Ukazh Journal of Arabic Studies
E-ISSN : 2797-1910
Vol. 4, No. 2, Desember 2023, Hlm. 515-532
DOI: 10.37274/ukazh.v4i2.858
http://ejournal.arraayah.ac.id

Sketch Engine Implementation in Arabic Vocabulary Learning

Ahmad Ashfia*, Elyatul Mu'awanah , Abdullah Nur, Wati Susiawati
Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia
ahmadashfia86@gmail.com

Abstract

This study aims to understand how to analyse the language corpus and be able to apply it properly and to improve the Arabic vocabulary of new students of UIN Jakarta majoring in Arabic Language Education (PBA) through the application of sketch engine. The application of sketch engine in learning Arabic vocabulary is still very rarely applied. This sketch engine is the first time applied to new students of UIN Jakarta in the special guidance of al-'Arabiyah al-Mukasyafah. The research method used is descriptive method with qualitative approach (field research). While data collection techniques are done by observation, interviews, and documentation. Based on the results of the study, there are three stages carried out; 1) presentation of conceptual material 2) presentation of skill material 3) evaluation of activities 4) reflection of activities. From these four stages, students are able to apply their applicative knowledge in using the sketch engine. All students who were divided into 4 groups were able to follow the steps of corpus analysis very well through reports in the form of working papers. In addition, the application of this sketch engine is able to increase students' Arabic vocabulary skills through finding word collocations and finding various meanings that fit the vocabulary.

Keyword: *The Language Corpus; Sketch Engine; Vocabulary Learning*

I. INTRODUCTION

In the historical record of the development of Arabic language learning, there are four language skills that must be mastered by students. The four language skills include listening skills (*maharah al-istima'*), speaking skills (*maharah al-kalam*), reading skills (*maharah al-qira'ah),* and writing skills (*maharah al-kitabah*) (Fudhaili, 2022). Of the four skills are still divided into two parts, *first,* receptive skills (listening skills and reading skills), *second,* productive skills (speaking skills and writing skills) (Ahyar, 2018).

In addition to these four skills, there are also elements of language that must be considered in the Arabic language learning process, namely sound (*aswat*), vocabulary (*mufradat*), and grammar (*qawa'id*) (Hestiyani, 2019). These three elements of language are very important to obtain the objectives of learning Arabic, one of which can apply and develop Arabic language skills, both actively and passively. And vocabulary (*mufradat*) is

Diserahkan: 07-10-2023 **Disetujui**: 03-11-2023. **Dipublikasikan**: 28-12-2023

the second element of language that must be considered and is a very important principle in learning Arabic.

Vocabulary and language have a very strong and close relationship. A person's ability to master a language depends on his or her ability to master the vocabulary in that language (Razak & Samah, 2018). Its role as a support for its use can help students master the four language skills. Without a proportionate vocabulary, students will have difficulty speaking, reading, listening and writing in Arabic.

So far, many methods, strategies, and learning media have been carried out by an Arabic teacher to improve students' vocabulary, starting from listening, writing and then memorising it to the latest methods such as the card short method, hybrid learning and so on. However, these methods are assumed to be still very manual in this sophisticated era. So in this case, a teacher must have high creativity and innovation that is renewable according to the times.

Technology has become a major aspect of human life. It is undeniable that its existence has a huge effect on the way people live their lives, one of which is in terms of learning. One technology-based tool that is very useful in language learning is corpus linguistics. Corpus linguistics is a digital linguistic data processor that can be utilised in various fields of linguistics such as morphology, syntax, semantics, lexicology and other linguistic fields such as translation and others (Yuliawati, 2012-2013). According to some people, corpus linguistics is a new study in linguistics because of its rapid development in the 1960s along with the development of computer technology.

So far, there are several researchers who have conducted research related to corpus linguistics. Among the studies relevant to corpus-based vocabulary learning is a study conducted by Nur Hizbullah, Fazlurrahman, Fuzi Fauziah on corpus linguistics in Arabic language study and learning in Indonesia. The results of this study indicate the importance of the existence of an Arabic corpus in Indonesia. The initial step was carried out by compiling a corpus of Arabic learners in Indonesia then continued with corpus linguistic-based research gradually, continuously, and comprehensively on Arabic.

From the results of these research achievements, the researcher took the initiative to conduct research on the implementation of the *sketch engine* which is one of the leading corpus in Arabic vocabulary learning for UIN Jakarta students in the special guidance of al-'arabiyah al-Mukasyafah. The use of corpus in Arabic language learning, especially in improving Arabic vocabulary skills is still very rare. Corpus linguistics is still mostly struggling in the world of research only. This is a big homework for Arabic language teachers or lecturers to be able to utilise technological sophistication such as corpus linguistics in Arabic language learning.

Therefore, the researcher wanted to open a wider window of knowledge about the application of corpus linguistics in vocabulary learning in the classroom. In addition to improving students' vocabulary skills, this study was also conducted to provide an understanding of how to apply the sketch engine itself. So that students can also analyse language through the corpus tool.

