Home   About   Login   Register   Categories   Search   Current   Archives   Announcements

Home > Arabiyat : Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban | Vol. 8 No. 1 June 2021 > **Hendra**

COMPILING VOCABULARY LISTS FOR CORPUS-BASED ARABIC FOR TOURISM TEACHING

Faisal Hendra, Mujahidah Fharieza Rufaidah

ABSTRACT

Arabic for Tourism is a course that can be facilitated with innovative teaching materials along with the availability of learning resources on the internet rich in vocabulary and terms typical to the field of Arabic tourism. Therefore, the use of Arabic tourism websites
is an effective way to compile vocabulary lists. The articles published in this kind of websites can be collected to be used as a corpus, then processed using a corpus processing software, AntConc. This study used a combination of qualitative and quantitative approaches with descriptive and comparative methods using a corpus linguistic approach. Teachers can take advantage of the features in this application to identify vocabularies commonly used in the world of tourism. These features are word frequency, concordance, collocation, and N-Gram. The results can be used as a reference in compiling an Arabic for Tourism vocabulary list.

KEYWORDS

Arabic for tourism; vocabulary list; website; corpus processing software; AntConc

FULL TEXT:

PDF

DOI: https://doi.org/10.15408/a.v8i1.20818 Abstract - 0 PDF - 0

Refbacks

- There are currently no refbacks.

Indexed by:

About The Authors

Faisal Hendra
Universitas Al Azhar Indonesia
Indonesia

Mujahidah Fharieza Rufaidah
Universitas Al Azhar Indonesia
Indonesia

Article Tools

Print this article

Indexing metadata

How to cite item

Home   About   Login   Register   Categories   Search   Current   Archives   Announcements

Home > Archives > **Arabiyat : Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban | Vol. 8 No. 1 June 2021**

Table of Contents

ARTICLES

Editorial Team

Focus and Scope

Publication Ethics

Author Guidelines

Online Submissions

Open Access Policy

Indexing and Abstracting

Acknowledgments

Publication Fee

Journal Partnership

Contact

Citation Analysis

Scopus®

Journal Content

Search
[   ]

Search Scope
[All   ▾]

Search

Browse
- By Issue
- By Author
- By Title
- Other Journals
- Categories

Information

- For Readers
- For Authors

- For Librarians

Visitors

Visitors

ID	86,082		DE	41
US	15,542		SA	33
MY	5,042		BR	32
FR	798		ZA	32
IN	734		IE	29
CA	676		GB	27
SG	447		RU	27
VN	429			

Pageviews: 345,402

FLAG

Indexed by:

Arabiyat, p-ISSN: 2356-153X, e-ISSN: 2442-9473

COMPILING VOCABULARY LISTS FOR CORPUS-BASED ARABIC FOR TOURISM TEACHING

Faisal Hendra, Mujahidah Fharieza Rufaidah
Universitas Al Azhar Indonesia, Jakarta, Indonesia
Jl. Sisingamangaraja No. 2, Kebayoran Baru, Jakarta, 12110, Indonesia
Corresponding E-mail: faisalhendra2104@gmail.com

Abstract

Arabic for Tourism is a course that can be facilitated with innovative teaching materials along with the availability of learning resources on the internet rich in vocabulary and terms typical to the field of Arabic tourism. Therefore, the use of Arabic tourism websites is an effective way to compile vocabulary lists. The articles published in this kind of websites can be collected to be used as a corpus, then processed using a corpus processing software, AntConc. This study used a combination of qualitative and quantitative approaches with descriptive and comparative methods using a corpus linguistic approach. Teachers can take advantage of the features in this application to identify vocabularies commonly used in the world of tourism. These features are word frequency, concordance, collocation, and N-Gram. The results can be used as a reference in compiling an Arabic for Tourism vocabulary list.

Keywords: *Arabic for tourism, vocabulary list, website, corpus processing software, AntConc*

Introduction

The arrival of Arab tourists to several Muslim countries has increased since 9/11 incident on September 11, 2001, known as the terrorist attacks in New York City and Washington DC.[1] The Arab tourists and Muslim travelers change their travel destination from Europe and North America to other parts of the world, especially Muslim countries.[2]

Indonesia as a country with the largest Muslim population is one of the targets of foreign tourists. Since five years after 9/11 in 2006, Indonesia witnessed an increase in tourist arrivals from the Middle East as much as 51.479, particularly from Saudi Arabia, Yemen, and Egypt. Since then, the number of tourists has

---

[1] Ibrahim et al., "Travelling Pattern & Preferences of the Arab Tourists in Malaysian Hotels", *International Journal of Business and Management*, Vol. 4, No. 7, 2009.

[2] Hashim Bin Mat Zin & Tengku Ghani Tengku Jusoh, "The Potential of Arabic as a Tourism Language in Malaysia", *Journal of Educational and Social Research,* Vol. 3, No. 7, 2013.

continued to rise, followed by tourists from other Middle Eastern countries such as Qatar and the United Arab Emirates.[3]

The increase in the number of Middle Eastern tourists should be accompanied by an increase in human resources in the tourism sector, especially those who have the ability to speak Arabic. Here lies the significance of Arabic for special purposes or professions in supporting language skills, especially in enriching a certain work-related vocabularies which has its own uniqueness.

The specific objectives of learning Arabic are divided into two parts, namely: (1) scientific objectives, as a support for obtaining limited proficiency in a particular field of study or profession; (2) the objective of practical use, as a means to acquire Arabic communication skills both in spoken and written way, as well as receptive and productive.[4] Therefore, Arabic for tourism, as a branch of Arabic for specific purpose, is taught in educational institutions, including universities, as part of the core curriculum. This is useful to prepare students to face the strong emergence of globalization that requires foreign languages, one of which is Arabic as an introduction to various activities, including the field of tourism.

