Towards Representing Semantic Opposition Using Semantic Web Technologies

Hend S. Al-Khalifa[1], Maha M. Al-Yahya[1], Alia Bahanshal[2] and Iman Al-Odah[2]

[1] Information Technology Dept, CCIS, King Saud University
hend@ccis.edu.sa, malyahya@ksu.edu.sa
[2] King Abdulaziz City for Science and Technology
abahanshal@kacst.edu.sa, Iodah@kacst.edu.sa

Abstract. This paper presents a work-in-progress project for building a framework for recognizing and identifying semantic opposition terms in the Holy Quran using semantic web technologies.

Keywords: Holy Quran, Arabic Language, Ontology, Semantic Opposition, Semantic Web.

1 Introduction

The Holy Quran is the Muslims Noble Scripture; the rhetoric presented in the Quran is highly valued in terms of religion and language. Quran indicates that no text is even comparable to it in terms of its linguistic beauty; therefore, Quran text has been considered as classical Arabic similar to what is used in literary texts.

Many linguistic styles can be found in the Holy Quran such as metonymy, synonyms and semantic opposition. Semantic opposition as a sub-field of structural linguistics, studies the relationship between words based on their properties or components [1].

Linguists usually use semantic opposition to explain the semantics and meaning of words and/or enrich the quality of writing. Therefore, semantic opposition has lately gained some interest to Arab linguist experts (e.g. [2]) by studying the Holy Quran text and identifying the occurrence of semantic opposition manually.

In order to represent the domain of semantic opposition for a given field and automate the process of its identification, we propose the establishment of a computerized framework that exploits the power of semantic web technologies and natural language processing to represent and identify Semantic Opposition (SO) in the Holy Quran. The initial focus of the research will be on SO on the word level. Results of this research will provide a basis for building intelligent language processing tools that are useful in the analysis of Quranic Text. It also provides a foundation for building Web-based Arabic text analysis tools.

2 Background

Semantic opposition is considered as one of the semantics relationships. Any word can be seen as a number of distinct elements or components of meaning. In order to recognize that two words are semantically opposite, one needs to perform a componential analysis for each word, i.e. provide definition for a word in terms of a few components. For example, 'Man' can be marked as (male) (adult) (human), and 'Women' as (female) (adult) (human), we can conclude that 'Man' and 'Women' are semantically opposite, since the two words didn't agree on one component that is (male/female) [1].

Semantic opposition is also based on the concept of semantic fields/domains. A Semantic field is defined as *"a group of words that stand in paradigmatic opposition to one another and share at least one semantic component"* [3].

Semantic fields' classification has many variations and they usually depend on community and culture, so there is no clear representation of them. However, the most well-known semantic fields' classification was introduced in the Greek New Testament taxonomy (GNT), which classifies semantic fields into four main categories: *Objects, Relations, Events and Abstractions*, and represents the fields as a hierarchy of primary, secondary and tertiary terms[1].

Semantic fields are very useful in determining the applicability of semantic opposition between two terms. For instance, 'Man' and 'Bull' are not semantically opposite because 'Man' belongs to the 'Human' semantic field while 'Bull' belongs to the 'Animal' field.

Given the inner-structure and nature of semantic opposition we can assume that the best artifact for representing the function of semantic opposition is the use of ontologies. Ontologies are knowledge representation structures that describe the concepts, relationships, and constraints within a specific domain [4]. Given the power of ontologies in representing knowledge, we decided to utilize its power to explore this niche field of language from which a computer assisted mechanisms such as semantic web technologies can help.

3 The Proposed Framework

SemQ is the name of our proposed framework that takes as an input a Quranic verse (i.e. sentence) and outputs the list of semantically opposed words in the verse along with their degree of opposition.

The framework architecture consists of two major components: the domain ontology and the SemQ Tool.

The main goal for building domain ontology is to mimic how the human brain keeps the semantics stored. Given the lexical semantics and the ontological knowledge about verbs and nouns along with their components, we decided to build the domain ontology on top of the Greek New Testament (GNT) taxonomy.

[1] http://bible.org/seriespage/introduction-new-testament

GNT is an existing taxonomy specifically designed to provide the basis for more specific domain ontologies. However, we need to point out that building a complete ontology for all domains in GNT is not a trivial task and requires more time and effort. Therefore, we have decided to concentrate on representing a small domain in GNT, which is the domain of Women, given that we have a complete reference in this domain [5] compared to other domains.

The main classes related to the women field include: *Relations, Accessories and Essence*. Several other classes are either included to help the class organization, or to maintain consistency with other domains. As for properties, they will express the relationships between classes.

The tool is expected to work in two modes: identification mode and expert mode. In the identification mode, the tool works automatically to identify semantically opposite terms. In the expert mode the tool works as a manual identification tool that relies on a subject matter expert (SME) to populate the ontology with terms and their properties.

Figure 1 depicts the SemQ tool process pipeline, which consist mainly of two parts: 1) Arabic Text Preprocessing and 2) Ontology Engine.

Fig. 1. SemQ Tool Pipeline.

The list of stems from the pre-processing phase is entered into the ontology engine to search for the stem in the domain ontology and retrieve its components. Then the engine decides whether semantic opposition exists and determines its degree: absolute or scalar, based on a set of predefined rules.

Acknowledgments. This research project is supported by King Abdul Aziz City for Science and Technology (KACST).

References

1. F. R. Palmer. *Semantics: A New Outline*. Cambridge University Press. 1976
2. Al-Helwa, N. "Semantic Opposition: A Theoretical Applied Study on Surat Al-Nisa". 2006. In Arabic.
3. P. Fronzaroli, *Componential Analysis*, ZAH 6, p. 79-91
4. Gruber, T. *Toward principles for the design of ontologies used for knowledge sharing*. International Journal of Human-Computer Studies, Vol. 43, Issues 4-5, November 1995, pp. 907-928. 1995.
5. Al-Anbar, J. *Women Semantic Field*, 2008. PhD Thesis. 508 pages. In Arabic.