SOURCE-BASED ARABIC LANGUAGE LEARNING: A CORPUS LINGUISTIC APPROACH

Nur Hizbullah[1*], Zakiyah Arifa[2], Yoke Suryadarma[3], Ferry Hidayat[4], Luthfi Muhyiddin[5], Eka Kurnia Firmansyah[6]

[1*]Universitas Al Azhar Indonesia, Jakarta, Indonesia; [2]UIN Maulana Malik Ibrahim, Malang, East Java, Indonesia; [3]Universitas Darussalam Gontor, Ponorogo, East Java, Indonesia; [4]Pondok Modern Tazakka, Pekalongan, Central Java, Indonesia; [5]Universitas Darussalam Gontor, Ponorogo, East Java, Indonesia; [6]Universitas Padjadjaran, Bandung, West Java, Indonesia.
Email: [1*]nurhz@uai.ac.id, [2]arifazakiyah@uin-malang.ac.id, [3]yoke.suryadarma@unida.gontor.ac.id, [4]hidayatferry02@yahoo.com, [5]luthfimuhyiddin@unida.gontor.ac.id, [6]eka.kurnia@unpad.ac.id

Abstract

Purpose: The study explores the process of using Arabic websites for Arabic language learning, utilising the Arabic Corpus Linguistic approach. This approach enables data-mining out of websites, systematically compiling the mined data, as well as processing the data for the express purpose of Arabic language teaching including its clusters, such as Arabic pragmatics, Arabic linguistics, and Arabic translation teaching as well.

MethodologyThe research is written descriptively and utilises qualitative methods used for analysing the process and step-by-step procedures to be executed to make good use of the data.

Main Findings: This study is conducted based on the theory of source-based teaching, while the process of utilising the websites is systematically elaborated through the corpus linguistic mechanism. The research concludes that almost all Arabic websites can be employed to be authentic, reliable teaching sources. The sources can be made good use of for teaching the four language competencies, for being the object of linguistic studies and for translation through the particular use of websites whose contents are bilingual or multilingual.

Implications/ Applications: The utilisation of the Corpus for teaching and learning has still been needing wide-spreading and promoting either among practitioners or among researchers of the Arabic language in Indonesia.

Novelty/Originality of this study: This study highlights that almost Arabic-language websites are one of the richest sources of learning. These learning resources can be used for language learning and various other dimensions of scientific Arabic. Corpus linguistics has many benefits for learners and teachers in Arabic language learning. This study gives the new approach of Arabic teaching-learning using website resources, and the dynamic of Arabic learning using technology.

Keywords: *Arabic Language in Indonesia, Source-Based Teaching, Arabic Websites, Arabic Corpus Linguistics.*

INTRODUCTION

The dynamics of teaching and learning in modern Arabic is increasingly rapid, along with the development of information and communication technology (Al-khresheh, Khaerurrozikin & Zaid, 2020). Various conventional approaches that are often used even relied on learning Arabic in the past have slowly begun to shift with the presence of a variety of cutting-edge approaches that are creative, innovative, even revolutionary (Hamed, 2012; Mansouri & Mhunpiew, 2016; Usman, 2013). It has a big impact on the overall process of learning Arabic starting from the preparation of the curriculum to the evaluation process (Hilao, 2016; Lubis, 2009). As a result, there has been an overall and continuous update of the Arabic language learning system.

This innovation does not only occur at several levels of Arabic language education but also covers all levels, especially the level of higher education or university. Along with the increase in maturity, learning ability, and academic abilities of students, the development of an Arabic learning system at a higher level is more open and complicated, when compared to learning at a lower level (Kasem, 2016). This allows teachers to be able to promote if they want to develop Arabic learning. For Example, one aspect that can be updated or developed is a learning resource. So far, the books have become the main source of learning and perhaps, in some places and times, the books taken are irreplaceable. Independent to learn Arabic without books as a source. However, now, with various learning resources and problems in the world of Indonesian graphics, the role of books begins to be rivalled or even shifted. On the one hand, it might make anyone sad, but on the other hand, technology-based learning resources are also like a favour that must be grateful and utilised (Hamed, 2012; Lubis, 2009; Meidrina, Mawaddah, Siahaan, &Widyasari, 2017).

The use of technology in learning began to be developed along with the development of technology in the era of revolution 4.0 Afrianto (2018) One of the sources of learning Arabic that should be taken into account because it is available in large numbers and broad coverage is an Arabic-language website (Ghani, 2015; Malroutu, 2017). Various webs containing news, scientific writings, popular features, cartoons and so on can now be found in an almost uncountable number, but are easily accessed by fingers, through a sophisticated technology media such as cellphones and tablets (Halimah, Ibrahim, & Lustyantie, 2018; Chadyiwa&Mgutshini, 2015). Both the teacher and the learner

should see this as an opportunity and its benefit, thus giving a new spirit and mutual commitment to improving the quality of learning Arabic in the digital era (Afrianto2018).

In concordance with this phenomenon, that Corpus is a systematically electronic collection of text, which came to be utilised in language teaching and learning (Boulton& Cobb, 2017; Vyatkina& Boulton, 2017). The Corpus provides original and potentially rich and interesting material, there is a fundamental discrepancy between the textual nature and the discoursal and communicative nature of language learning (Braun, 2007; McEnery & Xiao, 2011). These corpora contain textual records of discourse situations and recontextualisation of these records which is important if anything is to be learned from them, could be difficult, and requires a pedagogical source (Taher, Shrestha, Rahman, & Khalid, 2016; Widdowson, 2003). Corpus linguistics is one of technology-based that can be useful in learning and teaching, and in the last 30 years, the use of it in the classroom has started to increase and develop (Dazdarevic& Fijuljanin, 2015).

The learner corpus might be used as a new resource which is recently bringing learner language back into focus and enjoying growing interest from the language learning community at large. Among the many educational application by using this Corpus that would potentially benefit from learner corpus informed insight. A few could boast several concrete achievements: pedagogical lexicography, courseware, and language assessment(Granger, 2008; McEnery & Xiao, 2011; Pradhan, 2016).

The corpus linguistics has some advantages for users; it produced word lists and counts occurrences of individual search items. It allows for the presentation and reorganisation of data in a way that facilities the identification of patterns, it automatically produced clusters and collocation lists (Farr, 2008). Many research has been conducted to investigate the effectiveness of using corpus linguistics as a teaching technique, e.g to highlight the native speaker use the natural language form, vocabulary items, and expression (Almutairi, 2016; Donesch-jezo, 2013; Roca Varela, 2012).

Research Gap and objectives

These studies and research provide the positive of the usage, utility, significance, and the benefit of corpus-based in language learning. Therefore the study of corpus-based in Arabic learning potentially needs to explore and explain to be applied and practically used in classroom activities.

The objective of the study

 this study intends to how does the Arabic sources can be utilised in linguistic corpus format and how does the process of using Arabic-language websites for language learning using the Arabic linguistic approach. This approach allows the collection of data from the website, compiles it systematically, and processes the data for learning Arabic in various sub-fields, such as language skills, linguistics, and translation.

The finding of this study will redound to the benefit of Arabic learning sources, considering that linguistic Corpus has an essential role in technology and education. The demand for utilising technology and using big data the need for a more life-changing learning approach. Thus, the teachers and learners that apply this recommended approach of this study will be able to explore the data-mining more. For researchers will help them to explore the more educational resources and technological processes.