With regard to teaching Arabic for specific purpose, in this case Arabic for tourism, several important instruments such as curriculum, lesson plans, preparation of teaching materials, methods, teaching approaches, and so on should be taken into account.[5] This article focuses on discussing the preparation of teaching materials, especially the components of compiling a vocabulary list, considering that each special field has its own unique terms or vocabulary.

As the rapid development of cutting-edge technology emerges, there is an increasing trend in language teaching material development, especially with regard to the process of compiling Arabic for specific purposes vocabulary lists. The internet as one of the largest sources of foreign language vocabulary can make it easier to compose this vocabulary list, by utilizing sites that always provide the latest information for their readers. This site can be processed to be more specific by issuing vocabulary generally used in it and the context in which the vocabulary is used.

In creating a vocabulary list, the corpus linguistic approach can facilitate data processing. Corpus linguistics is the language database statistically taken for the purpose of investigation, description, application, and analysis related with the branches of linguistics.[6] The corpus comprised of a collection of data, both ordinary and digital data, in written form containing linguistic information like word level,

---

[3] Misran, "Dialek 'Ammiyyah dalam Pengajaran Bahasa Arab untuk Pariwisata di Indonesia", *Adabiyyât: Jurnal Bahasa dan Sastra,* Vol. 12, No. 2, 2013, 398–423

[4] Dahlan, *Metode Belajar Mengajar Bahasa Arab*, (Surabaya: Usaha Nasional, 1992).

[5] Halim, "Bahasa Arab Dengan Tujuan Khusus Berbasis Komunikatif Wisata Travelling." *Bintang,* Vol. 2, No. 3, 2020, 230–241.

[6] Suryadarma & Fakhiroh, "Optimalisasi Penggunaan Corpus Linguistics Dalam Penyusunan Kamus Az-Ziro'ah Sebagai Media Pembelajaran Bahasa Arab", *International Seminar on Language, Education, and Culture (ISOLEC)*, 2020.

structure, meaning, and discourse that can be used for research.[7] The corpus is created from not only hard file sources such as a collection of articles, textbooks, literary works, and newspapers, but also from the internet particularly websites, online news, and social media conversations.[8]

The compilation of Arabic vocabulary lists can be taken from a large body of online sources which can be then processed using corpus processing software one of which is called AntConc. This software can be downloaded freely and has useful features such as (1) word frequency, used to determine the number of occurrences of words in the corpus; (2) concordance, containing a list of combinations of a word that is in a context; (3) collocation, containing the occurrence of words that are paired with other words in a context; and (4) N-Gram or clusters, containing sequences of two or more words that appear repeatedly in the text.

Based on the need for compiling Arabic for tourism vocabulary lists, this study aimed to provide information related to Arabic tourism websites that can be used as a reference in compiling a vocabulary list for teaching Arabic tourism, as well as to provide examples of the results of a corpus processing application, namely AntConc with it features, namely word frequency, concordance, collocation, and N-Gram.

Method

This study used a combination of qualitative and quantitative approaches with descriptive and comparative methods using a corpus linguistic approach. A qualitative approach was used to describe Arab tourism sites and the steps for compiling a corpus-based vocabulary list. The quantitative approach is relevant with this study which aimed to describe the frequency of words and phrases contained in the corpus. The data sources for this research were articles on tourism websites in Arabic and various previous literatures.

Result and Discussion

Linguistic Principles in the Preparation of Teaching Materials

One of principles in the preparation of teaching materials in Arabic is linguistic principle. The use of *qaaimat al-mufradat* (vocabulary list) in the preparation of teaching materials in Arabic is crucial to provide students with vocabulary close to them, especially concrete vocabularies which can be seen by the sense of sight, can be found in the learning materials, has a relationship with the vocabulary of other specific topics. The vocabulary list is composed from the concrete to the abstracts words and from low frequency to high frequency words, and is repeated and reduced gradually.[9]

---

[7] Nur Hizbullah, Fazlurrahman, & Fauziah, "Linguistik Korpus dalam Kajian dan Pembelajaran Bahasa Arab di Indonesia", *Prosiding Konfererensi Nasional Bahasa Arab,* Vol. 1, No. 2, 2016.

[8] Azzahra, "Penyusunan Kamus Kedokteran Arab-Indonesia dengan Pendekatan Linguistik Korpus", *Tsaqofiya: Jurnal Pendidikan Bahasa dan Sastra Arab,* Vol. 2, No. 2, 2020, 60–66.

[9] Aliwafa, "Revitalisasi Asas Penyusunan Bahan Ajar Bahasa Arab Untuk Perguruan Tinggi", *Islamedia: Jurnal Komunikasi Dan Informasi Kegamaan,* Vol. 13, No. 1, 2012.

In addition, error analysis (*tahlil al-akhtaʾ*) of students in several skills, such as reading and writing Arabic can be used as a reference in compiling teaching materials to guide students to avoid the mistakes they have made, for example, grammatical errors, vocabulary writing, or vocabulary pronunciation.[10]

Corpus Linguistics

Corpus linguistics is an empirical mindset linguistic latest in a rapidly growing since the early 90s.[11] The term corpus itself is defined by Hunston as a collection of natural language examples consisting of several sentences from a series of written texts or several records that have been collected for linguistic studies.[12] The text which consists of two forms, namely written and spoken, is then arranged systematically. The corpus is called natural language because the text collected is produced and used fairly and as it is.

