

Airline Collocations: Frequency Based Analysis with COCA as a Corpus

Bonifasia Ekta Fima Natalia

Sekolah Tinggi Ilmu Ekonomi SBI

***Corresponding Author**

Email: bonifasiaekta.sbi@gmail.com

Abstract

Collocations are words that co-occur together in any text and have definite association. The use of collocations is part of linguistic awareness which defines how language is used naturally. This study aims to analyze the collocations used in airlines by using Corpus of Contemporary American English (COCA). The data were analyzed both quantitatively and qualitatively. Quantitatively, the data were calculated regarding the frequency, topics and cluster. Qualitatively, the data were verbally analyzed, described, and discussed. The big data is taken from 1990-2019. The frequent words in area of airlines management found in COCA are used as keywords to find concordance of other words or phrases. The results of the study shows that from the three words taken from the script of flight announcement in airport 'flight', 'passenger', and 'boarding' it was found that the frequency risen is 73,870. The topics related to the three words of airlines found is 22,798. The clusters found from the three words by COCA is 34,006 which 'flight' word has the largest clusters (18,730). From the COCA analysis, the new words or idioms related to airlines can be used as teaching material in vocabulary for aviation or airlines students.

Keywords: *corpus, collocations, airlines, COCA, frequency*

INTRODUCTION

The contributions of technology towards the linguistic over current decades have been evolved. One of them is the corpus-based technology in language instruction. A corpus is virtually worthless without some kind of computer software tool to process it and display results in an understandable way (Anthony, 2004). A large English lexical data base can be a big data used for developing language instructional model.

The term collocation was first introduced by Palmer, an English teaching expert in 1933, to explain the succession of two or more words which must be studied as an inseparable unity and cannot be defined based on the meaning of each component (Kennedy, 2002). The term collocation was later popularized in the context of linguistics by Firth in 1957 to explain the combination of words that have a permanent association (Gabrielatos, 1994). From the point of view of native speakers, the competence to use collocation is a common aspect of language and is part of metalinguistic consciousness, a phrase that uses to explain "the knowledge of language percept, parts of language, and how language works, or even a simple self-recognition of one's own language and the forms being used" (Shaw, 2011).

May Fan (2009) in his research on the use of collocations by English learners concluded that English foreign speakers use collocations in smaller amounts than native speakers and have difficulty in using collocation appropriately. Fan also found the effect of the combination of words forming the collocation of

their mother tongue when using English. Such conditions, according to Fan, must be accommodated by English language teachers with vocabulary teaching which increases the awareness of language about idiomatic and problematic language aspects.

The use of collocation orally is as important as in written. In oral situation, a person's words will be more simply understood if the speaker uses collocation properly. One of authentic usage of collocation orally is announcement. Announcement in airport, for instance, changes text to speech system. In the airport, communication takes place at different levels - from the time a traveller arrives to check-in to the time he boards the flight, till he disembarks the aircraft. In Indonesia, the use of English language for flight announcements is not just limited to international flights but also domestic ones. A flight announcer's duty is to give information about flight schedules, make boarding calls for travellers, inform the public about an inbound flight, as well as make other overhead announcements.

In studying vocabulary, very large data is needed. Nowadays data search is greatly facilitated by technology. For example, an online dictionary is a practical dictionary for digital learners. With the help of the application, they easily find out the meaning and context of the use of a collocation. The lists needed to be ordered, so learners could learn the more common words first. To analyze the collocation, the researcher used the Corpus of Contemporary American English (COCA). From this perspective, the corpus can provide various examples in authentic context, so language learners can explore and discover the meaning, form, and use of words (Ahn, 2014).

The analysis of corpus in this study covers frequency (which found in blogs, webs, TVs/movies, spoken, fiction, magazines, newspapers, and academic), topic, collocates, synonym, and cluster. The big data is taken from 1990-2019. The frequent words in area of airlines management found in COCA are used as keywords to find concordance of other words or phrases.

The studies related to corpus analysis have been widely conducted, whereas the study on airline collocation is still limited. By using COCA, this study, therefore, aims to analyze the English language frequently used by flight announcers in Indonesia based on corpus. It also to find out the most frequent words relates to airline that can be used by pre-service airline students in learning English.

LITERATURE REVIEW

Studies on Corpus Linguistics

There were some previous studies related to the use of corpus. Mansour (2017) studies on how to direct students to a more effective use of Corpus of Contemporary American English (COCA) in order to foster their learning process of different English collocations, and use English collocations accurately in their academic writings. The research illustrates search mechanisms and search strings that could be implemented effectively on COCA to get more representative and accurate collocation results. The research results show significant development of using collocations when the students translate academic texts from Arabic into English, or when they write an academic essay.

Hussein and Sawalha (2016) do the research on the forms of similes and the types of nouns following them and their frequency in different simile constructions.