LITERATURE REVIEW

A corpus describes how language is used in a real situation and makes an end to the so-called needs of relying on a native speaker institution to tell what is commonly used in language, and the computer enables to prove corpus search rapidly (Dazdarevic & Fijuljanin, 2015). The kinds of literature that are underlying this research are the studies of source-based learning. Some studies show the utilisation of Corpus to source in learning. The use of corpus-based in a language classroom began when (Vyatkina& Boulton, 2017) was published a direct application of corpus-derived examples in language learning and the same decade was followed by a few other researchers. Johns in the lates 1980s - 1990s identified a distinct teaching method and a distinct subfield of teaching research by drawing computer science and improving them namely data-driven learning (DDL) and outlined its benefit in the pedagogical field. Recently, not only research journal published and reported this field, but also theses and dissertations, online publication, and article conference proceedings. Some journals have also dedicated special issues to DDL.

Several published research illustrates the utility of the DDL application. It has been applied in foreign and second language teaching learning with learners at different competency levels in many countries and many institutions. It has included the work of teacher prepared corpus-derived materials and direct corpus searches by learners, and has focused on various linguistic elements; vocabulary, grammar, pragmatic, and discourse. Furthermore, Boulton & Cobb[9] have reported an effective and efficient approach of quantitative studies trough DDL in language learning used meta-analysis.

Braun (2007) applied an empirical case study conducted to investigate the integration of corpora and corpus-based activities into language learning. The positive finding emerging from the study create optimism especially concerning the students' interest and adaptability and original corpus materials. Ellis (2017) approaches the study by the perspective of usage-based linguistics that describes the essential contribution trough experimental, computational, and corpus-based study to the establishment of usage-based theories of language learning. Rebuschat, Meurers and McEnery (2017) illustrate how language learning research use corpus-based approaches.

Moreover, the Corpus also used in peer-assessment of the development of spontaneous interactive speaking skills at English class, the researchers reported that the Corpus encouraged participants to develop three strategies; willingness to improve, use of compensatory strategies, and construction of personalised version of the Corpus. On the other hand, the Corpus produced the emergence of detrimental traits; underassessment and dependency on the Corpus (Mily& Sará, 2015).

Additionally, Almutairi (2016) investigated the effectiveness of a corpus-based approach to language description in writing skills. The study has the pedagogical implication of using corpus-based activities and provides the weakness and the strengths of the Corpus as a resource in language learning, and it proved that a corpus-based approach to language learning and teaching can integrated help both the learners and teachers to decide and explore authentic language usage. The potential corpus-based source can investigate grammatical and lexical patterns as well as proven and established. The technological and advancement can be popularised by facilitating and motivating the benefits and advantages of using it (Roca Varela, 2012).

In the case of using corpus-based in Arabic learning, Ghani et.al (2015) reported the research about the effectiveness of using the website in Arabic learning, they used case study among the learners in Arabic for tourism. The research showed evidence of using the website in Arabic learning is effective and efficient and provides the new technology that represents valuable support for the learners in Arabic language learning. These studies and research above provoked the positive of the usage, utility, significance, and the benefit of corpus-based in language learning (Dazdarevic & Fijuljanin, 2015).

While those studies show the benefit of using the website in Arabic learning but did not cleared describe the process of the of them, this research will explore the kind of resources that can be used and apply the process of utilising of Arabic website with linguistic corpus approach for learning resources.

METHOD

This research is applied research with two methods, namely the experimental method and the descriptive method. The data source is taken from the internet. The experimental method is used to test the making of a simple corpus sample with text material taken from the source on the internet.

The trial of making this Corpus uses an MS Word processing application with Unicode UTF-8 based file conversion techniques. In the next stage, the trial processing of the Corpus is done with a free Anthony (2019) application that is downloaded from the web http://laurenceanthony.net/software/antconc/. Using AntConc also can be used for learning corpus-based (Anthony, 2004).

The qualitative descriptive analysis used as the second step. The description of the stages of making the Corpus was carried out descriptively, as well as projections for the application of corpus-based learning in various fields of learning Arabic. The two steps can be simulated and integrated to explore and explain the use of corpus-based as a source in Arabic language learning.

RESULTS AND DISCUSSION

Website as a Source of Learning Arabic and its Classifications

In general, various types of Arabic-language websites are available in huge quantities in cyberspace. Some of them can be used as a source of learning Arabic in the context of language proficiency, ranging from listening skills, speaking, reading, writing, and translating. Various learning resource websites can be classified as follows:

1. General sources, which contain a variety of reference books, textbooks, and learning materials that can be accessed independently and free.

2. Special source for learning the Arabic alphabet.

3. The source of Arabic learning material Arabic Fusha and its grammar.

4. Arabic language sources based on dialects, such as Egypt, Levantine, Morocco, Tunisia, Gulf Arab, and so on.

5. Online monolingual and multilingual dictionaries.

6. Online news media.

7. Online television media.

8. Online newspapers and magazines.

9. Website for online music, film, and entertainment providers.

10. Online radio broadcasts.

11. Digital book provider website.

Various websites can be used as a source of material used in learning Arabic at various levels, ranging from elementary to high level. Referring to the format of the material available, these sources provide material in the form of written text, sound/audio material, and audiovisual material. The format of the material is in line with the corpus format which can be made based on existing formats, namely the text body, audio corpus, and audiovisual body. Furthermore, the process of trying to make a corpus will be described, along with a description of the steps taken.

The Techniques for Making Simple Corpus Files

Based on the classification of the Corpus above, there are three types of the Corpus, namely the body of text, audio corpus, and audiovisual Corpus. The following explanation is only focused on making the text body only. In this section, we will briefly describe the practical steps of making a simple corpus containing text taken from a website to be used in the learning process.

The Selection and Copying of Text

Teachers must first choose a relevant theme or topic from a website source. After the theme or topic is determined, then a text is selected with certain criteria. Like the format, the text body only contains writing or orthography and does not contain elements of an image. Therefore, what will be copied from the website source is only the text, including the title, the release date of the article, and the link to the source address in question. In particular, some parts of the text may need to be marked with square brackets close and open, for example, the release date of the text link and address link, to be used as metadata which is the basis of information about the file concerned and keywords for calling the data from the system retrieval on the corpus database.

Figure 1: The Examples of copying text from the website to the MS Word page with the inclusion of metadata on the page of the target file(File taken from Aljazeera News)

Source: Developed by authors

The certain text is copied from the source website then pasted on the word processing application page (MS Word and the like). In making the Corpus, the writing format on the word processing application page does not have to be like a standard format that is usually made for the benefit of other writings. That is because the corpus file that will be created will be completely different from the initial MS Word version format and so on. So, the text copied from the website can be directly affixed without having to change the font type, size, space, etc.

