The Meaning Explorer

A Distributional Semantic Analyzer For The Holy Quran Words

User Manual

Maha Sulaiman Alrabiah

Computer Science Department

King Saud University

2014

Content

1 Introduction

Distributional Semantic Modeling is considered as an empirical approach that is mainly concerned with modeling words' meanings using words distribution statistics gathered from very large corpora. It is basically built on the notion of the *Distributional Hypothesis*, which dates back to Zellig Harris stating that "difference in meaning correlate with difference in distribution". In other words it states that "words which are similar in meaning occur in similar contexts", and hence "words that occur in the same contexts tend to have similar meaning".

Although Harris did not specify in his theory what kind of difference one should look for nor how it affects meaning, this missing information was revealed by an earlier linguist called Ferdinand de Saussure who depicted that this difference originates from two kinds of relations, which are syntagmatic and paradigmatic relations. Syntagmatic relations are linear combinatorial relations that are established between words that co-occur together in sequential text, while paradigmatic relations are substitutional relations that are established between words that occur in the same context, share neighboring words, but do not co-occur in the same text [1].

The Meaning Explorer is a tool that is mainly concerned with analyzing the distributional semantics of the words in the holy Quran and extracting the syntagmatic relations between them.

1.1 Purpose & Scope

The main purpose of this tool is to help users in extracting syntagmatic relations between words, lemmas and roots available in the holy Quran; these relations include identifying significant collocates and words' co-occurrences, which are the words that tend to appear together in the same context. In addition, the tool also provides other helpful functionalities that complement the primary purpose, which include a Key Word In Context (KWIC) concordance, in addition to frequency lists of all words, lemmas and roots in the holy Quran. The main intended users of this tool are Arabic Quranic scholars and linguists.

The Meaning Explorer applies a new distributional semantic model to extract words' significant co-occurrences from the holy Quran. This model is based on the Refined MI association measure [2] applied to all words within a symmetric sliding window of five words surrounding the node word. The Refined MI association measure is chosen because of its remarkable ability to extract syntagmatic relations between the words of the holy Quran [2].

2 Installation

1- Download the tool from the King Saud University Corpus for Classical Arabic site using the following link:

http://ksucorpus.ksu.edu.sa/wp-content/uploads/2014/09/MeaningExplorer.zip

2- Unzip the folder.
3- Run the MeaningExplorerEnfVr.jar file, and the tool will start.

3 Functionality

When you start the tool, you will see the main user interface showing the "Frequency" tab, as in Figure (1).
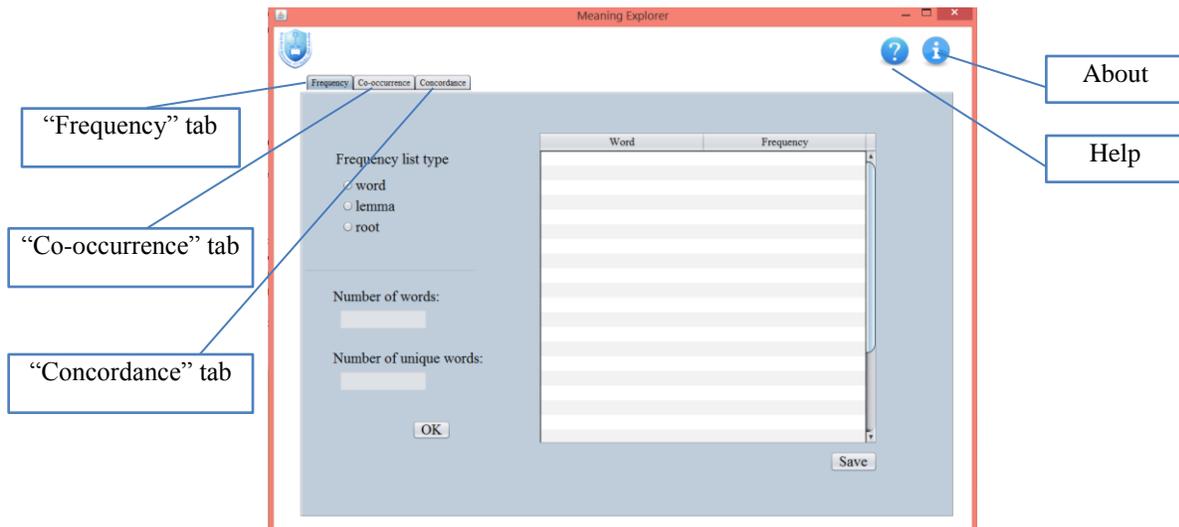

Figure1: Main user interface

3.1 Frequency

Using the "Frequency" tab, you can generate frequency lists of all the available words, lemmas and roots in the holy Quran sorted in ascending order, as in Figure (2), using the following steps:

1- Determine the type of the frequency list by choosing "word", "lemma" or "root"; the default is "word".