The study focuses on the most frequent nouns following similes in different constructions in two generalized corpora, namely the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). The research results indicate that similes, unlike other multi-word units are changeable as different nouns with varying frequencies ranging from one to fifty-three are shown to follow each of the similes under investigation. There may not be only one correct simile form which is given priority over others but rather several correct and perhaps equally acceptable forms. They suggest the liberal tendency towards the use of similes in native English be extended to the context of English as a foreign language so that students can enrich the language by their creative, innovative, and idiosyncratic forming of similes based on the totality of their experience in language and culture.

Bardovi-Harlig, et al (2015) studied how language teachers can use a corpus to develop pragmatically authentic materials. Using the context of academic discussion, they explain how an online corpus can be applied in teaching pragmatic routines found in the cognition of four speech acts generally found in academic discussion: disagreements, agreements, and self- and other-clarifications.

Frankenberg-Garcia (2014) explored the benefits of separate corpus examples for comprehension and production in secondary school in Portugal. The research problems are [1] to examine whether corpus examples will help a different population of language learners and [2] to find a clearer and more specific picture of the value of separate encoding and decoding examples. The examples were taken from some dictionaries such as: Oxford advanced learner's dictionary (OALD), Longman dictionary of contemporary English, COBUILD dictionary, BNC, COCA and UKWaC.

Corpus Linguistics

The term corpus (plural: corpora) according to Leech (McEnery, et al, 2006) refers to the collection of verbal material that is set with a specific objective. The corpora have been well evolved. It is a corporation which is set as a source language learning and research, both oral and written. Romer (2011) studied pedagogical corpus applications and analyzed recent publications in the area of corpus linguistics and language teaching. It intent to describe how general language corpora, such as the British National Corpus (BNC) or the Corpus of Contemporary American English (COCA), and specialized corpora, such as the Michigan Corpus of Academic Spoken English (MICASE) or the International Corpus of Learner English (ICLE), can be applied in these applications.

A corpus analysis toolkit is purposed particularly for practice in the classroom (Anthony, 2004). The place of corpus linguistics in applied linguistics is a tool for developing material for classroom language instruction (Anthony, 2004) meanwhile Romer (2011) states that in the field of applied linguistics, more researchers and practitioners search what corpus linguistics has to offer to language pedagogy.

Linguistics being the scientific study of language and its structure, corpus linguistics is the study of language on the basis of text corpora (Sinclair 1991, Romer, 2011). A corpus is a collection of naturally-occurring language text, chosen to describe a state of a language variant (Sinclair, 1991).

The corpus-based frequency method is one of the methods used in science corpus linguistics. This branch of linguistics studies how linguistic analysis is applied to batch of texts that set a corpus (Meyer, 2002). High frequent words can reflect aspect of language accuracy by native speakers whose patterns can be imitated by foreign speakers. According to Romer (2008), high frequent vocabulary in a corpus can help students to understand the meaning and use of certain English words deeply both in receptive and productive language skills.

Farr (2008) argued that corpus software has a number of advantages for users: it produces word lists and counts occurrences of individual search items, it allows for the presentation and (re)organization of data in a way that facilitates the identification of patterns, it automatically produces cluster and collocation lists, and most software has a keyword tool, which considers a comparison of lexis between corpora to identify overall significant items.

Frequency information is automatically encoded, the second factor that should be allowed is how the frequency information is accessed. Researchers are investigating the subjective frequency estimation that starts out with the opposite assumption of some corpus linguists, i.e., that frequency estimates are reliable. However, as noted, there is a prominent qualification, estimates may only be as well as the presence of the key information compulsory to make the decision severely (Mcgee, 2009).

COCA Corpus

The Corpus of Contemporary American English (COCA) is a more than one billion words corpus of American English. It was created by Mark Davies, Professor of Corpus Linguistics at Brigham Young University. The corpus is used by approximately tens of thousands of people each month, which may make it the most widely used "structured" corpus currently available.

COCA is the largest free English corpus and has prominent benefit over other free corpora in terms of vocabulary study (Davies, 2009). It has comprehensive data from the whole period of the last 30 years – 20 million words per year from 1990-2019. COCA also has a large range of genres. The following table shows the genres in the corpus.

Table 1. Genre in the Corpus (COCA Corpus, released on March 2020)

Genre	#texts	#words
Spoken	44,803	127,396,932
Fiction	25,992	119,505,305
Magazines	86,292	127,352,030
Newspapers	90,243	122,958,016
Academic	26,137	120,988,361
Web (General)	88,989	129,899,427
Web (Blog)	98,748	125,496,216
T/Movies	23,975	129,293,467

COCA provides useful information about the frequency of words, phrases, and grammatical constructions across the genres in which they can be very informal (e.g. TV and movie subtitles or in spoken transcripts), or they are more formal (e.g. academic articles), or they are somewhere in between (e.g. magazines and newspapers). The great general balance (and large size) of COCA also enables the

users to see the frequency of related phrases across genres. The users can also compare two genres or sets of genres. The ability to focus in on specific genres means that the users can find “just the right word” for a particular concept in particular concept in a particular genre (The COCA corpus overview, 2020).