The Conversion *.doc format to *.txt

After the text in the document format is ready, especially the file is stored in the form of *.doc or *.docx according to the standard document file format. After being saved as a document, the file is redirected in plain text format. The steps are as follows.

a. Use the "Save As" facility on the File menu in the MS Word application

b. Put the file in the destination folder for storing the intended file

c. Name the file, for example, "sample corpus 1", then in the "Save as type" column, select the type "Plain Text".
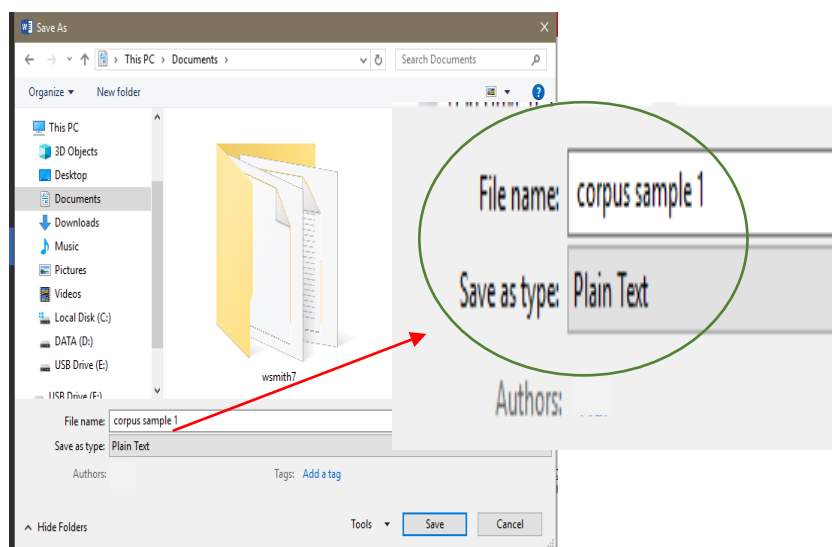
Figure 2: The Examples of placement of folders for file storage, naming, and selection of file types that will be used as examples of corpus

Source: Developed by authors

After clicking the "Save" button, the following file conversion display will appear. To complete the text encoding process, select the type of Unicode encoding (UTF-8) found in the "Other encoding" menu. The use of this type of encoding is indeed following with the standards for making corpus files that can be adapted and processed by most corpus processing applications, such as WordSmith, AntConc, and so on.
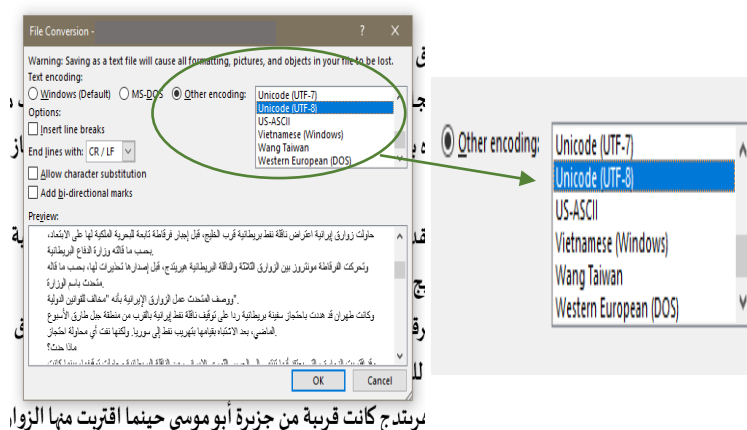

Figure 3: The process of converting document files into "Plain Text" using Unicode (UTF-8) text encoding

Source: Developed by authors

After the process is complete, close the document file from the MS Word application. The conversion process has been completed and the next will be seen the results of the intended file conversion.

The Files that have changed the format to "Plain Text" will also change the logo, different from the original logo, and when the file is still in document format, as in the File Explorer view as follows:
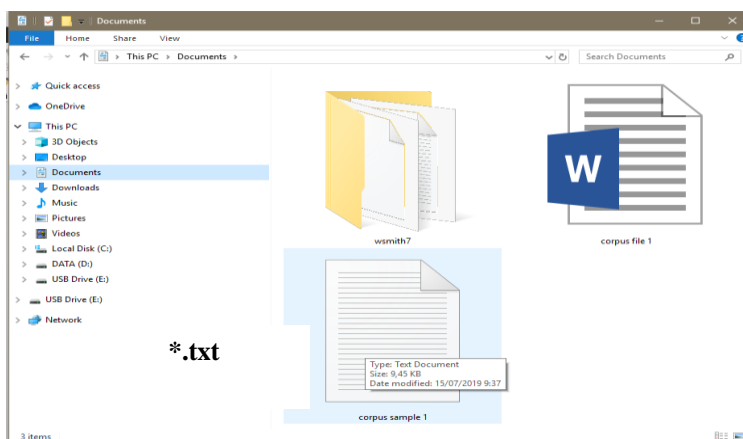
Figure 4: Change the document file to "Plain Text" which is marked with a change in the file type logo

Source: Developed by authors

Open the "corpus sample 1" file from the previous conversion by clicking on the "corpus sample 1" file that has changed to "Plain Text". The contents of the file will look like the following picture.
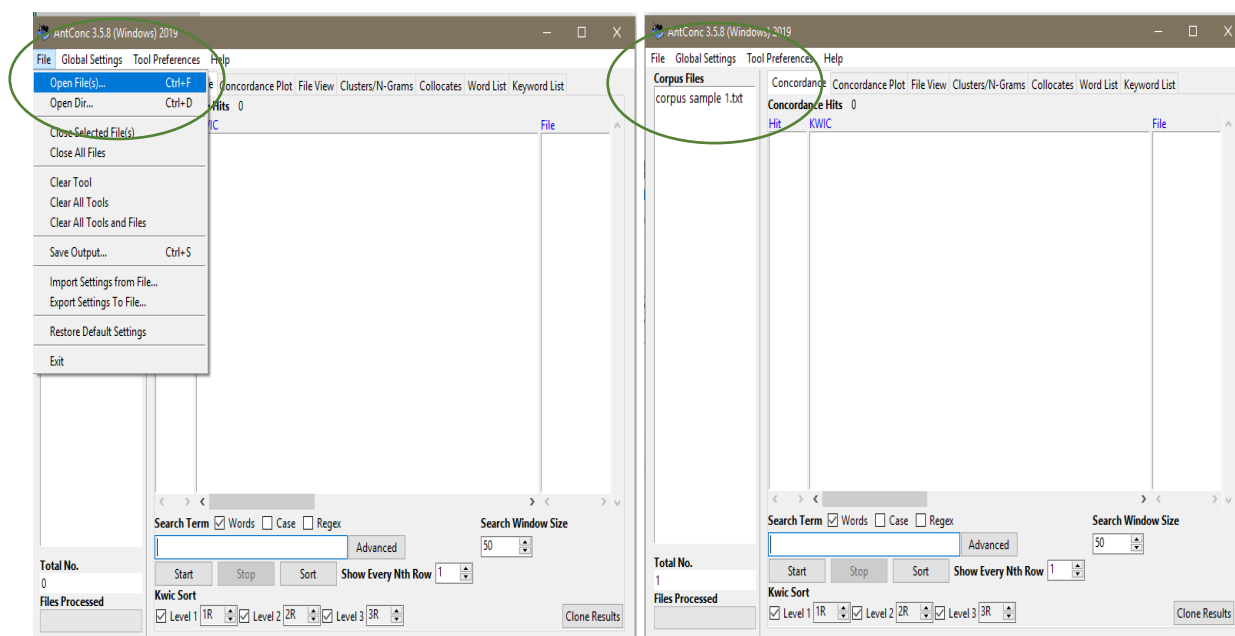

Figure 5: Display the contents of the "Plain Text" file that is opened through the Notepad application

Source: Developed by authors

The Example of processing Corpus with a free application "AntConc"

After the sample corpus file is created through the MS Word application and standard conversion mechanism by using UTF-8 text encoding as described in the previous step, the following process and corpus processing mechanism will be described below with a simple AntConc application.