Baker classified three aspects as considerations in understanding the concept of the corpus:[13]

1) The main corpus is a collection of texts generated electronically and can be analyzed automatically or semi-automatically.

2) The corpus is not only a collection of written texts but also speech.

3) The corpus probably includes most of the text that comes from a variety of sources.

At first, the corpus was in the form of hard files, which came from a collection of articles, journals, textbooks, text books, literary works (poems, short stories, and novels), newspapers, magazines, or it could also be in the form of broadcast recordings of conversations, interviews, and others. When digital technology grows fast, the corpus can be collected from the internet, such as websites, online news and newspapers, social media conversations, and so on.[14]

Corpus in written form can be obtained through collection from various sources, such as articles in newspapers, journals, literary works (poems, short stories, novels) and correspondences. Meanwhile, corpus materials in oral form can be obtained through recordings of several activities such as face-to-face informal conversations, telephone conversations, lectures, interviews, debates and discussions.[15]

---

[10] Aliwafa, "Revitalisasi Asas Penyusunan Bahan Ajar Bahasa Arab Untuk Perguruan Tinggi".

[11] Rajeg & Rajeg, "Pendekatan Linguistik Korpus Untuk Kajian Metafora Konseptual Bahasa Indonesia", *INA-Rxiv*, 2019.

[12] Arum & Winarti, "Penggunaan Linguistik Korpus Dalam Mempersiapkan Bahan Ajar English For Specific Purpose Di Bidang Radiologi", *Jurnal Teras Kesehatan,* Vol. 2, No. 2, 2020, 58–69.

[13] Azzahra, "Penyusunan Kamus Kedokteran Arab-Indonesia Dengan Pendekatan Linguistik Korpus".

[14] Azzahra, "Penyusunan Kamus Kedokteran Arab-Indonesia Dengan Pendekatan Linguistik Korpus".

[15] Setiawan, "Korpus dalam Kajian Penerjemahan", *Dipresentasikan Pada Seminar Nasional Perspektif Baru Penelitian Linguistik Terapan: Linguistik Korpus dalam Pengajaran Bahasa,* UNY, 2017.

Format and Characteristics of Corpus-Based Teaching Materials

The corpus can be defined as a collection of data, both ordinary and digital data in written form containing linguistic information, ranging from word level, structure, meaning, and discourse, which can be used for research.[16] Corpus linguistics has been used successfully in studying all aspects of linguistics (from lexical to grammatical) and language use. Romer said that the use of corpus linguistics analysis can be applied at least on three major areas of research, namely the field of linguistics with all branches of study, field of literature with a variety of text, and the field of language teaching which includes all process from beginning to end.[17]

The corpus in language teaching can be used as a source of information for a language, and as a domain to explore more deeply foreign language. One aspect that makes the corpus important in language teaching is the systematically arrangement of its empirical data.[18] Furthermore, the ability of computers to process large amounts of data is also the reason why the corpus can be an important and practical analytical tool in language teaching and research.[19]

Corpus-based teaching materials have been widely used in language teaching. One of them is the use of dictionaries. Dictionaries can be compiled using corpus software with the initial step of collecting a vocabulary list first. The data entered into the corpus must also be updated in accordance with the development of society, culture, and technology, all of which can certainly have a major influence in the realm of language.

Studies on a variety of linguistic aspects such as speech, vocabulary (lexicon), the meaning of the word / phrase / clause (semantics), language and society (sociolinguistics), the use of language (pragmatic), language and culture, and others have been using the corpus as accurate and representative database.[20]

The Arabic language corpus has been made in various countries with all their own uniqueness.[21] In the Sketch Engine application, for example, there is a corpus containing approximately 7.4 billion words taken from a number of sources. Then the Alsubaiti corpus was released more complete with 18 types of corpus taken from various sources and used for various specific fields of study that are on the University of Leeds sub-page, including the Corpus of Contemporary Arabic, Arabic Gigaword,

---

[16] Nur Hizbullah et al., "Source-Based Arabic Language Learning: A Corpus Linguistic Approach", *Humanities & Social Sciences Reviews,* Vol. 8, No. 3, 2020, 940–954.

[17] Hizbullah et al., "Source-Based Arabic Language Learning: A Corpus Linguistic Approach", 940-954.

[18] Wirza, "Aplikasi Software Concordance Program dalam Pengajaran dan Penelitian Bahasa (Studi Kasus Pada Mahasiswa Semester 6, Jurusan Bahasa Dan Sastra, Universitas Pendidikan Indonesia)", *Majalah Ilmiah UNIKOM,* 2011.

[19] Milroy, *Observing and Amalysing Natural Language,* (UK: Basil Blackwell Ltd, 1987).

[20] Wirza, "Aplikasi Software Concordance Program dalam Pengajaran dan Penelitian Bahasa (Studi Kasus Pada Mahasiswa Semester 6, Jurusan Bahasa Dan Sastra, Universitas Pendidikan Indonesia)".