II. METHOD

The research approach used in this research is a qualitative approach with the type of *field* study *research* (*field research*). Qualitative research is a research design that refers to the process of collecting and analysing data in the form of text, images, or various other types of data that are not in the form of numbers (Setyadi, 2022). Here the researcher chose the descriptive method in analysing the data. This descriptive method is used to describe or analyse research results but cannot be used to draw broad conclusions. This method aims to solve a problem that is solved systematically, accurately, and factually relating to some facts and characteristics of a particular population (Anggraeni et al., 2023).

This research was conducted at Syarif Hidayatullah State Islamic University Jakarta. The samples used in this study were new students majoring in Arabic Language Education (PBA) in the special guidance activities of al-'Arabiyah al-Mukasyafah level 4 with research collaborators, namely teachers or supervisors in these activities. The number of students involved was 20 people. The research instruments consisted of the researcher himself, observation sheets, interview guidelines, and pictures.

The procedure used in data collection is as follows:

1. Observation

   Through this observation technique, researchers try to understand the context of the data in the overall situation in the classroom in al-'Arabiyah al-Mukasyafah learning activities so that researchers can understand the data in the field as a whole.

2. Interview

   With this technique, researchers try to find deeper information related to the research topic with related parties, namely teachers or supervisors and also students through questions and answers related to the implications of the *sketch engine* in the learning process of Arabic vocabulary.

3. Documentation

   Documentation in this study is in the form of photographs of learning activities carried out during data collection in the classroom.

III. THEORITICAL STUDIES

Language Corpus

A. Definition of Corpus

Corpus comes from Latin meaning "body" or collection. According to a closer meaning, it means a collection of texts from a language variety that can represent as much as possible a trend or phenomenon of that variety (Wirza, 2016). Therefore, we are not talking about just a few texts, but a collection of texts that may consist of hundreds of thousands or even millions of words to be able to see a language phenomenon.

Corpus linguistics is an empirical method in linguistic analysis and description to examine the language spoken by natural speakers (Cheng, 2011). This method is used to examine various linguistic problems in a broad scope and is even able to reveal interesting, fundamental, and even surprising things from existing linguistic data. The linguistic data that can be analysed linguistically with this method is as explained by Nasselhauf that in general the linguistic data is divided into two large groups, namely those obtained from intuition, both the intuition of the researcher and the intuition of the informant, and language data used daily naturally, can be from light conversation and can also be a collection of texts that are collected and managed systematically. Nasselhauf also explained that the corpus is the text of practical and natural use of spoken and written language and is collected systematically. The meaning of systematic is the preparation of the structure and content of the text based on principles outside the language, such as the context of time, place, and others, and allows for further research on the aspects that are in it.

The data is organised systematically and easily accessible electronically via computer. This method is used in descriptive linguistics as well as applied linguistics, such as dictionary construction to ensure that the data used is trully derived from a wide range of usage and avoids subjective usage. In addition, a corpus can help provide, collect and organise linguistic data for other research and academic purposes, including pedagogy (Nasselhauf, 2011).

B. Corpus Approach

The corpus approach consists of four main characteristics (Bennett, 2010):

1. Empirical, i.e. analysis based on actual patterns of use. The corpora consist of all acts of communication, such as collections of texts of books, fiction, nonfiction, magazines, papers, literary works, newspapers, telephone conversations, mobile phone conversations, meetings, lectures, radio broadcasts, and television shows. In other words, any situation in a linguistic context can be a corpus.

2. Using a large, structured collection of texts as the basis for analysis; this characteristic refers to the corpus itself which can take the form of a written corpus, spoken corpus, etc.
3. Leveraging the use of technology in analysis; The use of technology in this context is not only to store corpora, but also to help analyse the language in a corpus. The corpus is accessed and analysed by a computerised codebook program.
4. Using quantitative and qualitative analysis techniques; the quantitative results generated from the corpus will later be analysed qualitatively to find research conclusions.

In terms of language corpus construction, there are now many language corpus models available. More specifically, in terms of types, there are several types of corpus that can be used depending on the type of analysis. Here are the types (Hizbullah et al., 2016):

1. General/reference corpus, eg: British National Corpus (BNC)- vs. a specialised corpus, created to represent a language in a general and comprehensive way, consisting of written and spoken language data that is transcribed and sourced from a variety of text types.
2. Historical corpus, e.g. Helsinki Corpus-vs modern language corpus, which is organised to represent specific periods of a language's development.
3. Regional corpora, such as the Wellington Corpus of Written New Zealand English (WCNZE), are compiled to represent the language variety of a region.
4. Learner corpus, such as the International Corpus of Learner English (ICLE)-vs-native speaker corpus, which was created to represent the language produced by non-native learners of the language.
5. A multilingual corpus vs. an ecolingual corpus, which represents at least one phenomenon from two languages sourced from the same variety of text, is usually used for the purpose of contrastive analysis.
6. Spoken corpus vs written corpus vs spoken-written mix, to represent the spoken language of a language-speaking community.