As brief information, AntConc is one of the many simple corpus processing applications that can be downloaded free from the website http://laurenceanthony.net/software/antconc/. This application is available for basic Windows, Mac, and Linux operating systems. Especially for Windows, there are two choices of applications by 32-bit or 64-bit system types. Applications can be directly downloaded and used without installation or set up on a computer.

After the application file has been downloaded and opened, firstly, the corpus file to be examined is opened in the application via the "Open File(s)" submenu from the "File" menu. Next, look for a corpus file that is formatted *.txt, clicked, and opened. The process of opening a new file will be declared successful after the corresponding *.txt file name appears in the "Corpus Files" column.

Figure 6: AntConc application page views in the process of entering corpus files to be processed

Source: Developed by authors

After the corpus file has been successfully entered and opened in the application, here is an example of processing the Corpus-based on several menus in the application, such as word list, collocation, N-grams, and concordance.

Wordlist

The word list is a list of all words contained in a text or Corpus which generally become a database for the compilation of dictionaries. This list of words can be sorted alphabetically as well as the frequency of occurrence in the text (Baker, 2006).

a. To compile a list of words from the Corpus whose files have been made in *.txt format and have been entered into the application, the "Word List" menu can be selected.

b. After the menu is selected, see the "Search Term" submenu in the lower left and check the "Words" selection box then click the "Start" button below it.
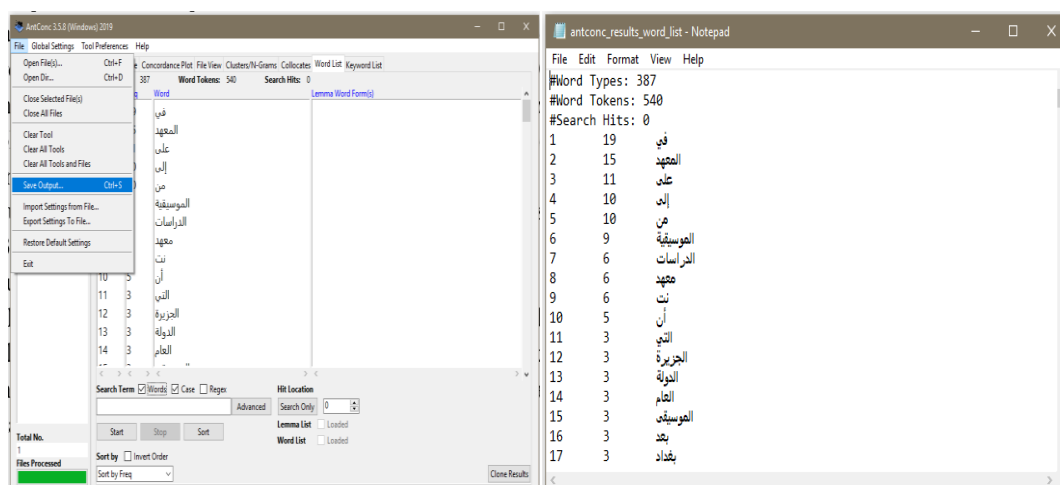

Figure 7: The Display of word list sorting method in the AntConc application

Source: Developed by authors

c. For sorting this list of words, there are several choices of methods available, which are based on frequency (many and vice versa, word [alphabetical], and letters at the end of the word [alphabetical]). There is an additional choice of other menus for sorting, whether to be sorted from the beginning of the back/end, namely in the "Invert Order" checkbox. If sorting is selected based on frequency from a lot to a little, without "flipping" the sequence by checking the "Invert Order" menu along with the display of results given by the AntConc application.
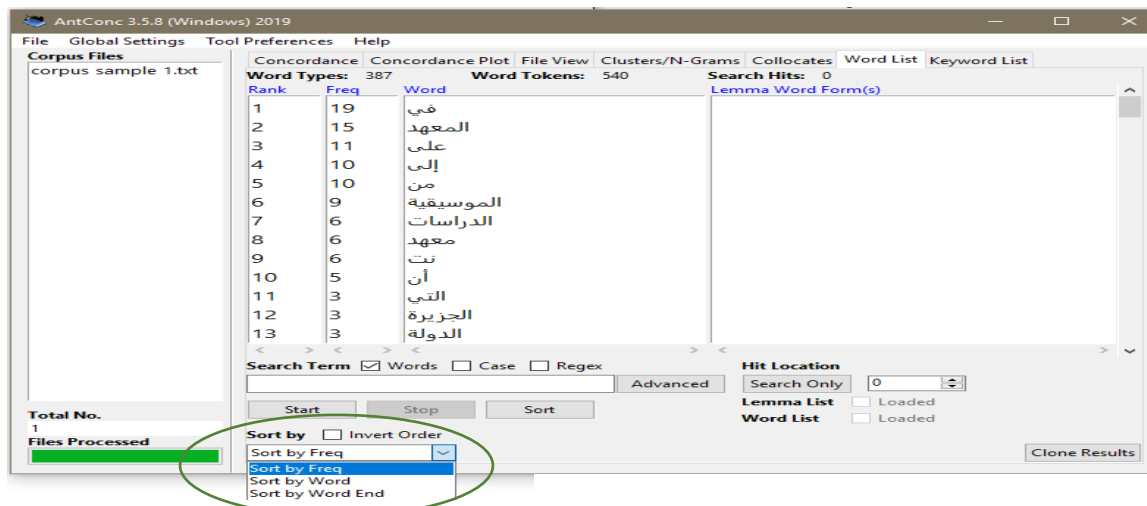
Figure 8: Display the results of sorting a list of words based on frequency (many-few) in the AntConc application

Source: Developed by authors

d. The sorting results are then saved into a separate file. According to the standard AntConc application, the file will be saved in the form *.txt. The storage step starts by opening the "File" menu and then clicking on the "Save Output ... submenu". The file is then stored in a folder that can be selected by the user. The application's default file name is "antconc_results" and we can modify the name ourselves. The following is the display of the contents of the file stored in the list of words obtained from the AntConc application.
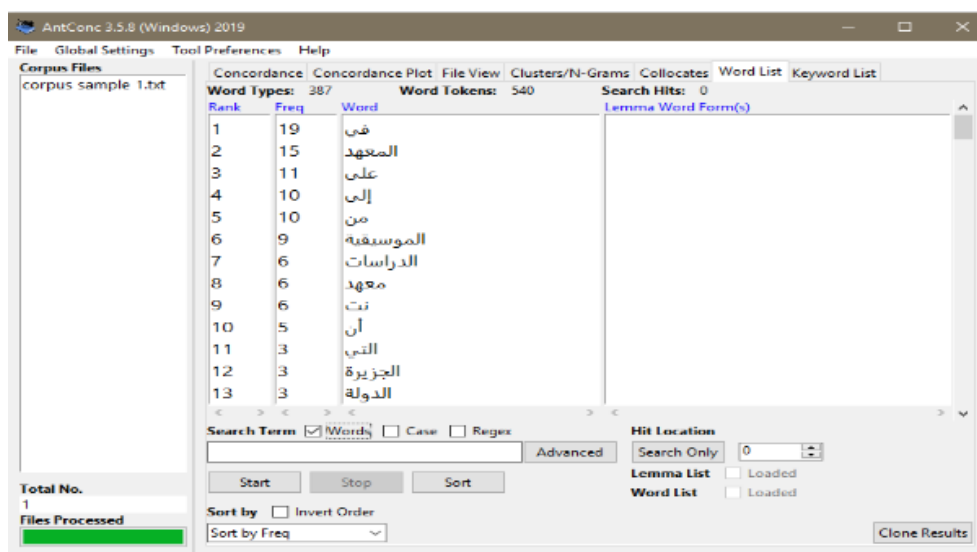