2- Press "OK" button.

3- The desired frequency list will appear in the table sorted in ascending order.

3

4- To save the current list as a text file, press the "Save" button and the table will be saved in the same folder by the name "Frequency.txt".
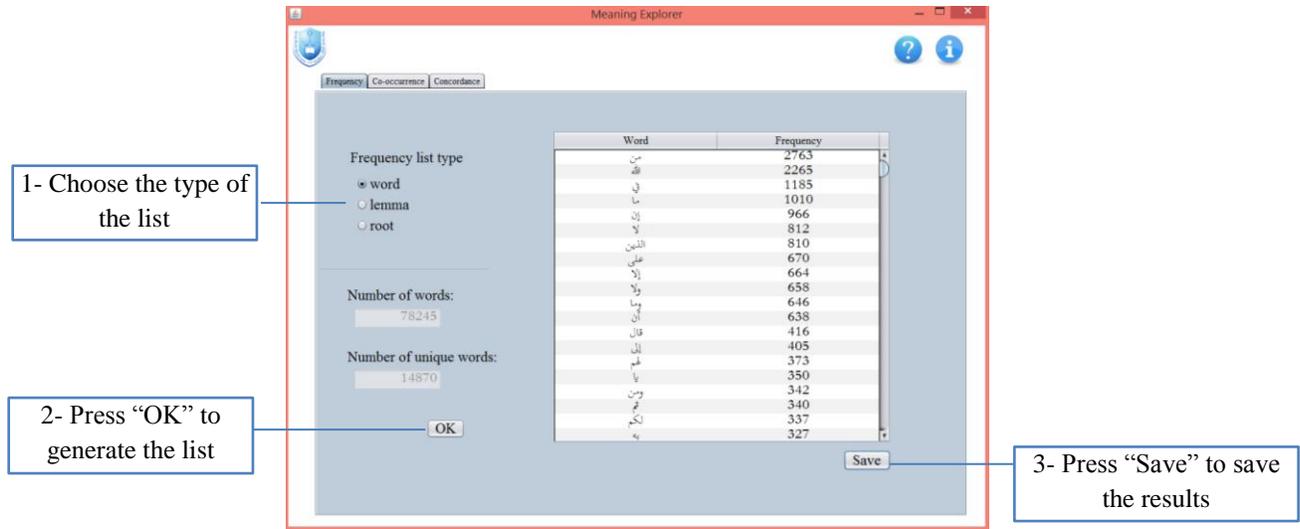

1- Choose the type of the list

2- Press "OK" to generate the list

3- Press "Save" to save the results

Figure2: The "Frequency" tab

3.2 Co-occurrence

Using the "Co-occurrence" tab, you can extract the words that tend to co-occur with any desired word in the holy Quran conveying a significant relation between them. These relations are measured using the Refined MI association measure applied to five words surrounding the desired target word. The extracted co-occurring words are displayed in the table in descending order, meaning that the ones with higher and more significant relation are shown first. You can also extract significant co-occurring lemmas and roots all using the following steps:

1- Determine the type of co-occurrences extracted by choosing "word", "lemma" or "root".

2- Enter the desired "word", "lemma" or "root" in the text box.

3- Press the "Search" button.

4- In the "Word frequency" text box, you will see the frequency of the "word", "lemma" or "root" that you have entered.

5- A list of the most significant co-occurring words will appear in the table sorted by the value of the association measure in descending order, in addition to the frequencies of those co-occurring words in the holy Quran and the number of times both the target

and the co-occurring words appeared together in the holy Quran. In case no co-occurring words were found, a message will be displayed indicating this.

6- To save the results, press the "Save" button, and the table will be saved in the same folder with the name "CoOccurrences.txt".

Figures (3-5) shows how to extract the significant co-occurrences of the word "الليل", the lemma "ليل" and the root "ليل".