C. The Arabic Corpus and its Studies

There are now several Arabic corpuses created with all their specificities, advantages, and not to forget their respective disadvantages. In the preparation of an Arabic corpus, through an application on the Sketch Engine website, for example, there is a corpus containing approximately 5.8 million words (2012 version) and updated to 8.3 million words (2015 version) taken from a number of sources (sketchengine.eu).

In addition, al-Sulaiti inventoried a more comprehensive list of about 18 corpus types of corpus that come from various sources and are used for various spesific fields of study in the University of Leeds Sublaman, including the Corpus of Arabic by the

Univercity of Alexandria, Egypt (eps.leeds.ac.uk). The many corpus are divided into two major groups, which can be downloaded for a fee and not for a fee.

Besides the Arabic corpus, there is also a specialised corpus on the Qur'an. So far, according to our observations, the most representative one seems to be the University of Leeds' The Quranic Arabic Corpus (https://corpus.quran.com/).Not only the Qur'ānic corpus, the website contains models of Qur'ānic linguistic analysis, such as word-by-word translations in English, Qur'ānic vocabulary dictionaries, syntax trees of concepts and word structures, and Qur'ānic grammars.

The website also provides digital files of the Qur'ān in txt format that can be downloaded for analysis. However, the file only contains the text of the Qur'an in Latin transliteration, not Arabic, although it is equipped with the grammatical status of each word. Studies on Arabic corpus linguistics, and specifically the Quran, have been conducted by Sharaf (2009). This expert introduces a series of researches along with the results of conceptual mapping of the contents of the Qur'an with certain applications. However, on his website, there is no digital material that can be downloaded and utilised for further research (http://textminingthequran.com/).

One of the steps in compiling a generalised corpus of Arabic that includes the Qur'an as one of its sources was undertaken by Al-Rabia (et.al., 2014). The corpus is called the King Saud University Corpus of Classical Arabic (KSUCCA). As the name implies, they only focus on taking corpus from classical period Arabic and the Qur'an is assumed to belong to that period.

They argued that classical Arabic patterns are the basis of Arabic linguistic theory and should be viewed as a scientific-academic resource. The corpus contains 50 million words from the classical Arabic literature and is used, among other things, for the study of the lexical semantic distribution of the Qur'an and classical Arabic using a computational linguistics approach.

But besides describing the structure of the corpus, their research also describes some symptoms of language change from classical to modern times. There are other works that are more specific, for example, Sharaf and Atwell (2012). They created Qur'an, an annotated corpus of Qur'an specifically on the case of pronominal anaphora. The research was directed towards analysing pronominal anaphora in the Qur'an and the patterns of occurrence of anaphora associated with pronominal anaphora and their antecedent links in Qur'anic verses.

The analytical model of this study is useful to apply to the analysis of pronomina, anaphora, and their antecedents in modern Arabic discourse and the comparison of the two periods. One other example of more specific Qur'ānic corpus research is the work of Imad and Abdelhak.They conducted research to compile a Qur'ānic corpus specifically for

research in Arabic morphology. They used the text of the Qur'an and also utilised Al-Khalil's morphological tables. They used data from the University of Leeds' Quranic Arabic Corpus and the University of Haifa's Quranic Corpus of Haifa as guidelines (Imad & Abdelhak, 2016).

In the next step, they analyse and critique the concept of word roots, and end with a comparison between the two and an explanation of the researcher's own formulation. In conclusion to the previous research, it can be said that the use of corpus linguistics approach among Arabic language researchers has been so widespread. Experts and enthusiasts of Arabic studies in Indonesia should enter this arena and enliven corpus-based Arabic studies.

Sketch Engine

Sketch engine is one of the leading corpus tools produced by Lexical Computing. Lexical Computing is a research company founded by lexicographer and linguist Adam Kilgarriff in 2003 (sketchengine.eu). The company works at the intersection of corpus and computational linguistics and is committed to an empirical approach to the study of language, with Adam Kilgarriff being the only central Sketch Engine figure in the company before he was diagnosed with cancer and died in November 2014.

The sketch engine was first used in the creation of the Macmillan English Dictionary and presented at Euralex 2002. After the presentation, questions were raised about whether the sketch engine could be used in their language. Following this, the company Lexical Computing developed the sketch engine as a corpus tool that accepts as input any language corpus with appropriate linguistic markings and then in between generates word sketches for those words (Kilgarriff et al., 2004).