There is a very wide range of searches, including words, phrases, substrings, lemmas, part of speech, synonym, and customize wordlists. It just takes about one second to search through the billion words corpus to find matched strings.

METHOD

In this study both quantitative and qualitative approaches were used to understand and analyze the findings and quantitative results generated from the corpus were analyzed qualitatively. In this study, both quantitative and qualitative approaches were used to understand and analyze the findings. Quantitative research involves the collection and analysis of numerical data to identify patterns, trends, and relationships among variables. It focuses on objective measurements and statistical analysis to provide a broader perspective and generalizability of the findings. Qualitative research involves the exploration and understanding of complex phenomena through in-depth examination of subjective experiences, beliefs, attitudes, and behaviors.

By employing both quantitative and qualitative approaches, this study benefits from the strengths of each method. The quantitative data generated from the corpus, which comprises a large collection of texts, was analyzed quantitatively to derive numerical results and statistical patterns. These quantitative results provide valuable information about the frequency, distribution, and statistical significance of certain linguistic features or phenomena within the corpus.

However, it is essential to go beyond the quantitative results and delve into the qualitative analysis. The qualitative analysis involves a deeper exploration and interpretation of the quantitative findings, seeking to understand the underlying meanings, contexts, and nuances associated with the data. This qualitative analysis helps uncover insights that may not be apparent through numbers alone, providing a more comprehensive and nuanced understanding of the research topic.

This research used a simple random sampling data. The words being analyzed were randomly selected from the script of flight announcement in Adisucipto airport. Adisucipto International Airport Yogyakarta is known as the 4th busiest airport in Indonesia. It serves around 8 million passengers per year with around 103,000 aircraft movements (www.wikipedia.org, 2020). The announcement in Adisucipto airport presented in three languages: Indonesian, English and Javanese languages. The airport announcement was then transcribed into text. From the script, the researcher took three words which were discovered the most closely related to airlines. The researcher analyzed the English announcement by recording and doing transcription.

The script of announcement is as follows:

*“Attention please. This is a final **boarding** call for Lion Air -- **passengers on flight** number JT-650 leaving for Denpasar. Passengers should be on board through gate number 5 immediately. Thank you”.*

The researcher chose the words in bold: *flight*, *passenger*, and *boarding* since they are the closest words related to airlines. The words were analyzed by using COCA tools (<https://www.english-corpora.org/coca/>). By using COCA tools, the word ‘*passenger*’ can be analyzed its frequency, topics, collocates, synonyms, and clusters. However, in this study the researcher only limit the analyze on frequency, topics and clusters.

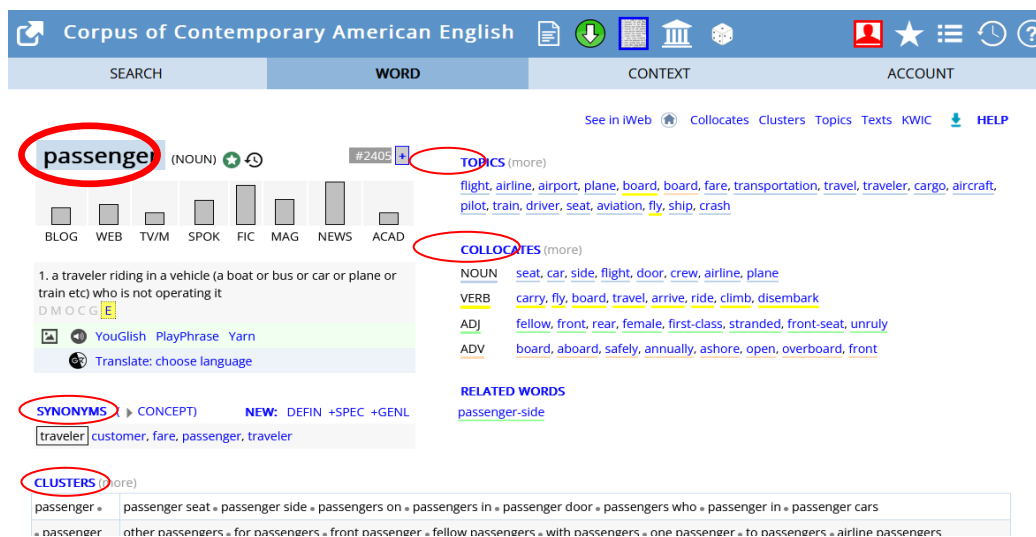


Figure 1. Searching word ‘passenger’ by COCA tool

The researcher input the word ‘*passenger*’ and found matching strings. From the word ‘*passenger*’ it can be found some collocations and topics, for example the word ‘*flight*’, which subsequently being analyzed as the previous word. Then the researcher analyzed the collocation in its frequencies, topics, collocates, synonyms, and clusters. The from the search results, the researcher only took the 10 highest frequency samples. They consist of 2, 3, or 4 clusters. The clusters found are only taken the 30 highest top to be analyzed.

