach in Latent Semantic Analysis.

3rd Phase: Investigating Semantic Relationships between Qur’anic Words
The proposed model will be validated by testing a concept of “Sight” in Quran as a sample. (an expected analysis is shown as below)

4th Phase: Validating the proposed model
The result then will be evaluated by two groups of experts:
1. An Arabic linguist who will examine Phonological Root Analysis and the merging process with multi-relational latent semantic analysis, and
2. Qur’anic experts who examine holistically the concept of Sight in al-Quran, according to Islamic perspective the evaluation will be conducted in focus group interview. Any suggestion or feedbacks will be taken into consideration to improve the model. A sample of proposed model in using SYNTAXLSEM to extract Qur’anic concept.
4. Discussion and Conclusion

Policies for government agencies to help them curb and solve the educational system in teaching and learning Arabic linguistic training and practices in tricking consumers into thinking that particular product by apply a new innovative method for whole level of students and society. This is a new invention thus; no comparable product existed in the market. Impact on human being and socio economic. The “SYNTAXSEM” have improved knowledge and linguistic skills in the production of higher quality of human capital. Furthermore, the researcher was sales from text and reference, more formula books, CDs and Courses, training, workshop generate additional income for. Harmonized society and community where Muslim and non-Muslim can seat together enjoying their communication with multiple lingual; that Muslim be it Malaysian or international could be unity of the nation local or foreign without any doubt or miscommunication. Academically closing the gap on communication between society and community related research, providing the aspect of language, culture and society perspective. As academics, we aim to contribute in closing the gap on miscommunication related research by looking from the aspect of language, culture and society perspective. The research will result in publications and we target to bring about the knowledge into seminars nationally and internationally and symposium to share with universities and industry players and those who are interested. We have to find that the research output will useful to attract foreign direct investors, mainly Asian countries, and other potential interested investor from Middle East, and Europe to understand the opportunity in linguistics education, to join in the development of Malaysia and other countries of Renewal Energy (SCORE) specifically in two identified clusters Linguistics Education with a new innovative method, society, cultures—Hub sustainability and Tourism.

References