The main goal of the programme is to create a tool that learns the lexicographic properties of words (Gulshat, 2019). In addition, the sketch engine also aims to contain coverage of various languages with a large number of speakers in the world requested by specific users. Therefore, this Sketch Engine offers more than 200 packages in 82 bilingual languages (Kilgarriff, 2014). While the data provided reaches more than 700 text corpora with a size of up to 60 billion words and covers more than 100 languages. One of the features in Sketch Engine is the variety of languages, including Arabic, Chinese, Czech, English (there are two versions of English: American and British), Estonian, Finnish, Japanese and others.

The functions of this system consist of the following tools:

1. Word Sketch, produces a summary of the grammatical and collocational behaviour of a word using "sketch grammar".
2. Word Sketch Differences, offers a comparison of two words based on collocation.

3. Dictionary, creating a distribution thesaurus based on common collocations. The resulting word list includes items in various semantic relationships.
4. Concordance, searching the corpus for word forms, lemmas, phrases, part tags, and others. The system converts all queries into Corpus Query Language (CQL) that can be used directly.
5. Word List, generates a frequency list of words, lemmas, n-grams or keywords.
6. Keywords and terms enable the extraction of core lexis in the corpus using a "keyness score".
7. Collocation, counting words that are statistically associated with query terms. This system uses a "sketch grammar".
8. Treen, helps conduct diachronic analyses of word usage.
9. WebBootCaT, is a set of programmes to complicate the user's web corpus (Thomas, 2016).

Vacabulary Learning (*Mufradat*)

A. Definition of Learning

Learning is inseparable from two learning events and teaching events, both of which have a close relationship and even interactions that influence and support each other. Learning is the main determinant of educational success. Learning is a two-way communication process. Teaching is carried out by the teacher as an educator, while learning is carried out by students or students. In the Law of the Republic of Indonesia. No. 20 of 2003 concerning the National Education System Article 1 paragraph 20 "Learning is the process of interacting students with educators and learning resources in a learning environment" (Hasnah, 2015).

B. Definition of Vocabulary

In the Big Indonesian Dictionary, the word vocabulary is given the meaning of vocabulary (http://kbbi.web.id). However, apart from the information in the Big Indonesian Dictionary, there are several definitions of vocabulary (Chaer, 2007) namely:

1. All the words in a language.
2. Words mastered by a person or group of people from the same neighbourhood.
3. Words or terms used in a field of activity or science.
4. A number of words of a language organised alphabetically with some explanation of their meaning, like a dictionary.
5. All the morphemes that exist in a language.
6. Based on the concepts of vocabulary above, it appears that the language units included in vocabulary are very broad and diverse (Ardiyanti et al., 2018).

C. Vocabulary learning stages

According to Sarinah Hardjono, of all the aspects of foreign languages that learners must master in the process of teaching and learning foreign languages, the vocabulary aspect is the most important thing. With the mastery of language that has a function to communicate well, a language learner must master vocabulary, because vocabulary will

help students in learning foreign languages, especially in mastering the four language skills, namely, listening (istima), speaking (muhadatsah), reading (qirā'ah), and writing (kitābah) (Hasnah, 2015). There are several stages according to Ahmad Fuad Effendy in learning vocabulary or students' experience in recognising and acquiring the meaning of words, namely:

1. The first stage is listening to the word. The teacher says a word and then the students are given the opportunity to listen. The spoken word is free, whether it stands alone or is in a sentence. If the sound they receive is clear, then students will be able to listen properly and correctly.
2. The second stage is saying the word. After the students hear the word that has been pronounced by the teacher, then the students are given the opportunity to pronounce the word that has been heard before. Saying the word that has been heard makes the student able to remember the word for a long time.
3. The third stage is getting the meaning of the word. At this stage the teacher is strongly advised not to use translation in giving meaning to students. Because, if this is done, students will forget the meaning of the word more quickly and there will be no direct communication in the language being learnt. One way to avoid this is by using pictures, providing sentence context, synonyms, antonyms, simple definitions, gesture demonstrations, and so on. And the last alternative if students really don't know it then the teacher provides a translation.
4. The fourth stage is reading the word. After doing the three stages above, students are given the opportunity to read the words that have been heard, pronounced, and understood the meaning using a loud voice.
5. The fifth stage is word writing. The sentence that has just been spoken by the student is certainly still fresh in his memory. The next step is for students to be directed to write the words they have learnt. The final stage of a series of vocabulary learning stages is making sentences.

The purpose of mufradat learning is none other than to introduce new vocabulary to students which can be done by using several media or materials such as: reading books, listening to Arabic stories, and so on. In addition to this, learning mufradat can also train students in understanding the meaning of vocabulary, both in the form of connotative and denotative meanings. And the most important thing is that students can express and function mufradat properly in a sentence (Mufidah & Rohima, 2020).

D. Vocabulary Function

In terms of function, vocabulary (mufradât) can be divided into two, among others:

1. Al-Mufradât al-Mu'jamiyah (المفردات المعجمية), i.e. vocabulary that has a meaning in the dictionary such as the word قلم، قمر ، بيت.
2. Al-Mufradât al-Wadzîfiyah (المفردات الوظيفية), namely vocabulary that carries a certain function, for example hurûf al-jar, asmâ al-Ishârah, asmâ al-Maushûl, dlamâir, and others like it.