Figure 9: The Display of the contents of the sorted file word list on the Corpus through the AntConc application

Source: Developed by authors

Collocation

Collocation is the phenomenon of the appearance of words that are coupled with certain other words in a context and field of meaning. In a corpus processing application system, collocation can be stretched from two-word pairs and/or more (Baker, 2006).

a. To compile a collocation list of corpus files that have been created, use the "Collocates" menu in the application.

b. After the menu is selected, for example, the word to be searched for from the corpus file is the preposition or particle "في" / fī /. The trick is to type the word in the column under the "Search Term" submenu in the lower left by checking the "Words" selection box and then clicking the "Start" button below it.

c. For the display of the order of search results, just like the display of word list search results, there are several choices of methods available, namely based on frequency (many and vice versa) and the word (alphabetical). Another menu that can be used also is "Invert Order" to make a sequence from the beginning of the numbers or alphabetically and vice versa.
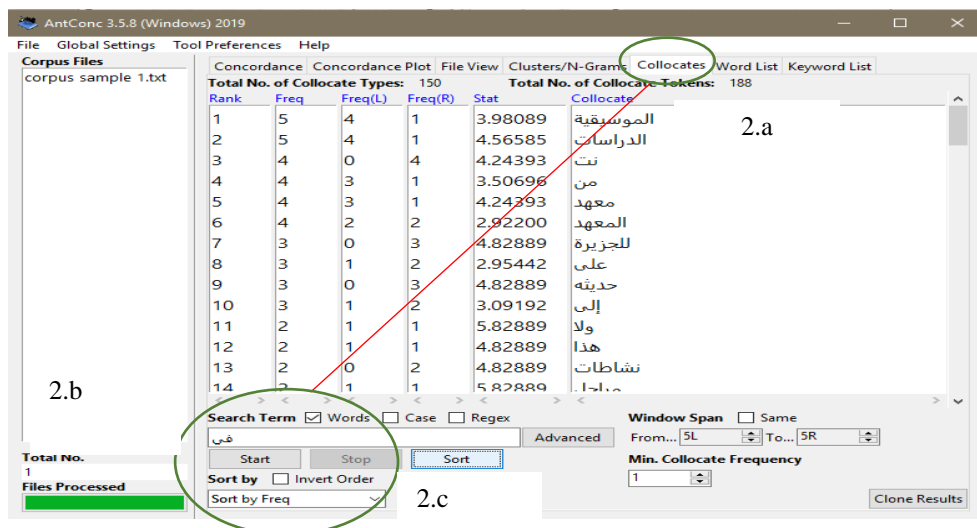
Figure 10: The display of collocation search file words "في" / fī / from the corpus file

Source: Developed by authors

In the display of the image, there is some special information in the frequency column followed by certain information, namely (L) and (R). Description "Freq (L)" indicates the frequency of occurrence of the word "الموسيقية" / al-mūsīqiyyat / as much as four times on the left (left) of the particle or preposition "في" / fī / in corpus data and one time to the right of the preposition. The total number of occurrences of collocation of the words "الموسيقية" / al-mūsīqiyyat / is as many as five times, according to what is shown in the "Freq" column on the left, and so on for other collocation words like those found in the Corpus.

d. Collocation search results can also be saved as separate files in the form of *.txt. The storage step is the same as in the process in number 1.d by opening the "File" menu then clicking on the "Save Output ... submenu". The file is then stored in a folder and named "antconc_results ..." by modifying the name. The following display of the contents of the collocation storage word "في" / fī / is obtained from the AntConc application.
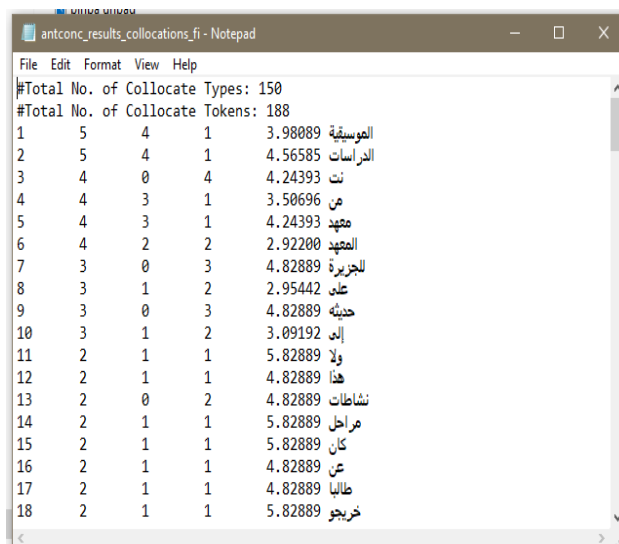

Figure 11: The Display of the contents of the collocation sequencing file on the Corpus through the AntConc application

Source: Developed by authors

N-Grams

N-grams are a sequence of two or more words that appear repeatedly in a text or body with a significant amount to be examined with certain assumptions (Baker, 2006).

a. To compile a list of n-grams of a corpus file, in the AntConc application the "Clusters / N-Grams" menu is used.

b. After the menu is selected, related to the "Search Term" menu, select the "N-Grams" column that is checked and other options in the column on the left are not active. After that, click the "Start" button to start searching n-grams. In the N-Gram Size menu, there are minimal range options and a maximum number of pairs or constant sequences appear in the text. The size can be modified depending on how many cluster members you want to study. The following is a search for the two-word pair you want to look for constantly.

c. To display the order of search results, there are several choices of methods based on frequency and words (alphabetical) with other menu options, namely "Invert Order" to make a sequence from the beginning of the number or alphabetically and vice versa.
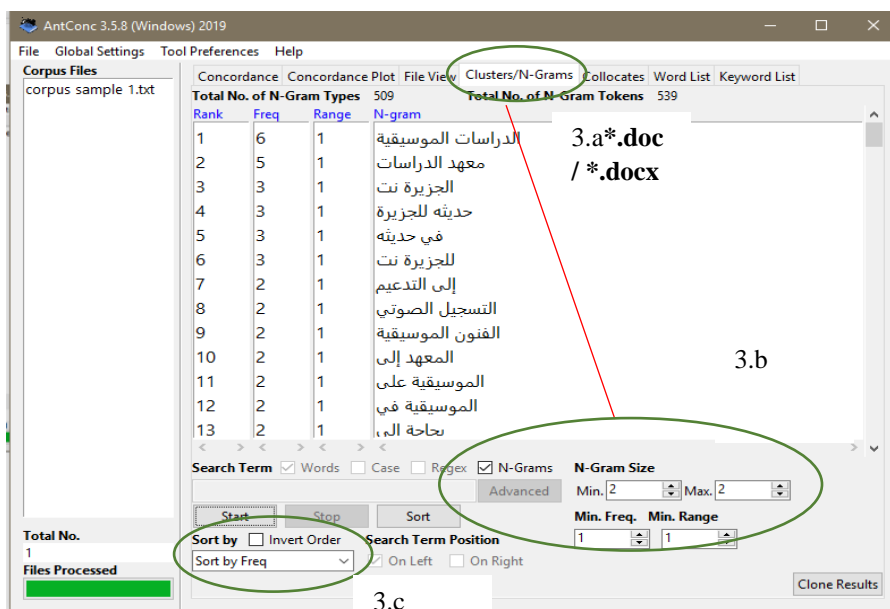

Figure 12: The display of n-grams sequence results file from the corpus file

Source: Developed by authors

d. Search results for n-grams can also be saved as *.txt files. The storage step is the same as the process in numbers 1.d and 2.d by opening the "File" menu and then clicking on the "Save Output ..." submenu. The file is then saved and named "antconc_results ..." by modifying the name. The following shows the contents of the n-grams file obtained from the AntConc application (Anthony, 2019; Rawi, 2017).

