

## Conference Paper

# Utilizing the Quranic Arabic Corpus as a Supplementary Teaching and Learning Material for Arabic Syntax: An Overview of a Web-based Arabic Linguistics Corpus

Muhammad Lukman Arifianto

Universitas Negeri Malang, Malang

**ORCID:**

Muhammad Lukman Arifianto: <https://orcid.org/0000-0001-7004-6360>

**Abstract**

This article discusses the plan to implement the Quranic Arabic Corpus for the development of Arabic syntax teaching and learning materials (TLM), especially at the university level. The study explores an alternative Arabic syntax TLM development by utilizing a corpus, one of the most important linguistic research resources. In recent years, corpora have been widely utilized in developing teaching materials. The Quranic Arabic Corpus is beneficial not just in terms of additional references or examples, but also in facilitating grammatical understanding and discussion. The highly adequate syntactic annotations recorded in this corpus can also be used to discover and explore the syntactic analysis of every word in *Quran*, so that it can be easily implemented in the analysis, examples of each chapter, or section of the discussions in Arabic syntax. Based on the analysis of this corpus' content and its relevance with the existing Arabic syntax teaching materials, it is shown that this corpus can be incorporated as a supplement of Arabic syntax TLM. The following six steps should be followed to utilize this corpus in Arabic Syntax learning: (1) prepare the necessary tools, (2) select learning materials, (3) determine keywords, (4) access QAC, (5) search for keywords in QAC, and (6) select search results and proceed with syntactic analysis.

**Keywords:** Arabic syntax, linguistic corpus, Quranic Arabic corpus, teaching and learning materials

Corresponding Author:

Muhammad Lukman Arifianto  
muhammad.arifianto.fs@um.ac.id

Published: 1 March 2021

Publishing services provided by  
Knowledge E

© Muhammad Lukman

Arifianto. This article is  
distributed under the terms of  
the [Creative Commons](#)

[Attribution License](#), which

permits unrestricted use and  
redistribution provided that the  
original author and source are  
credited.

Selection and Peer-review under  
the responsibility of the ISOLEC  
Conference Committee.

## 1. Introduction

Arabic linguistic corpus (plural: "corpora") has been gradually built and explored in recent years, and it has become a research trend in the field of computational language, natural language processing, and modern linguistics (McEnery, Hardie, & Younis, 2019). This exploration can be seen from the emergence of several Arabic corpus project sites such as the Quranic Arabic Corpus (<http://corpus.quran.com/>) by Dukes (2009) from

 OPEN ACCESS

University of Leeds, International Corpus of Arabic (<http://www.bibalex.org/ica/ar/>) by Bibliotheca Alexandria, The Arabic Learner Corpus (<https://www.arabiclearnercorpus.com/>) by Abdullah Alfaifi and Eric Atwell from University of Leeds, The Arabic Web Corpus (arTenTen) (<https://www.sketchengine.eu/artenten-arabic-corpus/>), the workshops on Arabic linguistic corpus—such as The 3<sup>rd</sup> Workshop on Arabic Corpus Linguistics (WACL-3) by Cardiff University—and a number of research papers or articles related to this topic. The linguistic corpus can provide teachers, learners, second language acquisition researchers, lexicographers, language materials writers, and many others, with a valuable data resource (Alfaifi & Atwell, 2013).

As stated by Whitcomb & Alansary (2018), in order to exploit the utility of corpora, language instructors do not have to answer questions related to the debate inside those linguistic corpora on the question of “representativeness” or discussion about the method, tool, and theory of language. So far as the corpora’s compilation is transparent in its methodology and criteria, language teaching professionals should feel comfortable using language data gathered in corpora to enhance classroom instruction, materials development, and independent study.

Discussions on the existence of the Arabic language corpus and its use as teaching and learning material have been conducted by Hizbullah, Fazlurrahman, & Fauziah (2016), which looks at the importance of the existence of an Arabic corpus in Indonesia. Hizbullah et al. (2016) mentioned that the initial stage is preparing the Arabic learner corpus in Indonesia. In their opinion, this step should be done at the beginning because it is very important to map the extent of Arabic learning and its studies at various levels of education in this country. Eventually, this corpus can be linked to several linguistics fields, such as morphology, syntax, semantics, pragmatics, discourse, lexicography, literature, translation, and many more.