[21] L al-Sulaiti and Atwell, "The Design of a Corpus of Contemporary Arabic (CCA)", *Research Report,* The University of Leeds, 2003.

and the International Corpus of Arabic by the University of Alexandria, Egypt. A number of the corpus is divided into two groups, paid download and unpaid ones.[22]

The corpus can also be processed using some applications available on the internet with a variety of features in it. There are some that can be downloaded for free, such as Nooj, TextStat, MonoconcEsy, Aconcord, and AntCont. There are also paid ones, such as WordSmith.[23] The results given by the corpus processing application are in the form of statistical data. So interpretation is needed in order to make the results understandable. Several features of technical tools (*corpus tools*) in this corpus processing application are very useful for reviewing the content of a text in a particular language. This tool can help learning Arabic, more specifically helping the preparation of Arabic for Tourism vocabulary. Among these features and their uses are:

1. Word frequency

The frequency feature is useful to know the number of occurrences of a word in a corpus or text.[24] An analysis on frequency of a word can help researchers identify the words that most frequently appear in a corpus, then comparing and distinguish it from other words. This can help the preparation of language teaching materials, where students know which vocabulary is often used in a subject and which vocabulary is rarely used. The list of basic vocabulary becomes the starting point for the preparation of teaching materials.[25]

2. Concordance

Concordance is a list or sequence of several examples of a word, part or combination of a word that is in a context and sourced from the corpus text.[26] This feature is an important aspect of corpus linguistics that helps the qualitative analysis of corpus data. Concordance can analyze data by looking at the linguistic features attached to a word in addition to looking at the form of the word itself as well the words around it.[27] Concordance provides a real example of how a word is used in context. This makes it easier for the teacher to explain the meaning of vocabulary as teaching material in depth, along with examples of its use in a real sentence.

---

[22] Nur Hizbullah, Fazlurrahman, & Fauziah, "Linguistik Korpus dalam Kajian dan Pembelajaran Bahasa Arab di Indonesia".

[23] Nur Hizbullah, Fazlurrahman, & Fauziah, "Linguistik Korpus dalam Kajian dan Pembelajaran Bahasa Arab di Indonesia".

[24] McEnery and Hardie, *Corpus Linguistics: Method, Theory and Practice*, (Cambridge University Press, 2011).

[25] Al-Naqah, "Khittah Muqtarahah Li Ta'lif Kitab Asasiyy Li Ta'lim Al-Lughah Al- 'Arabiyah Li Al-Natiqin Bi Ghayriha", *Dalam Waqai' Nadawat Ta'lim Al- Lyghah Al-'Arabiyah Li Duwal Al-Khalij,* 1985.

[26] Baker, Hardie, and McEnery, *A Glossary of Corpus Linguistics*, (Edinburgh: Edinburgh University Press, 2006).

[27] Arum and Winarti, "Penggunaan Linguistik Korpus dalam Mempersiapkan Bahan Ajar English for Specific Purpose di Bidang Radiologi", 58-69.

3. Collocation or word sketch

Collocation is the occurrence of words that are paired with certain other words in a context and field of meaning. Baker states that in corpus processing applications, collocations are usually pairs of two or more words.[28] Collocations or word sketches help the preparation of teaching materials in determining which vocabulary should be paid more attention, based on how many word pairs occur in a corpus.

4. N-Gram atau *clusters*

N-Gram contains a sequence of two or more words that appear repeatedly in a text with a significant number to be studied with certain assumptions.[29] N-Gram is useful to group frequently used phrases in the corpus text. This is useful for adding to the vocabulary of teaching materials, making it more interesting materials.

Arabic Tourism Website

Website can be used as a general learning source which provides the readers with various references such as books, modules, and teaching materials which teachers can access for free and use it as teaching material sources of Arab *fusha* and its grammar and structure.[30] The format of the material provided by the website is usually presented in written text, sound / audio material, and audiovisual material. If converted into a corpus, the material format has conformity, including the format of a text collection, an audio corpus, and an audiovisual collection.

These various websites can be used as a source of material by students as well as teaching materials for teachers in learning Arabic at various levels, from basic to advanced levels.[31] More specifically, it is very useful in the preparation of Arabic for specific purposes teaching materials, one of which is in the realm of tourism. Given that the world of tourism has many reading references on the internet with the availability of sites that discuss tourism topics in Arabic, and there is a special vocabulary that is only used in the world of tourism. Tourism webites in Arabic are considered beneficial to facilitate in classifying Arabic vocabularies frequently used and in the field of tourism. The authors found the following Arabic language tourism websites:

---

[28] Nur Hizbullah, Fazlurrahman, and Fauziah, "Linguistik Korpus Dalam Kajian Dan Pembelajaran Bahasa Arab Di Indonesia."

[29] Baker, Hardie, and McEnery, *A Glossary of Corpus Linguistics.*

[30] Nur Hizbullah et al., "Source-Based Arabic Language Learning: A Corpus Linguistic Approach".

[31] Idris Mansor & Ghada Salman, "Arabic for Tourism: Guidelines for Linguists and Translators", *Arab World English Journal,* Vol. 7, No. 3, 2016.

1. Mawdoo3.com (https://mawdoo3.com/)

Mawdoo3 is an online platform that provides thousands of Arabic articles with the latest information in various fields, one of which is tourism, not only tourism in the Arab region, but throughout the world. This site provides a voice feature to search for and listen to articles in Arabic by native speakers. This can be useful in learning Arabic in the preparation of teaching materials through access to copy of tourism articles to be included in the corpus processing application and for students by accessing this site to improve language skills.

2. Ootlah.com (https://www.ootlah.com/ar/blog/category/where-to-go.html)
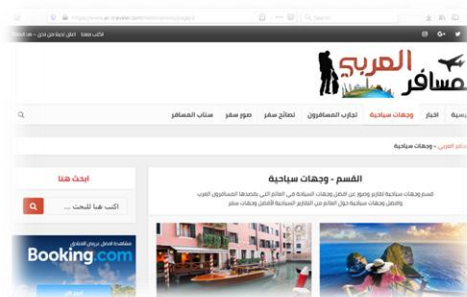
The Ootlah website offers a large selection of the best tourist spots for vacations along with cost deals with the best travel agencies in the destination city. This site has two languages, namely Arabic and English, in the sense that all articles can be read in two languages at once. This can help students improve their skills in both languages by accessing this site. In addition, teachers can also use this site as a reference in compiling a Arabic for Tourism vocabulary list.