Figure 13: The display of the contents of the n-grams sorting the file on the Corpus through the AntConc application

Source: Developed by authors

Also, related to the storage of word list (figure 9) search results, collocation (figure 11), and n-grams (figure 13) in plain text (*.txt) format, the contents of the file can still be moved and processed again with other applications such as number processing (MS Excel and the like). This will make it easy for users or researchers to process data with various models of sequences available in a number processing application. Thus, researchers can carry out various approaches to further analyse data that has been obtained from the corpus file. The following is an example image of the results of moving the contents of a text file (*.txt) to a number processing page (MS Excel).
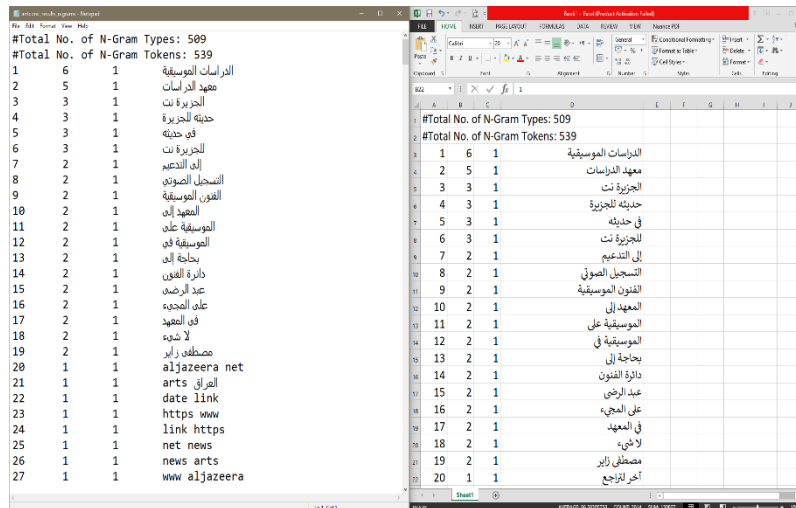

Figure 14: The Display of the results of the transferor copy of the contents of the n-grams file to the MS Excel application

Source: Developed by authors

Concordance

Concordance is a list of occurrences of a word that is in a certain "environment" context. Elements of this context can be identified by looking at several words that are in the "around" (before and after) words that are being analysed (Baker, 2006).

a. To see the concordance of a word, use the "Concordance" menu in the application.

b. After the menu is selected, for example, the word to be searched for is the preposition or particle "في" / fī /. The trick is to type the word in the "Search Term" column by checking the "Words" selection box then clicking the "Start" button below it.

c. To help identify contexts, users can modify the limit to how many words will be a reference around the left and right words to be analysed. The trick is to determine how many words in each of the three levels available on the KWIC sort menu that will be limited to identifying the context. For Example, the following particle concordance search "في" / fī / will be carried out by observing three words around the left and right of the intended particle. The application then puts the analysed word in the middle of the word line, while the third word on the left is marked with red writing and the three words on the right are marked with purple. Thus, researchers can be helped in identifying the context by limiting the range of the concordance.
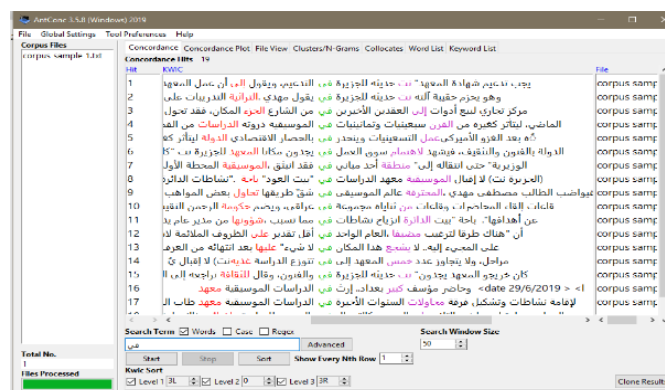

Figure 15: The display of the concordance search result file word "في" / fī / from the corpus file

Source: Developed by authors

d. The concordance search results can also be saved as *.txt files. The storage step is the same as in the previous process by opening the "File" menu then clicking on the "Save Output ... submenu". However, unlike other plain text format results that are relatively regular and easily copied or moved to files that will be processed with applications such as MS Excel, the concordance output results look irregular on the Notepad application page. As a result, the results in the *.txt format are difficult to process further into table-based applications such as Excel and others. Another obstacle is the irregularity of the position of the text which should start from right to left according to the rules of Arabic writing. AntConc still has weaknesses in this matter and it is risky to be relied upon to do concordance analysis because of these fairly basic technical constraints (Anthony, 2004). For concordance analysis, it's good for researchers to look for alternative applications that are more reliable.
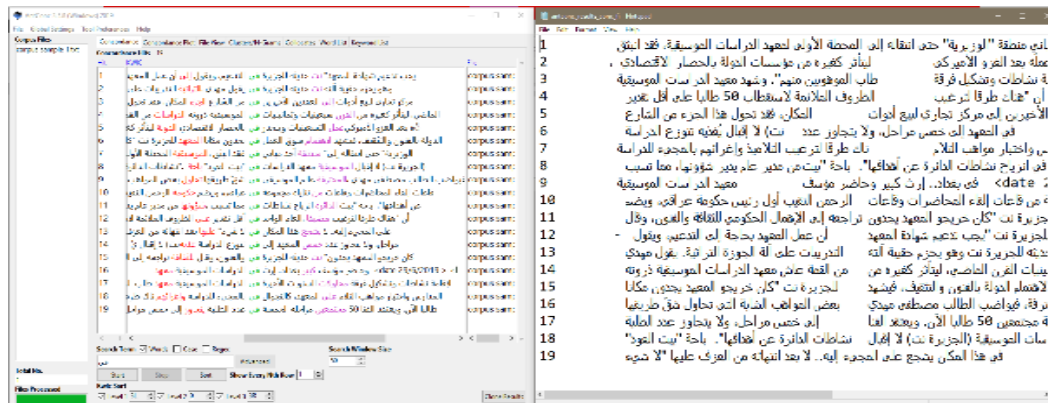

Figure 16: The comparison of page views of the AntConc version (left) and the save output results as plain text in the Notepad application