The corpus ‘*passenger*’ is composed of 1,001,610,938 words in 485,179 texts, found during 1990 up to 2019 and been classified into blogs, webs, TV/movies, spoken, fictions, magazines, newspapers, and academics.

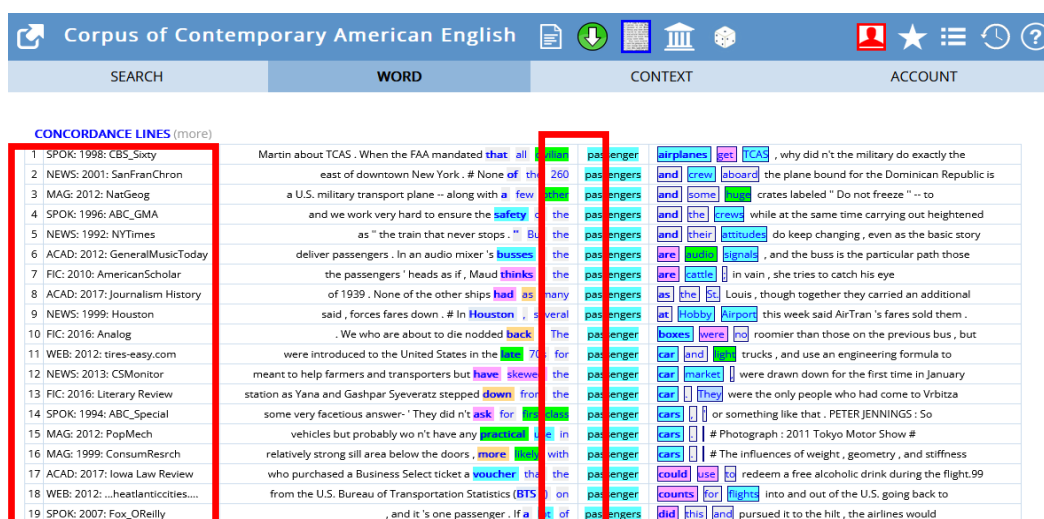


Figure 2. Key Words In Context (KWIC) Concordancer

In the same time, by using COCA tool, it can be found KWIC (Key Word in Context) related to the word 'passenger'. The data is presented in tables classified into 4 aspects. The most frequent words arisen can be used as teaching material in English especially for aviation or airlines students.

RESULTS AND DISCUSSION

From the flight announcement in Adisucipto International Airport, there are three words are taken and being analyzed, namely, 'flight', 'passenger', and 'boarding'. These three words are analyzed by using COCA tool in order to know their frequency, topics, collocates, synonyms, and clusters.

1. Frequency

The highest frequency of use 'flight', 'passenger', and 'boarding' can be seen in Table 2.

Table 2. Frequency and Topics

Collocation	Frequency	Related Topic
<i>flight</i>	51,869	<i>airline, plane, passenger, airport, fly, pilot, aircraft, flying, aviation, travel, airplane, jet, board, crew, carrier, fare, attendant, traveler, book</i>
<i>passenger</i>	15,469	<i>flight, airline, airport, plane, board, fare, transportation, travel, traveler, cargo, aircraft, pilot, train, driver, seat, aviation, fly, ship, crash</i>
<i>boarding</i>	6,532	<i>board, passenger, airport, flight, airline, Indian, luggage, plane, pass, reservation, taxi, passport, fare, native, policeman, traveler, tribe, crowded</i>
TOTAL	73,870	

The result shows that the word 'flight' is the most frequent collocation (51,869) compared to the word 'passenger' (15,469) and 'boarding' (6,532). The frequency of the words is found in blogs, webs, TVs/movies, spoken, fiction, magazines, newspapers, and academic. The total frequency appear from these three

words is 73,870. The details in which genres they are belonged is shown in Table 3, 4 and 5.

The findings indicate that the word *'flight'* occurs most frequently as a collocation within the corpus, with a count of 51,869 instances. This suggests that the term *'flight'* is highly associated with other words in the context of the analyzed data. It plays a significant role in the dataset, potentially indicating a focus on aviation-related topics or discussions.. In comparison to the word *'flight'*, the collocations with the words *'passenger'* and *'boarding'* occur less frequently. This comparison highlights the differences in usage and the relative prominence of these terms within the corpus. The contrasting frequencies between the collocations indicate varying usage patterns. The prominence of *'flight'* may suggest that discussions or texts within the corpus heavily focus on aspects related to aviation.