3. Ar-Traveler.com (https://www.ar-traveler.com/destinations)
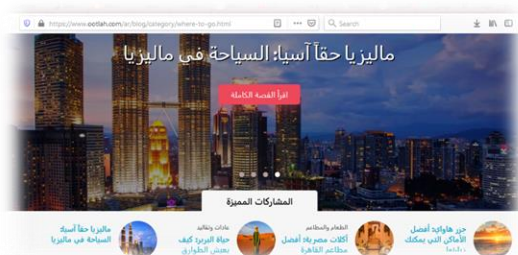
This ar-traveler website specifically discusses tourism around the world in Arabic. The site consists of several sections, including news section, tourist destinations, tips and directions related to the world of tourism, hotel and travel agency offers. With various sections on this site, the Arabic vocabulary and terms in it are very varied. This is useful for the preparation of corpus-based Arabic for Tourism vocabulary

4. TourFlag.com (https://tourflag.com/)
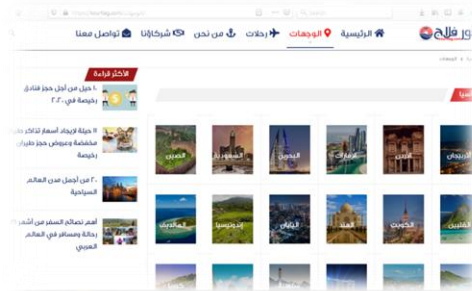
TourFlag is to provide information and referrals related to everything related to travel and tourism, the most famous tourism countries, and the cities of the loveliest in each country based on tourists' opinions, as well as show you tours of the most frequently visited city. The site also provides comparison of cost tourism among travel agents. With these features, teachers can find complete Arabic tourism vocabulary and terms.

The Compilation of Corpus-based Vocabulary List

A number of above mentioned websites can be used as material for compiling Arabic for Tourism vocabulary list through a corpus application called AntConc. The following are the stages of using the AntConc:

1. Copy a number of articles on the website into Microsoft Word by removing the editorial elements (numbers, symbols, images, etc.) and then converting them into plain text.
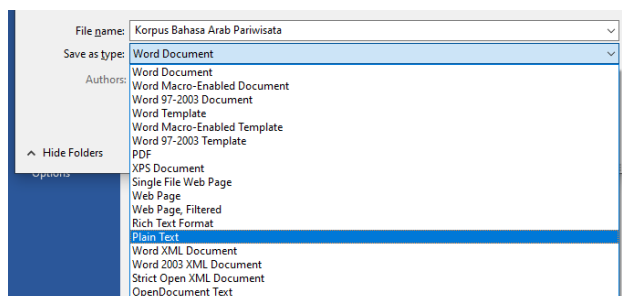

Figure 1.1: Selecting the *plain text* type when saving a copy of the article

2. Click the other encoding option then look for the UTF-8 *encoding* format. This format is used so that the document can be processed in the AntConc application.
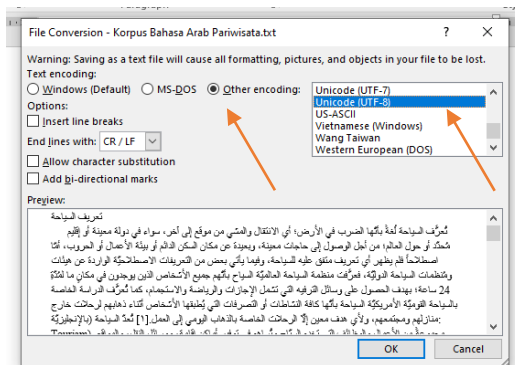
Figure 2.1: Selecting other encoding format, then click UTF-8

3. Open the AntConc application and enter the document in plain text format by clicking the file button, then open file.
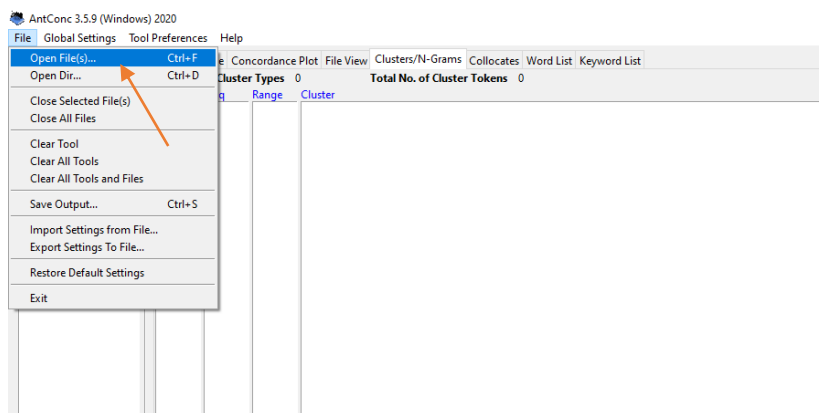
Figure 3.1: Inserting a copy of an article in plain text format into AntCont

4. Using word frequency feature or wordlist

Click *wordlist* feature and click *start,* then appears a list of words that can be selected sequentially in the feature *sort by* under button *start,* then order based on frequency of occurrence or alphabetical order.

Figure 4.1: word frequency feature

The following are some of the results of processing a number of tourism articles on the above website using the AntConc application. This data has been reduced by selecting the vocabulary of a verb and a noun related to tourism and given a transliteration and translation.