Source: Developed by authors

The Orientation of Utilising the Corpus in the Arabic Scientific Fields

The use of authentic and real-life examples with language learners is more crucial than examples that are made up by the teacher and simulate real-life use of language. The corpus linguistics can be integrated into the language classroom for language learners at various levels (Anthony, 2019; Dazdarevic & Fijuljanin, 2015). The use of the Corpus for learning, although it is quite popular in many circles, still needs to be socialised and encouraged among teaching practitioners and Arabic language researchers in Indonesia. As a new thing, it is only natural that this approach still needs time to be known, learned, understood, and used in the learning process. Hizbullah, et.al has studied the rich sources of Arabic Corpus to be made in Indonesia (Hizbullah& Muchlis, 2017; Hizbullah et al., 2019) contains seven distinct classifications by the availability of data in the Arabic field, they are 1.Scientific works Corpus, 2.Popular works Corpus, 3. Literary works Corpus, 4. Religious works Corpus, 5. Quranic Corpus and its translation, 6. Corpus of loan words from Arabic-Indonesian, 7. Learner Corpus (Alfaifi, 2015; Alansary, Nagi., & Adly, 2007).Therefore, the following is a description of the use of the Corpus in general in various fields of Arabic language with possible approaches or analysis in research and product projections that can be produced from the use of the Corpus for the benefit of learning Arabic.

Table 1: The Utilisation of the Corpus in the scientific fields of Arabic

No	Scientific Field	Corpus Function	Approach / Analysis	The Project of Product
1	Language Skills	as a recording and documentation of the appearance or language skills of Indonesian teak speakers as foreign learners of Arabic	- analysis of language skills -Synchronous and diachronic analysis	-student corpus, oral and written - language error data - learner dictionary
2	Language teaching	Teaching the language as real and factual data for teachers about the situation and language skills of the learners	- contrastive analysis -analysis of the development of language skills -Synchronous and diachronic analysis	- student campus - textbooks -accompanying textbook material
3	Linguistics	as data to help researchers identify and analyse certain patterns in various levels of language ranging from sound to discourse.	- linguistic analysis -Synchronous and diachronic analysis	- publication - general and specific dictionaries - big data

No	Scientific Field	Corpus Function	Approach / Analysis	The Project of Product
4	Translation	as data to help researchers identify certain aspects of the text in the source language and target language	- contrastive analysis - comparative analysis	-parallel corpus multilingual translation -multilingual dictionary
5	Culture	as data to help researchers analyse vocabulary, symbols, or linguistic expressions related to culture	- discourse analysis - comparative analysis -Synchronous and diachronic analysis	- publication -dictionary of languages and cultures
6	Literature	as data to help researchers analyse vocabulary, symbols, or linguistic expressions related to literature	- comparative analysis -Synchronous and diachronic analysis	- publication - literary dictionary

In line with Fligelstone (Fligelstone, 1993) describes the application or Corpus in language teaching-learning, the use of corpus tools and methods in its context, a useful distinction could be made between direct and indirect application. The indirect applications used in hands-on for researchers and material writers; effects on the teaching syllabus and effects on reference works and teaching materials. The direct application used in hands-on for teachers and learners (data-driven learning, DDL); teacher-corpus interaction, and learner-corpus interaction.

Hassan, Mat Daud, Atwell (2013) also used this approach in their study for Arabic language learning that showed that the material writers and teachers can write the textbook based on data sourced from a corpus, as it provides information on authentic language use which make teaching more relevant and useful to the learners of the language. The corpora can also be used to explore the learning activities by discovery learning in the classroom. The study also suggested the need to popularise the data-driven approach (Ghani et al, 2015). Furthermore, Ghani et al. (2015) used the Corpus to estimate the difficulty of reading Arabic texts. They found the formula of difficult text that can be easily understood by a language teacher as the argument is based on a language formula. This formula can be useful for the teacher to decide relevant text for Arabic learning levels.

Moreover, Dazdarevic and Fijuljanin(2015) explain some advantages of utilising the corpus based-approached in language learning are; (1) the learners eventually will be able to formulate their exploration of particular language use, (2) the learners will acquire the form of learned language in case they are engaged in exploring the real use of language based on authentic content, (3) it gives the independence to learners by providing them with computer assistant to answer the question during their learning, (4) the learners can be more active than they decide their own rule by scrutinising the concordance of any problematic vocabulary and grammatical items, whereas the teacher does not make the rule instead guide the learners to think and conclude more effectively.

CONCLUSION

This study concludes that almost all Arabic-language websites are one of the richest sources of learning. These learning resources can be used for language learning and various other dimensions of scientific Arabic. On the one hand, corpus linguistics has many benefits for learners and teachers in Arabic language learning. On the other hand, using corpus linguistics in Arabic language learning requires some technical in using a computer program, software and internet. The result of this study gives the new approach of Arabic teaching-learning using website resources, and the dynamic of Arabic learning using technology. Furthermore, linguistic Corpus offers a new type of research approach in language learning. Hopefully, this approach can be an innovation for the development of learning Arabic in Indonesia. More research is necessary to describe and examine the practice and the effectiveness of Arabic learning using a corpus-based approach.

LIMITATION AND STUDY FORWARD

This study has some limitations which must not be ignored and addressed in future studies. Though the current study advocated the use of corpus learning for Arabic language learning, this method has some constraints which must be taken into account while adopting this method and narrating the results. Authors would encourage researchers to search for ways through which the retrains offered by the corpus learning method can be overcome. A combination of mixed methods might enhance the effectiveness of Arabic learning. Since Arabic teaching and learning as a foreign language is extremely relen=vant in a current era; therefore, more research in a similar domain are highly encouraged and appreciated.

AUTHORS CONTRIBUTION

Nur Hizbullah, Zakiyah Arifa, and Yoke Suryadarma worked on the first phase of data collection from the internet and the formation of the Corpus. Ferry Hidayat and Luthfi Muhyiddin worked on the second phase of data analysis. Eka Kurnia Firmansyah drafted the initial idea, refined it, and completed the write-up as per journal guidelines.