Table 3. Frequency of word *'flight'*

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
FREQ	51869	4669	5334	7171	7167	7624	8862	8436	2606
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8
PER MIL	52.23	36.30	42.93	55.99	56.82	64.43	70.28	69.29	21.75

The word *'flight'* is found most frequent in magazine, which is 8,862 times and the least frequent appears in academic which is 2,606 times. The variation in the frequency of the word *'flight'* across different genres, specifically magazines and academic texts, suggests that the usage and context of the term may differ based on the genre. Magazines often cater to a broader audience and may include articles related to travel, aviation, or leisure, where the term *'flight'* is more commonly used. Magazines often utilize a more informal or conversational style of writing, allowing for a broader use of terms like *'flight'*. Academic texts, however, adhere to a more formal and specialized language register, which may lead to a more restricted usage of certain terms or a preference for more discipline-specific terminology.

Table 4. Frequency of word *'passenger'*

SECTION	ALL	<u>BLOG</u>	<u>WEB</u>	<u>TV/M</u>	<u>SPOK</u>	<u>FIC</u>	<u>MAG</u>	<u>NEWS</u>	<u>ACAD</u>
FREQ	15469	1418	1673	1001	1617	3969	1972	2977	842
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8
PER MIL	15.58	11.03	13.46	7.82	12.82	33.54	15.64	24.45	7.03

As shown in Table 4, the word *'passenger'* most frequently appears in fiction genre, which 3,969 times. Whereas, the least frequent word of *'passenger'* appears in TVs/movies which is 1,001 times. The variation in the frequency of the word *'passenger'* across different genres suggests that the term may be more relevant and central to the narratives and themes explored in fiction. Fictional stories often involve characters and plotlines that revolve around journeys, transportation, or interpersonal dynamics in confined spaces, which may explain the higher frequency of *'passenger'* in this genre. The lower frequency of *'passenger'* in the TV/movies genre could be attributed to the visual nature of this medium. In TV shows or movies, the focus is primarily on visual storytelling, and

the term *'passenger'* may be less frequently used or less central to the plotlines or dialogue compared to the visual elements, action sequences, or cinematography.

Table 5. Frequency of word *'boarding'*

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
FREQ	6532	631	747	1161	544	1182	735	925	607
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8
PER MIL	6.58	4.91	6.01	9.07	4.31	9.99	5.83	7.60	5.07

Same with the word *'passenger'*, the word *'boarding'* appears mostly in fiction genre. It is shown in Table 5. It appears 1,182 times of totally 6,532 (18.1%) found in fiction. Meanwhile, it appears least frequently in academic genre, which is only 607 times. The higher frequency of the word *'boarding'* in the fiction genre suggests that this term is often employed to describe and establish settings related to transportation, such as airports, train stations, or ships, which are commonly featured in fictional narratives. Fictional stories often incorporate detailed descriptions of characters boarding vehicles or vessels, emphasizing the process and creating a sense of movement and transition. Whereas, academic texts typically prioritize a formal and precise language register, which may result in a more restricted use of terms like *'boarding'*. Additionally, the academic genre may not extensively cover topics related to transportation or travel, further contributing to the lower frequency of *'boarding'* within academic texts.

Using the words arisen in high frequency shows the higher accuracy. High frequency can reflect the accuracy of an aspect of language by native speakers whose patterns can be imitated by foreign speakers (Meyer, 2002). Having information on the number of times that a word is observed in a representative corpus is essential, whereas assignment errors lead to easily recognizable noise for high-frequency words, they may result in misclassifications for low-frequency words (Brysbart, et al, 2012).

2. Topics

Topics show what words co-occur on a given web page, and they are a great way to find related words and concepts. The researcher only took 30 largest topics related to airlines. The total topics found is 22,798 topics. Based on COCA the topics related to the words of *'flight'*, *'passenger'*, and *'boarding'* can be seen in Table 6.

Table 6. Topics related to *'flight'*, *'passenger'*, and *'boarding'*

flight		passenger		boarding	
1073	<i>airline</i>	733	<i>flight</i>	87	<i>board</i>
970	<i>plane</i>	691	<i>airline</i>	36	<i>passenger</i>
860	<i>passenger</i>	586	<i>airport</i>	31	<i>airport</i>
845	<i>airport</i>	521	<i>plane</i>	31	<i>flight</i>
767	<i>fly</i>	509	<i>board</i>	29	<i>airline</i>
743	<i>pilot</i>	406	<i>fare</i>	25	<i>Indian (j)</i>
657	<i>aircraft</i>	353	<i>transportation</i>	24	<i>luggage</i>
574	<i>flying</i>	348	<i>travel</i>	23	<i>plane</i>
460	<i>aviation</i>	328	<i>traveler</i>	22	<i>pass</i>
452	<i>travel</i>	306	<i>cargo</i>	22	<i>reservation</i>