Of the two kinds of vocabulary, it should be noted that among Al-Mufradât al-Mu'jamiyah there are several things to note, as follows:

1. There are several words that have similar meanings, such as the words look, see, pay attention and witness in Arabic in the form of (شاهد, لاحظ, نظر, رأى).
2. There are some words that have the same denotative meaning but have different connotative meanings or differ in the context in which they are used, such as the words مات and توفـي which can be interpreted in Indonesian as "die, died, killed, passed away or mampus".
3. A word that has several different meanings, such as the word فصل which can mean "class" , "season" or "chapter" and "chapter".

The description of things related to vocabulary needs to be considered and known by people who work as language teachers, especially Arabic. In Arabic language learning, there are some problems in vocabulary learning called vocabulary problematics (مشكلات صرفية). It happens because vocabulary learning includes complex themes such as derivation changes, inflection changes, verbs, mufrad, tatsniyah, jama', ta'nîts, tadzkîr and lexical and functional meanings.

E. Types of Vocabulary

Rusydy Ahmad Thu'aimah provides a classification of vocabulary (mufradât) into 4 (four), each of which is further divided according to its duties and functions, as follows:

1. Vocabulary division in the context of Linguistic Proficiency

a). Vocabulary for understanding both spoken language and text.

b). Speaking vocabulary. In conversation, it is necessary to use the right vocabulary, both informal and formal conversation.

c). Writing vocabulary. Writing also requires the selection of good and appropriate vocabulary so as not to be misinterpreted by the reader. This includes informal writing such as diaries, agendas, etc. as well as formal writing such as books, magazines, newspapers, etc.

d). Potential vocabulary. This type of vocabulary consists of context vocabulary which can be interpreted according to the context of the discussion, and analysis vocabulary which can be analysed based on the characteristics of the derivation of the word to further narrow or expand its meaning.

2. Division of vocabulary according to meaning

a). Core words (content vocabulary). This is the basic vocabulary that makes a piece of writing valid, such as nouns, verbs, etc.

b). Function words. These words that bind and unite vocabulary and sentences so as to form a good exposure in a piece of writing. Examples hurûf jâr, adawât al-istifhâm, and so on.

c). Cluster words. These are words that cannot stand alone, but are always combined with other words to form different meanings. For example, the word رغب

can mean liking when it is combined with في to become رغب فى . When followed by the word عن, it becomes رغب عن , and the meaning changes to hate or dislike.

3. Division of vocabulary according to the characteristics of the word (takhassus)

a). Task words (service words) are words used to indicate tasks, both in the field of life informally and formally and are official in nature.

b). Special content words. This vocabulary is a collection of words that can transfer meaning to specific ones and are used in various specific areas of review, which are also called local words or utility words.

4. Division of vocabulary according to its use

a) Active words, i.e. vocabulary that is generally widely used in various discourses, whether speaking, writing or even widely heard and known through various readings.

b) Passive words, which are vocabulary that is only in a person's vocabulary but rarely used. This vocabulary is known through printed books that are usually used as references in writing papers or scientific works (Hijriyah, 2018).

IV. RESULTS

The activity of implementing the sketch engine in learning vocabulary (*mufrodat*) began on 20 October 2023 to 10 November 2023 for 4 meetings with a face-to-face learning format. The object of this research is UIN Jakarta students from the level 4 Arabic Language Education Study Programme in al-Arabiyah al-Mukasyafah learning activities. There are 20 students who actively participate in this activity. Based on the results of observations and interviews conducted by researchers with teachers or supervisors and students, it is conveyed that the process and steps taken in implementing the sketch engine in learning Arabic vocabulary consist of:

1. Presentation of Conceptual Material

Learning activities at this stage are in the form of activities to absorb conceptual material about the nature and types of data corpus and the benefits of understanding language data corpus. This activity received a positive response from students identified through interviews to students through a draft interview shared to the WhatsApp group. In general, all students in this activity did not know what a language corpus and sketch engine were. In this activity, most students stated that they only understood that the corpus is language data that can be accessed through digitised means.

Gambar 1: Presentattion of Comceptual Material

Al and can be accessed through existing web facilities and applications. In general, students stated that they gained new insights into the ins and outs of language corpus data that they had not known about.

2. Presentation of Skills Material

In the second stage of this activity, the teacher or supervisor divides 4 groups of 20 students. Each group is required to bring a laptop and create an account on the *sketch engine* website. Then the teacher or mentor gives a predetermined *mufrodat* taken from one of the texts in the book Arabiyah Baina Yadaik volume 2B by Abdurrahman page 355 about *al-harb wa al-salaam*. Each group is asked to find the collocation and meaning of the *mufrodat* that has been determined.