Furthermore, Zaki (2017) has also conducted a similar study regarding the use of corpora in TAFL (Teaching Arabic as Foreign Language) classes. Her paper presents a model of integrating corpus-based teaching within a textbook-based syllabus for both elementary and intermediate levels. It includes ideas for a material design for vocabulary building activities and enhancement of grammatical rules for Arabic learners.

However, in the real act of teaching and learning, using this linguistic corpus as teaching and learning material (TLM) is still merely a discourse, especially in Arabic teaching and learning. One of the reasons, according to Granath (2009, p. 47), is that learning how to use a corpus is seldom part of teacher training courses. As a result, teachers at the university level and at lower levels lack the skills needed to use this kind of “native-speaker consultant”. If the training in using corpora were integrated into

university-level courses such as syntax or written proficiency and translation, it could become just as natural to consult a corpus as to look up an item in a dictionary or a grammar book in time.

This paper attempts to test and investigate the utilization of the Quranic Arabic Corpus (QAC), primarily related to its grammatical discussions, for Arabic syntax teaching and learning materials development. As mentioned before, based on an in-depth overview and observation in the content of QAC and its relevance with the existing Arabic syntax teaching materials in university level, it can be noticed that this corpus can be well-utilized for Arabic syntax teaching and learning materials development, not just in terms of additional references or examples of discussion, but also in terms of facilitating grammatical understanding. It is because this corpus has quite complete, clear, and comprehensive grammatical discussion. The highly adequate syntactic annotations recorded in this corpus can also be well-used to find out and explore the syntactic analysis in every word in the Quran, so that it can be easily implemented in the analysis and examples of each chapter or section of the discussions in Arabic syntax.

As reinforced by Wahba (2018), Arabic language teachers or instructors need to have a comprehensive overview of the available language teaching and learning materials to be informed of the main trends in materials development and the gaps in the existing literature. In case of the various needs of Arabic language programs and language learners, surveying and analysing current language materials helps language educators develop principles that enable teachers to write, use, evaluate, and make proposals for future progress in material development and research.

## 2. Method

This research is qualitative research. It uses the descriptive-analytic method in delivering the result of the discussion because it is only focused on the circumstance of the QAC utilization as a supplementary Arabic syntax teaching and learning materials, not in the quantity or amount of this utilization. As Mohajan (2018) stated, the purpose of qualitative research is to systematically describe and interpret an issue or phenomena from the point of view of the individual or population being studied and generate new concepts and theories. Thus, this observation is intended to describe, illustrate, and portray this corpus' utilization in Arabic syntax teaching and learning materials based on the TLM (Teaching and Learning Materials) development theoretical framework. Nevertheless, it does not rule out the possibility of discovering or generating a new concept in Arabic syntax teaching and learning.

This research is conducted in three stages: (1) data collection, (2) data analysis, and (2) presentation of the result. The data needed for this research are the Arabic grammatical discussions and annotations recorded in the QAC website (*corpus.quran.com*) and the grammatical discussions generated from the Arabic syntax teaching materials at the university level, especially in Indonesia. The data collection is intended to see the potential and similarities in terms of the Arabic Syntax learning concept in these two learning sources. After the required data has been collected, it is then followed by an integrated analysis of the corpus' use in the usual Arabic Syntax learning steps. The combination of existing Arabic syntax learning materials and the use of data sets collected in QAC will produce new steps in learning Arabic syntax. After the analysis has been carried out, the final step is to present the analysis results in a simple step map and describe those steps in detail.

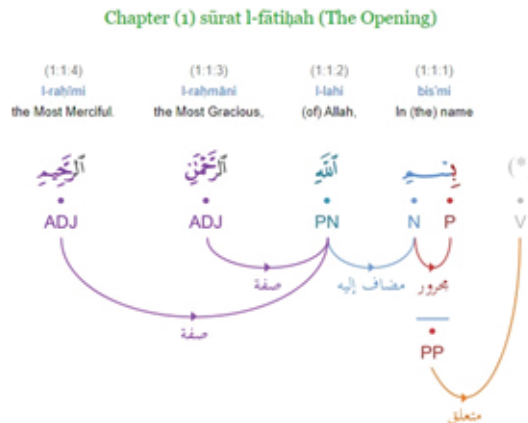
### 3. Findings and Discussion

This section presents the features of QAC that can be used as a supplement to Arabic Syntax learning materials. In addition, it also presents the steps to use the QAC in the real practice of learning Arabic Syntax, as well as an analysis in terms of advantages and disadvantages in the utilization process.