451	<i>airplane</i>	305	<i>aircraft</i>	22	<i>taxi</i>
450	<i>jet</i>	298	<i>pilot</i>	20	<i>passport</i>
425	<i>board</i>	297	<i>train</i>	19	<i>fare</i>
396	<i>crew</i>	284	<i>driver</i>	19	<i>native</i>
345	<i>carrier</i>	281	<i>seat</i>	19	<i>policeman</i>
345	<i>fare</i>	280	<i>aviation</i>	18	<i>Indian (n)</i>
327	<i>attendant</i>	268	<i>fly</i>	18	<i>traveler</i>
318	<i>traveler</i>	266	<i>ship</i>	18	<i>tribe</i>
316	<i>book</i>	255	<i>crash</i>	17	<i>crowded</i>
297	<i>space</i>	251	<i>depart</i>	17	<i>exit</i>
291	<i>crash</i>	248	<i>vehicle</i>	17	<i>gate</i>
288	<i>cargo</i>	246	<i>engine</i>	17	<i>roommate</i>
286	<i>engine</i>	241	<i>flying</i>	17	<i>suitcase</i>
266	<i>air</i>	240	<i>cabin</i>	17	<i>tuition</i>
266	<i>land</i>	237	<i>crew</i>	16	<i>blond</i>
259	<i>landing</i>	236	<i>airplane</i>	16	<i>financially</i>
258	<i>mission</i>	236	<i>carrier</i>	16	<i>open</i>
258	<i>united</i>	232	<i>route</i>	16	<i>polite</i>
252	<i>depart</i>	229	<i>jet</i>	16	<i>terrorist</i>

The result shows that between the three words ‘*flight*’, ‘*passenger*’, and ‘*boarding*’, most of topics found are same. Only in collocation of ‘*boarding*’ there are found some different topics which do not have close relation with airlines or aviation, such as ‘*Indian*’, ‘*tribe*’, ‘*blond*’, even ‘*terrorist*’. The similarities in topics associated with ‘*flight*’ and ‘*passenger*’ suggest that these words are closely related to the domain of airlines and aviation. It is expected that discussions, texts, or collocations with these terms would primarily revolve around aspects such as travel, air transportation, passenger experiences, airline operations, or related topics within the aviation industry. In teaching EFL, teachers can introduce COCA to students, and select some key words for the writing topic and then search lexical items closely related with the key words, which can greatly inspire students’ ideas and provide adequate lexical resources for them (Xiao & Chen, 2018).

3. Cluster

Word cluster is a group of words that are related to each other in some way. These related words may be grouped together because they have similar meanings. The clusters found in COCA are the words ‘*flight*’, ‘*passenger*’ and ‘*boarding*’, which each word consists of 2 words, 3 words, or 4 words of clusters. The clusters found are only taken the 30 highest top to be analyzed.

Table 7. Clusters of ‘*flight*’

Freq	2 Clusters	Freq	3 Clusters	Freq	4 Clusters
2669	<i>flight to</i>	907	<i>on the flight</i>	120	<i>on the next flight</i>
1965	<i>flight from</i>	812	<i>flight of stairs</i>	108	<i>how was your flight</i>
1858	<i>flight attendant</i>	733	<i>on a flight</i>	77	<i>on the same flight</i>
1476	<i>flight attendants</i>	513	<i>flights of stairs</i>	72	<i>flights of stairs to</i>
1445	<i>in flight</i>	283	<i>American airlines flight</i>	72	<i>have a nice flight</i>
1017	<i>space flight</i>	278	<i>for the flight</i>	72	<i>flight to and from</i>
842	<i>flights to</i>	270	<i>down a flight</i>	66	<i>on the first flight</i>
766	<i>airlines flight</i>	263	<i>flight back to</i>	57	<i>have a safe flight</i>
749	<i>first flight</i>	256	<i>for a flight</i>	56	<i>flight of stairs to</i>
648	<i>flight in</i>	226	<i>to the flight</i>	54	<i>have a good flight</i>

In Table 7, the word ‘*flight*’ is grouped with one, two or three other word(s).

It is combined with an article (ART), for example ‘*a flight*’, preposition (PREP) for example ‘*flight to*’, noun (N) for example ‘*airlines flight*’, adjective (ADJ) as in ‘*safe flight*’, or with verb (V) as in ‘*have flight*’. The position of the word ‘*flight*’ can be in the beginning, for example as in ‘*flight attendant*’, or at the end as in ‘*on the next flight*’.

The combinations of ‘*flight*’ with different word categories reveal the syntactic structures in which ‘*flight*’ is commonly used. The presence of articles, such as ‘*a flight*’, suggests that ‘*flight*’ is often treated as a countable noun. The combinations with prepositions, such as ‘*flight to*’ indicate the relationship between ‘*flight*’ and a specific destination or purpose. Combinations with nouns, adjectives, and verbs illustrate the diverse ways in which ‘*flight*’ can be modified or used within different syntactic contexts. It also reveals its flexibility in terms of word order. The word ‘*flight*’ can occur at the beginning of the phrase or at the end. This positional variation demonstrates the ability of ‘*flight*’ to function as both a head noun or a modifier within a noun phrase.