Word Types: 3.711 | *Word Tokens*: 10.601

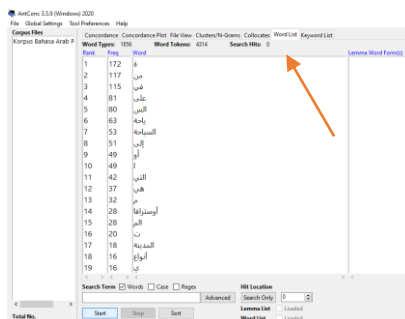| No	Frequency	Word	Transliteration	Translation
1	50	الأماكن	*al-amâkin*	Places
2	26	مول	*mûl*	Mall
3	19	الرائعة	*ar-râ`i'ah*	Amazing
4	18	الخلابة	*al-khallâbah*	Enchanting
5	18	ساحة	*sâhah*	Square/page
6	17	بوكينج	*bûkīnj*	Booking
7	16	قصر	*qashr*	Palace
8	15	تتمتع	*tatamatta'*	Enjoy
9	15	معالم	*ma'âlim*	Places
10	15	الاستمتاع	*al-`istimtâ'*	Enjoyment
11	14	المجانية	*al-majjâniyyah*	Free
12	12	الطبيعية	*Ath-thabî'iyyah*	Natural
13	12	المناطق	*al-manâthiq*	Region
14	11	التاريخية	*at-târīkhiyyah*	Historical
15	10	تتميز	*tatamayyaz*	Has privileges
16	10	تتوفر	*tatawaffar*	Available
17	10	فنادق	*fanâdiq*	Hotels
18	9	التراث	*at-turâts*	Heirloom
19	9	منتزه	*muntazah*	Park/recreation area
20	8	الجذب	*al-jazb*	Attractiveness
21	7	الترفيهية	*at-tarfîhiyyah*	Entertainment
22	7	المحيط	*al-muhīth*	That surrounds
23	6	المغامرة	*al-mugâmarah*	Adventure
24	5	السياح	*as-suyyâh*	The tourists
25	5	متنوعة	*mutanawwi'ah*	Varies
26	5	الاسترخاء	*al-istirkhâ`*	Relax
27	5	المذهلة	*al-muzhilah*	Awesome
28	5	الملونة	*al-mulawwanah*	Colorful
39	4	الشواطئ	*asy-syawâthi`*	Beaches
30	3	ممتع	*mumti'*	Fun
32	3	مميزة	*Mumayyizah*	Characteristic

Table 1. List of vocabularies resulted from AntConc

Word types is the number of types of words in the corpus. The number is 3,711 words which are then reduced as needed. Meanwhile, *words token* is the total number of words in the corpus, including the repeated words. *Words tokens* can also be interpreted as the total of all words resulting from the addition of the frequency of

occurrence of each word. This corpus contains a total of 10,601 words which were then reduced by choosing verbs or nouns related to tourism and at least three times the frequency of occurrence.

In table 1 it can be seen that the word الأماكن/ *al-amâkin* / 'places' occupies the first position as the highest frequency word with fifty times occurrence; while the word مميزة/ *mumayyizah* / characteristics is the word with the lowest occurrence, which is 3 times. This data can make it easier for teachers to choose vocabulary according to the desired field.
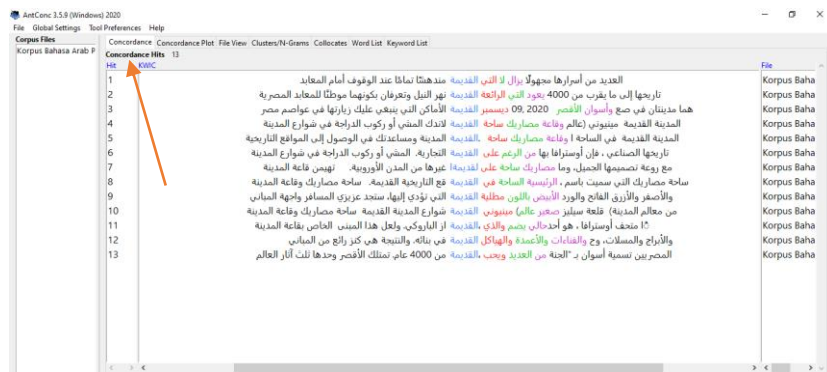
5. Using *corcondance* feature

Figure 5.1: Concordance feature

The following is the concordance of one of the vocabulary words, namely التفيهية/ *at-tarfîhiyyah* / 'entertainment' or can be interpreted as ' *recreation'* in certain contexts.

No	Word After	Word	Words Before
1	الأولى في دول الخليج	الترفيهية	بل إنها تعد الحديقة
	It is the first recreational park in the gulf countries		
2	المتنوعة، حيث روعة الحدائق والمتنزهات	الترفيهية	الفرص
	Diverse recreational opportunities, such as the splendor of parks and gardens		
3	في العين، بل إنها من أهم الأماكن السياحية	الترفيهية	تأتي حديقة الحيوان ضمن أبرز الأماكن
	The zoo is one of the most famous places of entertainment in Al Ain, and even one of the most important tourist attractions		
4	والنباتية الممتعة	الترفيهية	فضلا عن الحدائق
	As well as recreational parks and fun plants		
5	والألعاب وصالة البولينج والتزلج	الترفيهية	فمع وجود العديد من الأنشطة
	With lot of recreational activities, games, bowling and skiing		