REFERENCES

1. Al-khresheh, M., Khaerurrozikin, A.,& Zaid, A. (2020). The efficiency of using pictures in teaching speaking skills of non-native Arabic beginner students. *Universal Journal of Educational Research, 8 (2).* https://doi.org/10.13189/ujer.2020.080318
2. Afrianto (2018). Being a Professional Teacher in the Era of Industrial Revolution 4.0: Opportunities, Challenges and Strategies for Innovative Classroom Practices Afrianto Faculty of Teachers Training and Education (FKIP), Universita. *English Language Teaching and Research*, *2*(1), 1–13. https://doi.org/10.31258/ijebp.v1n1.p3-18
3. Alansary, S., Nagi, M., & Adly, N. (2007). Building an International Corpus of Arabic (ICA): progress of compilation stage. *Int. Conference on Language.*
4. Alfaifi, A. Y. G. (2015). Building the Arabic Learner Corpus and a System for Arabic Error Annotation," Imam Muhammad bin Saud Islamic University, Riyadh, Saudi Arabia.
5. Almutairi, N. D. (2016). The Effectiveness of Corpus-Based Approach to Language Description in Creating Corpus-Based Exercises to Teach Writing Personal Statements. *English Language Teaching*, 9(7), 103. https://doi.org/10.5539/elt.v9n7p103
6. Anthony, L. (2019). AntConc. Faculty of Science and Engineering at Waseda University, 2019.
7. Anthony, L. (2004). AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit. *Proceeding IWLeL 2004: An Interactive Workshop on Language e-Learning.*
8. Baker, P., Hardie, A., & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
9. Boulton, A., & Cobb, T. (2017). Corpus use in Language Learning: A Meta-Analysis. *Language Learning*, *2*(67), 348–393. https://doi.org/10.1111/lang.12224
10. Braun, S. (2007). Integrating Corpus Work into Secondary Education: From Data-driven Learning to Needs-driven Corpora. *ReCALL*, 19(3), 307–328. https://doi.org/10.1017/S0958344007000535
11. Chadyiwa, M., & Mgutshini, T. (2015). Using mobile handheld devices as tools of learning and teaching for Student EHPS: A blessing or a curse? International Journal of Humanities, Arts and Social Sciences, 1(2), 85-91.
12. Dazdarevic, S., & Fijuljanin, F. (2015). Benefits of Corpus-Based Approach To Language Teaching. *Baulk. Distance Educ. Netw. - Bad. News Lett.*
13. Donesch-jezo, E. (2013). Using Language Corpus in Teaching Foreign Language Vocabulary, 11–25.
14. Ellis, N. (2006). Cognition, Corpora, and Computing: Triangulation Research in Usage-based Language Learning. *Language Learning*, 67(1), 41–66, https://doi.org/10.1111/lang.12215
15. Farr, F. (2008). Evaluating The Use of Corpus-based Instruction in A Language Teacher Education Context: Perspectives from the Users. *Language Awareness*, 17(1), 25–43. https://doi.org/10.2167/la414.0
16. Fligelstone, S. (1993). Some Reflections on The Question of Teaching, from A Linguistics Perspective. *ICAME Journal1*, *17*(1), 97–109.
17. Ghani, M. T. A., Daud, W. A. A. W., & Sahrir, M. S. (2015). The Effectiveness of Using Website in Learning the Arabic Language for Tourism : A Case Study among Students at College Poly-Tech MARA (KTM). *International Conference on e-Learning.*
18. Granger, S. (2008). Learner Corpora in Foreign Language and Education. *Encyclopedia of Language and Education*. https://doi.org/10.1007/978-0-387-30424-3_109
19. Halimah, H., Ibrahim, G. A., & Lustyantie, N. (2018). Cello As a Language Teaching Method in Industrial Revolution 4.0 Era. *English Review: Journal of English Education*, *7*(1), 75. https://doi.org/10.25134/erjee.v7i1.1497
20. Hamed, A. (2012). The Utilisation of Technology in Teaching of the Arabic Language in Secondary Schools in Riyadh, Saudi Arabia. *Procedia - Procedia Social and Behavioral Sciences*, 64, 594–603. https://doi.org/10.1016/j.sbspro.2012.11.069
21. Hassan, H., Mat Daud, N., & Atwell, E. S. (2013). Connectives in the world wide web Arabic Corpus. *World Applied Sciences Journal,* 21(1); 67–72.
22. Hilao, M. P. (2016). Creative teaching as perceived by English language teachers in private universities. Journal

of Advances in Humanities and Social Sciences, 2(5), 278-286. https://doi.org/10.20474/jahss-2.5.4

23. Hizbullah, N., & Muchlis, M. M. (2018). Projected Characteristics and Content of Arabic Corpus in Indonesia," in *Advances in Social Science, Education and Humanities Research (ASSEHR)*. https://doi.org/10.2991/icclas-17.2018.42

24. Hizbullah, N., Mardiah, Z., Suryadarma, Y., Muhyiddin, L., Sofyan, O., & Hidayat, F. (2019). Arabic Learners' Corpora in Pesantrens for Developing Arabic Language Researches in Indonesia. *KnE Social Science*, *20*(19), 980–989. https://doi.org/10.18502/kss.v3i19.4922

25. Kasem, A.-H. (2016). Innovation in Language Education: Enriching Arabic Language Learner Experiences and Proficiency in in2D and 3D Environments. *IOSR Journal of Humanities and Social Science*, *21*(09), 40–46. https://doi.org/10.9790/0837-2109064046

26. Lubis, M. A. (2009). Teaching and Learning the Arabic Language through Multicultural Approach and Applying ICT in Pesantren Institution," *International Journal of Information and Education Technology*, *3*(3), 171–178.

27. Malroutu, Y. L. (2017). Enhancing student learning experience in blended classroom teaching. Journal of Advances in Humanities and Social Sciences, 3(6):324-331. https://doi.org/10.20474/jahss-3.6.4

28. Mansouri, S., & Mhunpiew, N. (2016). Leadership is skin deep: A new way of being through inside-out effect of leadership and its strategies in teaching. Journal of Advances in Humanities and Social Sciences, 2(3), 133-142. https://doi.org/10.20474/jahss-2.3.2

29. McEnery, R., & Xiao, T. (2011). What corpora can offer in language teaching and learning? In E. 191 Hinkel (Ed.), *Handbook of research in second language teaching and learning*.

30. Meidrina, T., Mawaddah, S. N., Siahaan, F. M. M., & Widyasari, P. (2017). The relationship between teacher efficacy and attitude towards inclusive education in private elementary school: A study based on teaching experiences. Journal of Advances in Humanities and Social Sciences, 3(3), 124-134. https://doi.org/10.20474/jahss-3.3.1

31. Miley, M., & Sará, G (2017). The Influence of Peer Assessment and the Use of Corpus for the Development of Speaking Skills in In-Service Teachers. *How*, 23(1), 103–128. https://doi.org/10.19183/how.23.1.142

32. Pradhan, S. (2016). English language teaching: The next gate to social awareness. International Journal of Humanities, Arts and Social Sciences, 2(4), 156-158. https://doi.org/10.20469/ijhss.2.20005-4

33. Rawi, M. (2017). Using Antconc: A Corpus-Based Tool, To Investigate and Analyse the Keywords in Dickens' Novel 'a Tale of Two Cities. *International Journal of Advanced Research*, 5(2), 366–372. https://doi.org/10.21474/IJAR01/3158

34. Rebuschat, P., Meurers, D., & McEnery, T. (2017). Language Learning Research at the intersection of Experimental, Computational, and Corpus-based Approach. *Language Learning*, 67(1), 6–13. https://doi.org/10.1111/lang.12243

35. Roca Varela, M. L. (2012). Corpus Linguistics and Language Teaching: Learning English Vocabulary through Corpus Work. *Corpus Linguistics and Language Teaching*, 33, 285–300.

36. Taher, M. A., Shrestha, P. N., Rahman, M. M., & Khalid, A. K. M. I. (2016). Curriculum Linked Video (CLV) as a tool for English Language Teaching (ELT) at secondary school classrooms in Bangladesh. International Journal of Humanities, Arts and Social Sciences, 2(4), 126-132. https://doi.org/10.20469/ijhss.2.20002-4

37. Usman, A. Y. (2013). Using Information and Communication Technology (ICT) to Enhance the Teaching and Learning of Arabic and Islamic Studies in Nigeria. *Journal of Teacher Education*, 2(3), 353–368.

38. Vyatkina, N., and Boulton, A. (2017). Corpora in language learning and teaching: Discussant's summary. *Language Learning and Technology*, *21*(3), 1–8.

39. Widdowson, H. G. W. (2003). *Defining Issues in English Language Teaching*. Oxford: Oxford University Press.