Table 8. Clusters of ‘passengers’

Freq	2 Clusters	Freq	3 Clusters	Freq	4 Clusters
1848	<i>passenger seat</i>	1066	<i>in the passenger</i>	99	<i>one of the passengers</i>
789	<i>other passengers</i>	316	<i>on the passenger</i>	97	<i>in the front passenger</i>
780	<i>passenger side</i>	374	<i>passenger and crew</i>	59	<i>sat in the passenger</i>
769	<i>passenger on</i>	316	<i>into the passenger</i>	58	<i>sitting in the passenger</i>
590	<i>passenger in</i>	208	<i>from the passenger</i>	57	<i>out of the passenger</i>
557	<i>passenger door</i>	155	<i>number of passengers</i>	51	<i>passengers and crew members</i>
550	<i>passengers who</i>	129	<i>as a passenger</i>	46	<i>passenger in a car</i>
418	<i>passenger in</i>	125	<i>to the passengers</i>	42	<i>passengers in the car</i>
392	<i>passenger cars</i>	98	<i>passenger on board</i>	29	<i>passengers on the plane</i>
362	<i>for passengers</i>	97	<i>opened the passenger</i>	23	<i>passenger bill of rights</i>

The word ‘*passenger*’ is clustered in two, three, four or more words. It is combined with Noun (N) as ‘*passenger seat*’, preposition + article (PREP + ART) as in ‘*to the passenger*’, and verb (V) as in ‘*sat in the passenger*’. The position of the word ‘*passenger*’ can be in the beginning, in the middle or at the end of the phrase. The most frequent cluster appears 1,848 times as ‘*passenger seat*’.

The clustering of ‘*passenger*’ with Nouns (N) indicates that ‘*passenger*’ often functions as a modifier within noun phrases. These combinations provide specific information about the role or location of the passenger, highlighting the relationship between ‘*passenger*’ and other elements in the phrase. The prepositional combinations indicate the direction, destination, or recipient associated with the passenger. The variations in the position of ‘*passenger*’ within the collocations (beginning, middle, or end) illustrate its syntactic flexibility and its ability to function in different positions within the phrase. This flexibility allows for different emphases or syntactic structures, depending on the specific context or intended meaning.

Table 9. Clusters of ‘boarding’

Freq	2 Clusters	Freq	3 Clusters	Freq	4 Clusters
1796	<i>boarding school</i>	270	<i>boarding school in</i>	32	<i>boarding call for flight</i>
466	<i>boarding house</i>	90	<i>boarding school for</i>	17	<i>I went to boarding</i>
400	<i>boarding schools</i>	83	<i>to a boarding</i>	17	<i>sent to a boarding</i>
346	<i>to boarding</i>	73	<i>in a boarding</i>	14	<i>back to the boarding</i>

246	<i>boarding pass</i>	65	<i>at the boarding</i>	12	<i>sent away to boarding</i>
136	<i>at boarding</i>	59	<i>boarding call for</i>	12	<i>went to a boarding</i>
117	<i>final boarding</i>	59	<i>at a boarding</i>	11	<i>boarding schools for girls</i>
104	<i>boarding call</i>	58	<i>to the boarding</i>	9	<i>him off to boarding</i>
95	<i>in boarding</i>	57	<i>off to boarding</i>	8	<i>it was a boarding</i>
77	<i>boarding houses</i>	40	<i>in the boarding</i>	7	<i>kicked out of boarding</i>

The word ‘boarding’ can be combined with Noun (N) as in ‘boarding house’, preposition (PREP) as in ‘at boarding’, Verb (V) as in ‘sent to a boarding’, and an article (ART) as in ‘a boarding’. The most frequent cluster appears 1,796 words for example ‘boarding school’.

The collocations of ‘boarding’ with Nouns highlight its usage as a modifier within noun phrases. These combinations provide specific information about the type or purpose of the boarding arrangement, such as a house where boarding is available or a school that offers boarding facilities. The collocations of ‘boarding’ with Verbs reflect actions or processes related to boarding. These combinations indicate the act of sending someone to a boarding arrangement, such as being sent to a boarding school or institution. The collocations of ‘boarding’ with an article, as in ‘a boarding’ signify a more general or indefinite reference to the concept of boarding. These combinations indicate a non-specific boarding situation, without providing further details or specifying a particular type of boarding arrangement. From the findings above, it can be discoverable that the clusters of three words ‘flight’, ‘passenger’ and ‘boarding’ by COCA totally is 34,006. The word ‘flight’ has the largest clusters (18,730). The cluster found can be used as idiom in teaching vocabulary especially for airlines or aviation students.