Table 2. concordance of word الترفيهية

The concordance feature can help teachers define the meaning of the vocabulary list and the context of its use correctly after going through several stages of analysis based on the example sentences. In addition, this feature is helpful to

distinguish vocabulary in the form of compound words or idioms and which ones are not. In table 2, it can be found that the word الترفيهية / *at-tarf hiyyah* / can be interpreted as 'entertainment' when paired with the word ال / *al-had qah* / 'garden' and can also be interpreted as 'recreation'. The choice of meaning is in accordance with the context of the sentence used. Furthermore, teachers can also provide some examples of the use of a word in a sentence accurately and in accordance with the application of the word in day-to-day lives.
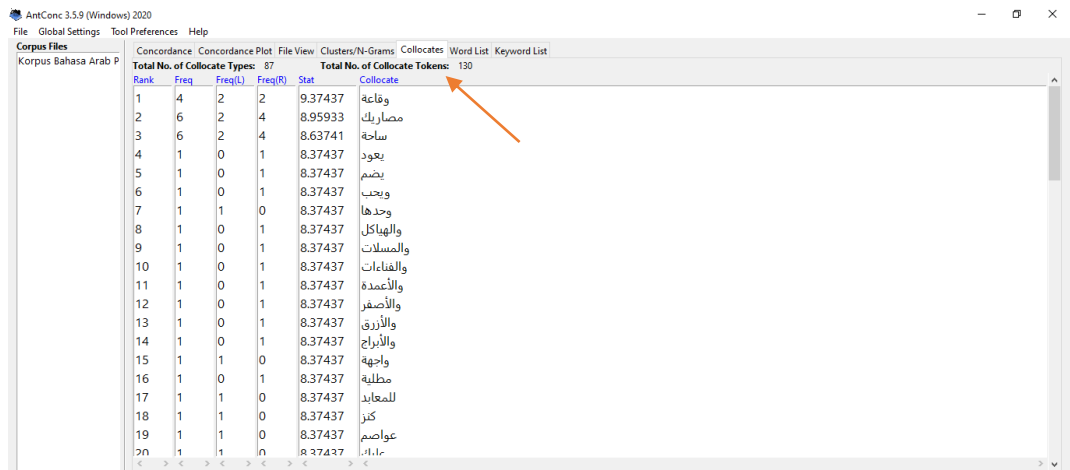
6. Using collocation feature

Figure 6.1: Collocation feature

The following is an example of a word that is collocation or paired with the word الخلابة /al-khallâbah/ 'amazing'. The right and left frequencies means the place where the collocation appears.

No	Frequency	Frequency (right)	Frequency (left)	Collocation
1	4	3	1	الطبيعي /*ath-thabî'i*/ 'natural'
2	3	3	0	المناظر /al-manâzhir/ 'scenery'
3	3	3	0	المشاهد /al-musyâhid/ 'scenery'
4	3	3	0	الطبيعية /ath-thabî'iyyah/ 'natural'
5	1	1	0	البركانية /al-burkâniyyah/ 'volcanic'
6	1	1	0	الشواطىء /asy-syawâthi`/ 'beaches'
7	1	1	0	الإطلالة /al-`ithlâlah/ 'good-looking'

				المائية	
8	1	1	0	/al-mâ`iyyah/ 'waters'	
				الحدائق	
9	1	1	0	/al-hadâ`iq/ 'gardens'	

Table 3. Collocation of word الخلابة

The collocation feature in the AntConc software can make it easier for teachers to identify collocation words and determine their meaning. For example, the collocation of the word الخلابة /al-*khallâbah*/ 'charming' in table 3. The data show that the word is most often paired with the word الطبيعي /*ath-thabī'ī*/ 'natural'.

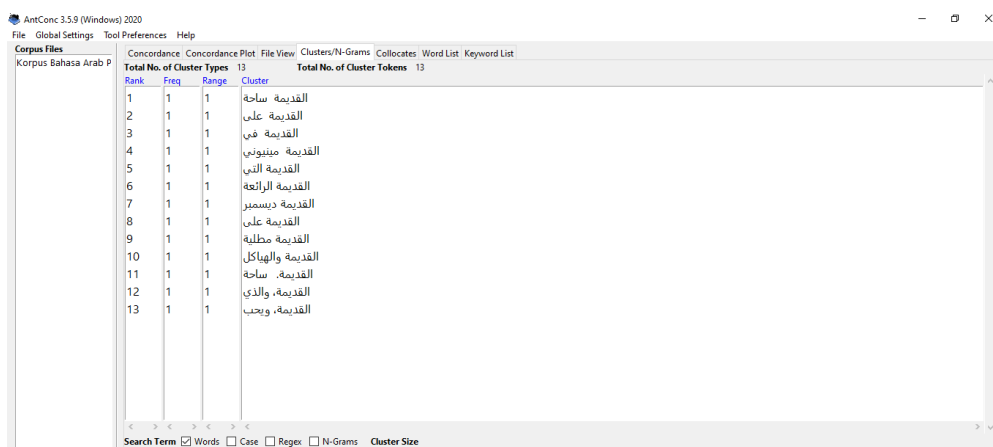7. Using N-Gram clusters feature

Figure 7.1 Features of N-Gram or Clusters

This feature is used to identify phrases from a vocabulary, example phrases قصر / *qishr* / 'palace' which appears repeatedly in the corpus along with the frequency with which it occurs. This feature can increase knowledge related to the list of popular Arabic for tourism terms used to those rarely used.

No.	Frequency	Cluster
		قصر المويجعي
1	8	/qishrul muwayji'ī/ 'al-muwaiji palace'
		قصر العين
2	5	/qishrul 'ain/ 'al-ain palace'
		قصر فينيسي
3	1	/qishr fīnīsiy/ 'venesia palace'
		قصر كورسيني
4	1	/qishr kûrsīnī/ 'istana corsini'
		قصر مجدد
5	1	/qishr mujaddid/ 'renovated palace'

Table 4. N-Gram of word قصر

After all the data was collected and reduced based on its suitability with the intended theme, the data classification process was then carried out based on the desired category. The results of the above analysis can be used as a reference in compiling an Arabic for Tourism vocabulary list. If it is further processed, it can be used as test practices by asking students to make sentences containing the vocabulary or phrases above, and be used to assist in the preparation of dictionaries as companion teaching materials.