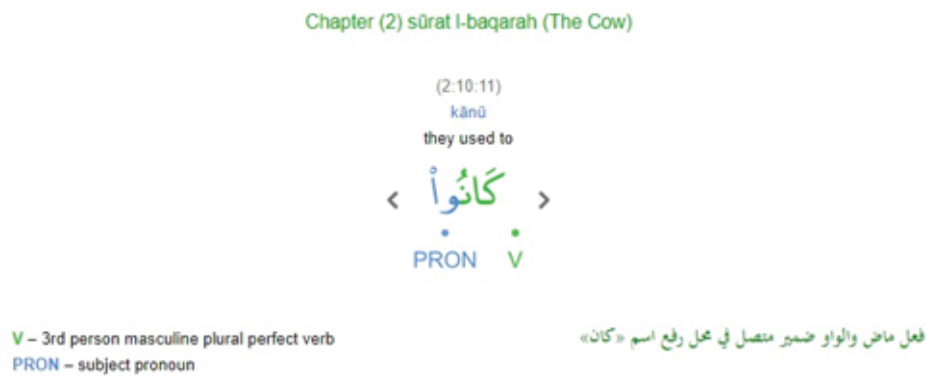
#### 3.1. The Quranic Arabic Corpus syntactic features

The Quranic Arabic Corpus (QAC) is a collaboratively built linguistic data source started at the University of Leeds (2009), with multiple layers of annotations including part of speech, morphological segmentation (Dukes & Habash, 2010), as well as grammatical analysis or syntax (Dukes & Buckwalter, 2010). This corpus is designed to be an open-source of research data (Dukes, Atwell, & Habash, 2013). Based on observations, as the supplementary of Arabic syntactic learning materials, the QAC has quite complete features in terms of grammatical analysis on the Quran verses (Dukes, 2009), both related to morphology (*Sharf*) and syntax (*Nahwu*), they are: (1) syntactic dependency treebank, (2) grammatical annotation (morphology and syntax), and (3) discussion of Quranic grammar.

As found in the corpus in general, this corpus (QAC) is also equipped with a root word search menu (Dukes, 2009) to make it easier for users (Arabic syntax teachers and learners) to find the desired word and context. This search menu will be one of the several essential steps in the stages of utilizing the corpus for Arabic Syntax teaching



**Figure 1:** The syntactic dependency treebank.



**Figure 2:** Grammatical annotation on each word of Al-Quran.

and learning because this will be a means of bridging the concepts or learning material that has been obtained and the use of the corpus to strengthen the learning outcomes.

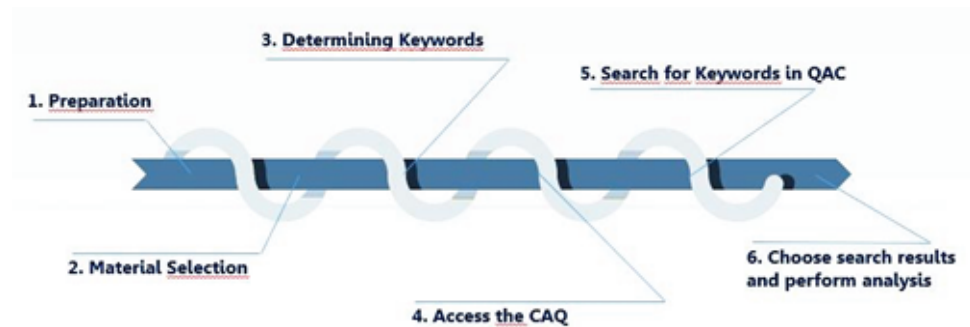
### 3.2. Implementation of QAC in Arabic syntax teaching and learning

There are at least 6 (six) steps in the process of utilizing the QAC for Arabic syntax teaching and learning, which is described in the following flow:

The following is a brief description of each of these steps:

#### 3.2.1. Preparation

Before using QAC, there is one main requirement that needs to be fulfilled: the availability of a device either in the form of a Smartphone or Laptop (PC) connected to a reliable Internet connection. Without a device connected to the Internet, the corpus—especially



**Figure 3:** The process of corpus utilization.

those based on the Web—will not be accessible to users and used for language teaching and learning.

### 3.2.2. Material selection

Determine the Arabic syntax learning material to be studied. For example, we will try to use the QAC to develop a discussion of the *an-Nawasikh* (النواسخ) material in the Basic Arabic Syntax course. More specifically, we will take the example material about *Kana wa Akhawatuha* (كان وأخواتها) as one of the discussions.