CONCLUSION

Using corpora is still in its development in foreign language classes despite its great advantages and potential to provide solutions to the varied challenges in foreign language instruction for teachers and learners (Kayaoğlu, 2013). Therefore, this study aims to analyze the collocations used in airlines by using Corpus of Contemporary American English (COCA). The analysis is focused on frequency, topics, and clusters. The results of the study show that from the three words taken from the script of flight announcement in airport ‘flight’, ‘passenger’, and ‘boarding’, there are various aspects, namely: word frequency, collocation patterns, genre and contextual variations, topic associations, and syntactic flexibility. The research provides valuable insights into the frequency, collocation patterns, and contextual associations of the studied words within the examined corpus. These findings contribute to a deeper understanding of language usage, semantic connections, and syntactic patterns in relation to the domains of flights, passengers, and boarding.

The conclusions drawn from the research findings can inform further investigations, language studies, and applications within the relevant fields. Further research could investigate whether extended exposure to the corpus, over a longer time frame, can facilitate learners' inductive learning and enhance their language skills. By examining the effects of prolonged corpus use on language acquisition, researchers can gain insights into the potential benefits and limitations of this approach. Furthermore, it would be interesting to compare the outcomes of learners

who engage in extended corpus-based learning with those who follow traditional language learning methods. By comparing the effectiveness of these approaches, researchers can provide evidence-based recommendations for incorporating corpus-based techniques into language curricula.

REFERENCES

- Ahn, S. (2014). The effectiveness of using corpus in developing learners' awareness of target lexis, grammatical structures, and collocations. *Journal of the Korea English Education Society*, 13(4), 1-20.
- Anthony, Laurence. (2004). AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit. *IWLeL: An Interactive Workshop on Language e-Learning*. 7-13.
- Anthony, Laurence. (2005). AntConc: Design and Development of a Freeware Corpus Analysis. Toolkit for the Technical Writing Classroom. *IEEE International Professional Communication Conference Proceedings*.
- Bardovi-Harlig, K., Mossman, S., & Vellenga, H. E. (2015). The effect of instruction on pragmatic routines in academic discussion. *Language teaching research*, 19(3), 324-350.
- Brysbaert, M., New, B., and Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behav Res (2012)* 44:991–997.
- Davies, M. (2009). The 385-million-word corpus of contemporary American English (1990-2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, 14, 159-190.
- Fan, May. (2009). An Exploratory study of collocational use by ESL students—A task-based approach. *System*, 37, 110-123.
- Farr, Fiona. (2008). Evaluating the Use of Corpus-based Instruction in a Language Teacher Education Context: Perspectives from the Users. *Language Awareness Vol. 17, No. 1*.
- Frankenberg-Garcia. (2014). The use of corpus examples for language comprehension and production. *ReCALL 26(2): 128–146*. 2014 © *European Association for Computer Assisted Language Learning* 128.
- Gabrielatos, C. (1994). Collocations: Pedagogical implications and their treatment in pedagogical materials. Unpublished essay. *Research Centre for English and Applied Linguistics. University of Cambridge*.
- Hussein, R., & Sawalha, M. (2016). A corpus-based study of similes in British and American English. *Arab World English Journal (AWEJ) Volume*, 7.
- Karimi, P. and Sanavi, R. V. (2014). Analyzing English Language Learning Needs among Students in Aviation Training Program. *Procedia - Social and Behavioral Sciences* 98 (2014) 852 – 858.
- Kayaoğlu, M. N. (2013). The use of corpus for close synonyms. *The Journal of Language and Linguistic Studies*, Vol. 9, No. 1.
- Kennedy, Graeme. (2003). Amplifier Collocations in the British National Corpus:

- Implications for English Language Teaching. *TESOL Quarterly* 37 (3), 467-477.
- Mansour, D. M., (2017). Using COCA to foster students' use of English collocations in academic writing. *The 3rd International Conference on Higher Education Advances*. DOI: <http://dx.doi.org/10.4995/HEAd17.2017.5301>
- McEnery, Tony et.al. (2006). *Corpus-Based Language Studies: An Advanced Resource Book*. London: Routledge.
- Mcgee, Iain. (2009). Adjective-noun collocations in elicited and corpus data: Similarities, differences, and the whys and wherefores. *Corpus Linguistics and Linguistic Theory* 5-1,79-103 1613-7027/09/0005-79 DOI 10.1515/CLLT.2009.00.
- Meyer, Charles F. (2002). *English Corpus Linguistics: An Introduction*. Cambridge: CUP.
- Romer, U. (2008) *Corpora and language teaching*. In A. Ludeling & M. Kyoto (Eds.), *Corpus linguistics: An international handbook* (Vol. 1, pp. 112-130). Berlin: Mouton de Gruyter.
- Romer, Ute. (2011). Corpus Research Applications in Second Language Teaching. *Annual Review of Applied Linguistics*, 31, 205-225.
- Xiao, G. and Chen, X. (2018) Application of COCA in EFL writing instruction at the tertiary level in China. *International journal of Emerging Technologies in Learning*, Vol. 13 No.09.
- Yusu, X. (2014). On the application of corpus of contemporary American English in vocabulary instruction. *International Education Studies*; Vol. 7, No. 8.