Figure 3: Extracting significant co-occurring words of the word "الليل"

Figure 4: Extracting significant co-occurring lemmas of the lemma "ليل"
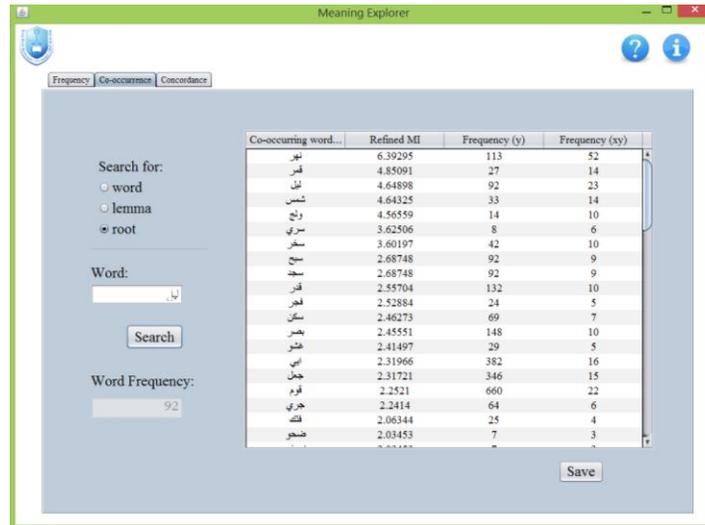
Figure 5: Extracting significant co-occurring roots of the root "ليل"

3.3 Concordance

Using the "Concordance" tab, you can generate a KWIC concordance of any word in the holy Quran, as in Figure 6 below using the following steps:

1- Type the desired word in the "Word" text box.

2- Press the "Search" button.

3- The frequency of the target word will appear in the second text box.

4- The concordance will be shown in the table, with the following fields:

- Left context: the Quranic text to the left of the target word.
- Word: the target word as it appears in the context.
- Right context: the Quranic text to the right of the target word.
- Surah: the name of the Surah in which the context appear.
- Ayah: the number of the ayah in which this context appear.

5- To save the results, press the "Save" button and the table will be saved in the same folder with the name "Concordance.txt".
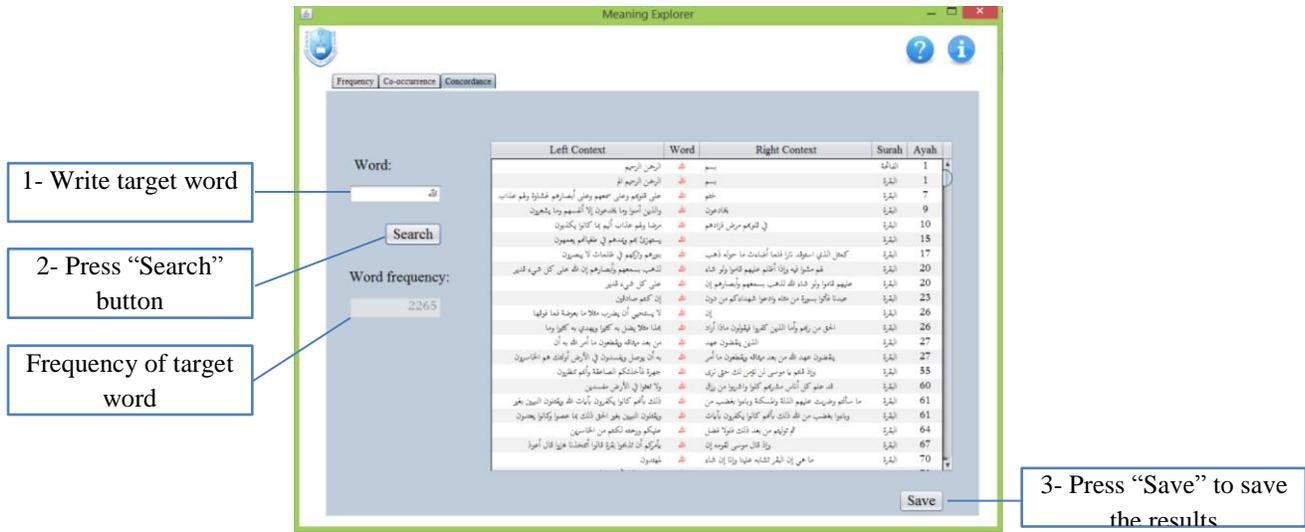
Figure 6: The concordance of the word "الله"

References
[1] Alrabiah, M., Al-Salman A. & Atwell, E., 2013*b*. KSUCCA a cornerstone to study thee semantics of the Quranic words in the light of distributional lexical semantics. In *Proc. NOORIC'1435-201*. Almadinah Almonawwrah, Saudi Arabia, (in Arabic).

 [2] Alrabiah, M., Al-Salman, A. & Atwell, E., 2014*b*. The Refined MI A Significant Improvement to Mutual Information. *In Proc. International Conference on Asian Language Processing 2014 (IALP2014)*. Kuching, Malaysia.