### 3.2.3. Determining keywords

After the *Kana wa Akhawatuha* learning material has been delivered well by the lecturer in a classical mode, we should determine the keywords from this topic. As for what is meant by keywords in this step, the core words are contained in the examples of sentences related to the learning material. These keywords will be used as a “key” in the process of searching for linguistic data in the QAC.

### 3.2.4. Accessing the QAC

The next step is to select and access the corpus of data that will be utilized; the Quranic Arabic Corpus (<http://corpus.quran.com>). In the QAC, users do not need to register to access the available corpus content to be used immediately.

Note: For the record, each corpus has different characteristics and features. Preferably, before choosing a corpus to use, it is a good idea for users to first study the characteristics of each corpus by reading the available instructions and carrying out a

simple exploration of the various information and features contained in each corpus so that users can recognize the corpus and can be applied per desired learning objectives.

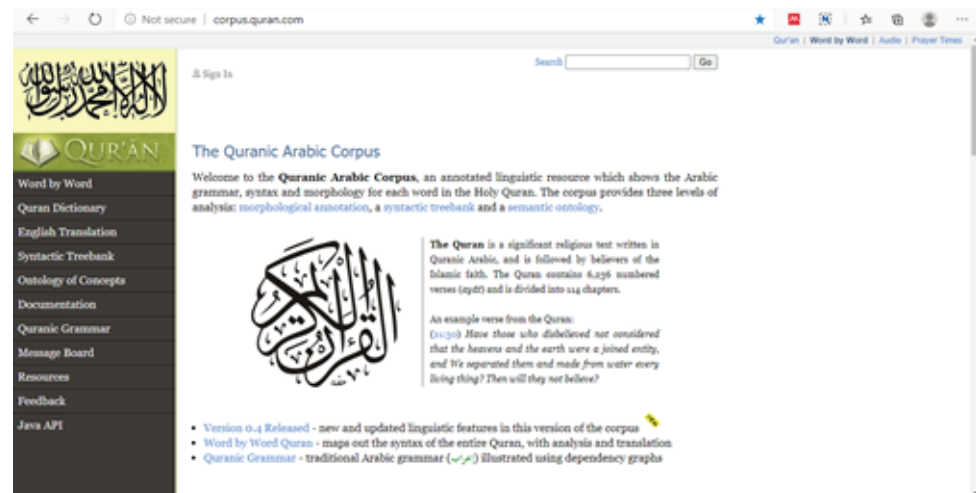


Figure 4: main page of QAC.

### 3.2.5. Searching for Keywords

After the corpus page is ready, the next step is to do a keyword search using the search menu provided by QAC located on the top right of main page. In general, the corpus has a data search feature to trace the use of words and their context (phrases, clauses, or sentences), because the main function corpus is providing language data for various needs. For example, the search for the word Kana (كان) in the Quran via the QAC is done in the following ways:

- Go to the main page of QAC, then type the keywords you want to search for (كان) in the search menu (Search), then click Go
- After that, the search results for the word *Kana* (كان) will appear along with various morphological-inflective changes in the Koran. The data displayed in this search result includes information about the frequency of the word, letters and verses, translations (in English), as well as sentences or verses containing the word

### 3.2.6. Choosing and profiling search results and performing an analysis

After searching for the word *Kana* (كان) has been successfully carried out, select the sentences that we want to highlight—either partially or completely—as an example of the application of theory or learning material of *an-Nawasikh (Kana wa Akhawatuha)*.

To perform a similar search on other keywords, it can be done by repeating the steps for using the corpus that has been described previously.

When we have found examples of applying theories that are suitable for learning, then we can continue by analysing the sentences that have been selected in terms of their *l'rab* (syntactic analysis) to find out the function of each word (part of speech) and the structure or pattern of the sentence. Apart from being carried out by lecturers as an enrichment of learning material, the search for linguistic data through this corpus can also be assigned to students as their learning process in exploring examples of applying Arabic syntactic theory that they have already studied.