Before they start the practice of applying the sketch engine, the teacher invites all students to read and translate the text about *al-harb wa al-salaam* first in turn. Then all students were told to join their respective groups to find the *mufrodat* collocation that had been determined on the *sketch engine* website. Then they discussed the suitable meaning for each word collocation they found. This activity also directs students to understand how to analyse the language corpus so that they know how extensive the

Arabic language is and also to realise that in one vocabulary it does not only mean one meaning.

The application of this *sketch engine is* not only to find the collocation of a word, students are also assisted by various digital dictionaries such as al-ma'ani dictionary, Arabic-Indonesian dictionary, and the like to find out what the meaning of the vocabulary is when collocated with different words. The following are the discussion results from each group of students:

Word Collocation Search Result in Word Sketch Feature in Sketch Engine

Group Name	Vocabulary	Kosakata					
		as a Verb	as a Noun	as Subject	as Object	as Modifier	as Or/Nor
Group 1 1. Shafa 2. Habibi 3. Ayu 4. Reyhan	حدث-يحدث	1.968.288x	1.886.499x	يحدث مرض (affected by disease) (378x)	أحدثك بحديث (I am discussing a discussion) (74x)	-	-
	إنسان	-	3.270.047x	يحتاج الإنسان (humans are creatures that need something) (564x)	اناخلقنا الإنسان من نطفة (Indeed we created man from a clot of blood) (78x)	الإنسان العاقل (Man is intelligent) (6.606x)	الإنسان أو الحيوان (human or animal) (3.514x)
	بحث-يبحث	1.197.770x	2.585.559x	وتبين البحوث (Doing search) (272x)	يقوم بالبحث (Explaining the explanation)	البحوث الزارعية (Agricultural research) (8.498x)	البحث أو المقالة (Article of paper)
	عاش-يعيش	1.675.300x	-	يعيش معظم (Organised life)	عشت عمري (Spend life)		
Group 2 1. Lulu Favira 2. Hanifa 3. Fadil 4. Sadam	وقع-يقع	2.210.019x	9.134x	تقع مدينة (Located in the city)	وهي تقع بالقرب من (And it is located close to) (374x)		
	عظم-يعظم	92.790x	298.625x	علم عظم (22x)	يعترف بعظمة لسانه	التهاب العظم والنقي	العظام والإنسان
	نقل-ينقل	1.148.502x	2.108.127x	انتهى النقل	نكتفي بنقل	النقل الجوي (Air transport)	العقل والنقل Aqli and naqli evidence)
	ذاق-يذوق	77.014x		ذقت العذاب (Feeling the torment) (12x)	تذوق الجمال (Enjoy the beauty)		
Group 3 1. Dafa	حفظ-يحفظ	358.976x	875.026x	يجب حفظ	فهو يقووم بحفظ المال	حفظا متقنا	الحفظ والفهم

2. Sahla 3. Nuriyah 4. Faizi 5. Rois				(Must be saved) (371x)	(He tried to save his wealth) (83x)	(Perfectly maintained) (324x)	(Memorisation and understanding) (193x)
	تمتع-يتمتع	514.194x	81.532x	تتمتع مدينة (Beautiful city) (411x)	وهويتمتع بصحة (He enjoyed his life) (319x)		
	حل-يحل	2.532.501x	372.004x	أريد حل (The solution in question) (866x)	وجت حلا (I found the solution) (96x)	حلّ سلمي (Peaceful solution) (21.248x)	حل أو تسوية (Solution or Settlement) (432x)
	أنشأ-ينشئ		218.675x	أنشئت كلية (Universities are established)	أنشأجمعية (Form an association) (10x)		
	فرض-يفرض	661.867x	564.668x	يفوض القانون (Law enforcement) (577x)	تفرض عقوبات (Determining the punishment) (32x)		
Group 4 1. Uswah 2. Nadiya 3. Tsurayya 4. Vhebe 5. Tengku	حقق-يحقق	1.289.694x		حققت الشركة (investigate) (4.309x)	يحقق الإنتصارات (Achieving victory) (142x)		
	نفذ-ينفذ	655.734x	763x	نفذت وحدات (Unit Implemented) (918x)	ينفذ تعليمات (Storing intructions) (94x)		
	وهم-يهم	36.469x	122.926	سقط أوهام (Loss of common sense) (13x)	تبيع الوهم (Selling the illusion) (33x)	الأوهام والظنون (Illusion and suspicion) (234x)	الخرفات والأوهام (Superstition and illusion) (167x)
	وقف-يقف	966.629x	1.187.334x	يجوز وقف	نحن وجب وقف (Stop) (17x)	الوقف الفوري (Immediate suspension) (8.721x)	الأوقاف والشؤون الإسلامية (1.300x)

Based on the group work results above, all groups were successful and able to apply their applicative knowledge in using the sketch engine. All groups were able to follow the steps of the Language corpus analysis very well and report it in the form of a working paper. Thus, it can be concluded that the implementation of sketch engine as a language corpus is suitable for Arabic vocabulary learning.