Conclusion

The preparation of a list of vocabulary of Arabic for tourism can take benefit of several Arabic websites such as Mawdoo3.com, Ootlah.com, Ar-Traveler.com, and TourFlag.com. In processing a number of these sites, it involves the help of a corpus processing application, namely AntConc. Articles in a number of these sites are copied into Microsoft Word and then converted into plain text with UTF-8 *encoding* format for processing in AntConc through its features: word frequency, concordance, collocation, and N-Gram.[]

REFERENCES

Al-Naqah. "Khittah Muqtarahah Li Ta'lif Kitab Asasiyy Li Ta'lim Al-Lughah Al-'Arabiyah Li Al-Natiqin Bi Ghayriha", *Dalam Waqai' Nadawat Ta'lim Al- Lyghah Al-'Arabiyah Li Duwal Al-Khalij,* 1985.

Aliwafa. "Revitalisasi Asas Penyusunan Bahan Ajar Bahasa Arab Untuk Perguruan Tinggi", *Islamedia: Jurnal Komunikasi Dan Informasi Kegamaan,* Vol. 13, No. 1, 2012.

Arum, E. R., and W Winarti. "Penggunaan Linguistik Korpus Dalam Mempersiapkan Bahan Ajar English For Specific Purpose Di Bidang Radiologi", *Jurnal Teras Kesehatan,* Vol. 2, No. 2, 2020,

Azzahra, S. F. "Penyusunan Kamus Kedokteran Arab-Indonesia dengan Pendekatan Linguistik Korpus", *Tsaqofiya: Jurnal Pendidikan Bahasa dan Sastra Arab,* Vol. 2, No. 2, 2020.

Baker, P., A Hardie, and T McEnery. *A Glossary of Corpus Linguistics.* Edinburgh: Edinburgh University Press, 2006.

Bin Mat Zin, Hashim., & Tengku Ghani Tengku Jusoh, "The Potential of Arabic as a Tourism Language in Malaysia", *Journal of Educational and Social Research,* Vol. 3, No. 7, 2013.

Dahlan, Juwairiyah. *Metode Belajar Mengajar Bahasa Arab.* Surabaya: Usaha Nasional, 1992.

Halim, N. "Bahasa Arab Dengan Tujuan Khusus Berbasis Komunikatif Wisata Travelling." *Bintang,* Vol. 2, No. 3, 2020.

Hizbullah, N., Z. Arifa, Y. Suryadarma, F. Hidayat, L. Muhyiddin, and E. Kurnia Firmansyah. "Source-Based Arabic Language Learning: A Corpus Linguistic Approach", *Humanities & Social Sciences Reviews,* Vol. 8, No. 3, 2020.

Hizbullah, Nur., F. Fazlurrahman, and F. Fauziah. "Linguistik Korpus dalam Kajian dan Pembelajaran Bahasa Arab di Indonesia", *Prosiding Konfererensi Nasional Bahasa Arab,* Vol. 1, No. 2, 2016.

Ibrahim, Zulkifli., Zulhan Othman, Mohd S. Zahar, Kamaruzaman Jusoff, and Maimunah Sulaiman. "Travelling Pattern & Preferences of the Arab Tourists in Malaysian Hotels", *International Journal of Business and Management*, Vol. 4, No. 7, 2009.

Mansor, Idris., & Ghada Salman, "Arabic for Tourism: Guidelines for Linguists and Translators", *Arab World English Journal,* Vol. 7, No. 3, 2016.

McEnery, T., and A. Hardie. *Corpus Linguistics: Method, Theory and Practice.* Cambridge University Press, 2011.

Milroy, Lesley. *Observing and Amalysing Natural Language.* UK: Basil Blackwell Ltd, 1987.

Misran, M. "Dialek 'Ammiyyah dalam Pengajaran Bahasa Arab untuk Pariwisata di Indonesia", *Adabiyyāt: Jurnal Bahasa dan Sastra,* Vol. 12, No. 2, 2013.

Rajeg, G. P. W., and I. M Rajeg. "Pendekatan Linguistik Korpus Untuk Kajian Metafora Konseptual Bahasa Indonesia", *INA-Rxiv*, 2019.

Setiawan, T. "Korpus dalam Kajian Penerjemahan", *Dipresentasikan Pada Seminar Nasional Perspektif Baru Penelitian Linguistik Terapan: Linguistik Korpus dalam Pengajaran Bahasa,* UNY, 2017.

al-Sulaiti, L., and Atwell, "The Design of a Corpus of Contemporary Arabic (CCA)", *Research Report,* The University of Leeds, 2003.

Suryadarma, Yoke., and Alinda Z. Fakhiroh. "Optimalisasi Penggunaan Corpus Linguistics Dalam Penyusunan Kamus Az-Ziro'ah Sebagai Media Pembelajaran Bahasa Arab", *International Seminar on Language, Education, and Culture (ISOLEC)*, 2020.

Wirza, Y. "Aplikasi Software Concordance Program dalam Pengajaran dan Penelitian Bahasa (Studi Kasus Pada Mahasiswa Semester 6, Jurusan Bahasa Dan Sastra, Universitas Pendidikan Indonesia)", *Majalah Ilmiah UNIKOM*, 2011.