### 3.3. Strengths and weaknesses in utilizing QAC

Based on the description above, it is concluded that the QAC can be used in the development of Arabic syntactic learning materials or materials. As an alternative to the development of Arabic syntactic learning materials, of course, the corpus has several advantages and disadvantages. Here is some strength (advantages) of using the QAC in the development of Arabic syntactic learning materials based on the author's experience:

1. The corpus is equipped by adequate syntactic annotations which can be well-used to find out and explore the syntactic analysis in every word in the Quran,
2. The corpus makes it easy for users to search for words they need or want (for example, in studies of the use of "*Kana wa Akhawatuha*", "*Inna wa Akhawatuha*", and so on),
3. The corpus can be a source of research material related to linguistic phenomena (use of language, language errors, etc.),
4. Students are directly involved in the development of learning materials.

The weaknesses (deficiencies) in the use of the QAC in Arabic syntax learning are as follows:

1. No corpus can search for the desired grammatical element (research materials already exist but have not been systematically compiled into a data corpus)
2. Not all keywords can be found in the corpus, especially if the word is rarely used (ancient)



3. In QAC, morphological and grammatical annotations are only available in English; this may make it difficult for Arabic learners from Indonesian speakers (however, on the other hand, it could be an additional learning material in terms of enrichment of linguistic terms)

## 4. Conclusion and Suggestion

Based on the findings and discussion, it can be noticed that this corpus is very useful, not just in terms of additional references or examples of discussion, but also in facilitating grammatical understanding and discussion. The highly adequate syntactic annotations recorded in this corpus can also be used to find out and explore the syntactic analysis in every word in the Quran, so that it can be easily implemented in the analysis and examples of each chapter or section of the discussions in Arabic syntax. Thus, the QAC can be well-used as a supplement to Arabic Syntax teaching and learning.

This research still needs to be followed up with the use of QAC in more real learning practices to determine the student's response to this utilization and their Arabic grammatical knowledge retention. In addition, it can also be explored the use of this QAC corpus for other language learning materials.

## References

- [1] Alfaifi, A. and Atwell, E. (2013, June). Potential use of the Arabic Learner Corpus. Presented at *Leeds Language, Linguistics and Translation PGR Conference 2013*. Leeds: Leeds University.
- [2] Dukes, K. (2009). *The Quranic Arabic Corpus*. Retrieved from <http://corpus.quran.com>.
- [3] Dukes, K. and Habash, N. (2010, May). Morphological Annotation of Quranic Arabic. Presented at *Proceedings of LREC 2010, Seventh International Conference on Language Resources and Evaluation*, pp. 2530-2536.
- [4] Dukes, K. and Buckwalter, T. (2010, March). A Dependency Treebank of the Quran using Traditional Arabic Grammar. Presented at *Proceedings of The 7th International Conference on Informatics and Systems (INFOS)*, Cairo, pp. 1-7.
- [5] Dukes, K., Atwell, E. and Habash, N. (2013). Supervised Collaboration for Syntactic Annotation of Quranic Arabic. *Language Resources & Evaluation*, issue 47, pp. 33–62.

- [6] Granath, S. (2009). Who Benefits from Learning How to Use Corpora? In K. Aijmer, (Ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins Publishing Company, pp. 47-66.
- [7] Hizbullah, N., Fazlurrahman, F. and Fauziah, F. (2016, October). Linguistik korpus dalam kajian dan pembelajaran bahasa Arab di Indonesia. Presented at *Proceedings of Konferensi Nasional Bahasa Arab (KONASBARA) 2*. Malang: Universitas Negeri Malang, Indonesia
- [8] McEnery, T., Hardie, A. and Younis, N. (Eds.) (2019). *Arabic Corpus Linguistics*. UK: Edinburgh University Press.
- [9] Mohajan, H. (2018). Qualitative Research Methodology in Social Sciences and Related Subjects. *Journal of Economic Development, Environment and People*, vol. 7, issue 1, pp. 23-48.
- [10] Wahba, K. M. (2018). Materials Development in Arabic Language Learning and Teaching: Realities, Needs, and Future Directions. In K.M. Wahba, *et al* (Eds.), *Handbook of Arabic Language Teaching Professionals in the 21<sup>st</sup> Century: Volume II*. New York: Routledge.
- [11] Whitcomb, L. and Alansary, S. (2018). Using Linguistic Corpora in Arabic Foreign Language Teaching and Learning. In K.M. Wahba, *et al*. (Eds.), *Handbook of Arabic Language Teaching Professionals in the 21<sup>st</sup> Century: Volume II*. New York: Routledge.
- [12] Zaki, M. (2017). Corpus-Based Teaching in the Arabic Classroom: Theoretical and Practical Perspectives. *International Journal of Applied Linguistics*, vol. 27, issue 2, pp. 514-541.