3. Evaluation of Activity

In this learning activity, evaluation is conducted in the following ways.

1. The evaluation was conducted through several questions asked by the researcher to the students. The draft questions were done through a Google form link and shared with participants through a WhatsApp group. Through the draft questions, students' understanding and interest in the *sketch engine* can be known. The results of the

comprehension evaluation showed very good success. Students were able to fulfil competencies in knowledge and proved to be able to apply knowledge about language corpus data. The results of the evaluation of the theoretical and applicative knowledge gained by students about the language data corpus have been summarised and can be observed in the description of the results and discussion.

2. Evaluation is conducted through practical performance results. The evaluation of performance in this learning activity involves assessing the applicative skills of participants or students, including (1) their ability to apply the sketch engine to find word collocation and (2) their ability to interpret vocabulary and discuss it with their group. At this stage, students are able to report their work well and are able to respond to various problems from other group participants.

4. Reflection of Activity

After the implementation of the sketch engine in Arabic vocabulary learning was completed, a reflection was conducted as the final part of this lesson. This reflection involved a question and answer session with the participants using an interview draft as a means to obtain feedback during the training. The reflection includes the responses and benefits felt by the participants on the knowledge material about language data corpus, language data corpus analysis exercises, as well as the follow-up of the sketch engine implementation. Based on the results of the reflection through the questionnaire, the responses from the participants were as follows.

1. Reflection on the conceptual knowledge material of the language corpus on the sketch engine website:
   students' responses after attending the training:

   a. better understand the language corpus; understanding the language corpus can encourage self-learning because it can be accessed anywhere and anytime. Understanding the language corpus can inspire or get ideas and discover something new;
   b. knowledge and understanding of the language corpus can recognise how to process the data;
   c. can clearly know what is meant by a language corpus, can distinguish between a raw and processed language corpus;
   d. gained explanatory knowledge about what a corpus is, the types of corpus, and the benefits of a corpus;
   e. The benefit of studying a language corpus is that it can be a source of descriptive data that provides information about how the language is used. This information can be utilised by teachers or learners as a tool to seek evidence or confirm their intuitions about the knowledge of the language.
   f. know how to create a corpus and distinguish between corpus types;
   g. provide knowledge related to how to manage data, and can find out the description of the data to be studied.

2. Reflection on the practice/application of language corpus analysis exercise using *sketch engine* in *mufrodat* learning. After carrying out the practice/application of language corpus analysis using sketch *engine*, students' responses are:
   a. can find out the collocation of *mufrodat* easily.
   b. The practice of this activity is very useful, from one vocabulary presented, it can bring up a variety of data, and from these data a conclusion can be drawn that one vocabulary can have many meanings when collocated with different words and in different discussions;
   c. The practice process is new knowledge for students in analysing *mufrodat*.
   d. from the corpus training can know the importance of learning the language corpus to know the collocation of a word and even there are synonyms and antonyms of a word;
   e. There is a need for mentoring in the use of the sketch engine for beginners, so that it is easier to understand the use of the language corpus website.

Reflection on feedback as a follow-up to the use of *sketch engine* in *mufrodat* learning. The students' suggestions in general stated that there should be continuous training in the use of sketch *engine* as the utilisation of digital language corpus. Students also suggested that the learning of this language corpus can be felt by all students of the Arabic Language Education study programme at UIN Jakarta.

V. CONCLUTIONS

Based on the results of the study, it shows that there are three stages carried out; 1) presentation of conceptual material 2) presentation of skill material 3) evaluation of activities 4) reflection of activity. From these four stages, students were able to apply their applicative knowledge in using the sketch engine. All students who were divided into 4 groups were able to follow the steps of corpus analysis very well through reports in the form of working papers. In addition, the application of the sketch engine is able to increase students' Arabic vocabulary skills through finding word collocations and finding various meanings that fit the vocabulary. However, in this sufficient research, there is still a need for continuous training in the use of the sketch *engine* as a digital language corpus utilisation. Students also propose that this language corpus learning can be felt by all students of the Arabic Language Education study programme at UIN Jakarta.

VI. REFERENCES

Abdul Razak,. Zainur Rijal,. & Samah Rosni. (2018). Arabic Vocabulary Saiz among Students in Higher Education: Issues and Teaching Strategies. *International Journal of Language Education and Applied Linguistics (IJLEAL)*, 08 (2), 61-70. https://doi.org/10.15282/ijleal.v8.394

Ahyar, Haerul (2018). Mastery of Mufrādat and Qawā'id as an Effort to Improve Arabic Writing Skills. *al Mahāra Journal of Arabic Language Education,* 4 (2),259-274. https://doi.org/10.14421/almahara.2018.042-06

Anggraeni, L,. Martin, A,. Puspita, D,. Kristiana, D, Novi, A,. Kristina, M,. Nagara, ES,. Utami, BHS,. Sari, NY,. Pratomo, PA,. Andewi, W,. (2023). *Research Methodology,* West Java: CV Adanu Abitama.

Ardiyanti, A,. Bandu, I,. & Usman, M,. (2018). Learning French Vocabulary with Flashcard Media (Case Study on French Literature Students). *JOURNAL OF CULTURAL SCIENCES,* 6 (1), 176-186. http://dx.doi.org/10.34050/jib.v6i1.4327

Bennett, Gena R (2010). *Using Corpora in the Language Learning Classroom: Corpus Linguistics For Teachers*. University of Michigan ELT Press. https://doi.org/10.3998/mpub.371534

Cheng, Winnie (2011). *Exploring Corpus Linguistics*. London: Routledge. https://doi.org/10.4324/9780203802632

Comp.leeds.ac.uk. accessed on 11 November 2023 at 11:06 pm. http://www.comp.leeds.ac.uk/eric/latifa/arabic_corpora.htm

Corpus.qura.com. accessed on 11 November 2023 at 11:09 pm http://corpus.quran.com/

Fudhaili, Achmad (2022). Kumon Model as a Learning Sequence: Improving Students' Imla' Writing Skills. *An Nabighoh,* 24 (2), 155-170. https://doi.org/10.32332/an-nabighoh.v24i2.5138

Hasnah, Sitti (2015). Learning Arabic Vocabulary (Mufradät) Through Picture Media to Increase Student Learning Interest in the Department of Islamic Education, Faculty of Tarbiyah IAIN Palu. *ISTIQRA, Journal of Scientific Research*, 3 (1), 197-225.

Hestiyani, Yosha (2019). Card Sort Method to Improve Vocabulary Mastery in Arabic Language Learning at School. *ALSUNIYAT: Journal of Arabic Language, Literature and Culture Research*, 2 (2), 149-161. https://doi.org/10.17509/alsuniyat.v2i2.23574

Hizbullah Nur, Fazlurrahman, & Fuzi Fauziah (2016). Corpus Linguistics in Arabic Language Study and Learning in Indonesia. *Proceedings of the National Conference on Arabic Language II* Malang, 385-393.

Imad, Z & Abdelhak L (2016), New Quranic Corpus rich in Morphological Information, *International Journal of Speech Technology*, 19(2), 339-346. https://doi.org/10.1007/s10772-016-9335-7

Kbbi.web.id. accessed 11 November 2023 at 11:03 WIB. http://kbbi.web.id

Kilgarriff, A,. Rychly, P,. Smrz, P,. & Tugwell, D,. (2004). *Sketch Engine.*

Kilgarriff, Adam,. Baisa, V,. Busta, J,. Jakubicek, M,. Kovar, V,. Michelfeit, Rychly, P,. & Suchomel, V,. (2014). The Sketch Engine: Ten Years On. *Lexical Computing Ltd, Brighton, UK & Masaryk University, Brno, Czech Republic*.

Mufidah, Nuril & Rohima, Intan Izha,. (2020) Teaching Vocabulary for Arabic Intensive Class Students. *Uniqbu Journal Of Social Sciences (UJSS)*, 1 (1), 13-24.

Nadja, Nesselhauf (2011), *Corpus Linguistics: A Practical Introduction*. Anglistisches Seminar

Safiullah, G,. Davitova, R,. & Lelakova, E,. (2019). Sketch Engine in Building a Lexical Minimum for Child. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 1(9), 5105-5108. https://dx.doi.org/10.35940/ijitee.A9206.119119

Setyadi, Bakti (2022). *Research Methods,* Yogyakarta: Jejak Pustaka.

Sketchegine.eu. accessed 11 November 2023 at 11:03 am. https://www.sketchengine.co.uk/artenten-corpus/

Sketchengine.eu. accessed 11 November 2023 at 11:03 pm. https://www.sketchengine.eu/#blue

Textminingthequran.com accessed on 11 November 2023 at 11:15 WIB http://www.textminingthequran.com/wiki/QurConcord

Thomas, James (2016). *Sketch Engine: a Toolbox for Linguistic Discovery*.

Uswar, Yenita, et al. (2023), *Improving Corpus-Based English Vocabulary Mastery Through Sketch Engine Media*, Abdimas Mandiri, Journal of Community Service, Vol. 3 No. 1 May 2023

Wirza, Yanti. Application of Concordance Program Software in Language Teaching and Research (Case Study on 6th Semester Students of Language and Literature Department, Universitas Pendidikan Indonesia). *UNIKOM Scientific Magazine,* 7(2), 217-222.

Yuliawati, Susi (2014). Corpus-Based Analysis; Collocation of Words Meaning 'Woman' in Sundanese Media (Mangle Magazine, 2012-2013). *Domain,* 03 (1), 107-123. https://doi.org/10.26499/rnh.v3i2.42