

A corpus-based analysis of English synonyms : acquire and obtain

¹Sita Sittironnarit, ²Kongkiat Khunasathitchai,
³Krishna Kosashunhanan and ⁴Soranut Kumdee

^{1,2,3}Faculty of Liberal Arts, Rajamangala University of Technology Thanyaburi

⁴Faculty of Science and Health Technology, Navamindradhiraj University

Corresponding Author, E-mail: kongkiat_k@rmutt.ac.th

Abstracts

Due to their similar meaning, L2 English language learners may struggle with the use of near synonyms. Corpus-based studies on English synonyms have been conducted over the last two decades to solve this issue. To the best of the researchers' knowledge, the distinction between the synonymous verbs *acquire* and *obtain* has not yet been examined. This study, thus, examines these two verbs in terms of genres and collocations using the Corpus of Contemporary American English (COCA). The findings indicate that they appear to have a high degree of formality and are prevalent in written rather than spoken genres. In general, *obtain* is more common than *acquire*. In terms of collocation, they may be distinguished by their specific noun collocates. Even though they share some noun collocates, they cannot be used interchangeably in all situations. Semantic preference and semantic prosody and pedagogical implications are also discussed.

Keywords: Collocation; Corpus Linguistics; A Corpus-Based Analysis; English Synonyms,

Introduction

English Synonyms are considered one of the challenges for language acquisition due to the difficulty in distinguishing their semantic meaning (Phoocharoensil & Kanokpermpoon, 2021). Each synonym cannot be used interchangeably due to its meaning (Edmonds & Hirst, 2002) as well as the context of the language (Liu, 2010). According to Thornbury (2002), L2 learners may write unnatural English sentences by unconsciously substituting a word with its synonyms, resulting in the unnatural English use. For example, the two synonyms, *fast* and *quick* can collocate with the noun *action* e.g. “His *quick action* in the gym saved her lives.” and “Doctors said Joey's training and *fast action* saved his mother's life (Aroonmanakun, 2015). However, *fast* cannot be replaced by *quick* in “this *fast* computer is worth buying”. Undoubtedly, knowledge of how to use English synonyms commonly and naturally is of importance for all English skills especially for writing ability (Yeh et al., 2007).

Boonraksa & Naisen (2022) studied English collocational errors of 285 Thai EFL undergraduate learners by classifying them into three groups of students based on their English proficiency, low, moderate, and high. It was found that all three groups have errors in two types of collocations, namely lexical and grammatical, which are caused by a negative transfer from their L1 or Thai, synonyms, and a lack of collocation understanding, resulting in incorrect collocation usage in their writing. It is due to the fact that in a Thai context in which English is regarded as a foreign language (EFL), exposure to authentic English is limited, resulting, collocational errors. It is no doubt that English language teachers need to equip Thai EFL students with collocational knowledge.

* Received: September 1, 2022; Revised: September 8, 2022; Accepted: September 9, 2022

In terms of English language teaching, vocabulary instruction mostly focuses on the use of a single word rather than multiple word phraseological units or its neighboring words since it is more convenient for teachers in terms of material preparation (Schmitt, 2010). These days have witnessed the call for deeper vocabulary teaching in terms of how to use a word naturally (Phoocharoensil, 2021). A corpus-based analysis of synonyms makes it possible to reveal both the similarities and differences of a pair or set of synonyms because of its size of words which is over millions or billions of authentic words (e.g., Phoocharoensil, 2010; 2020; 2021; Selmistraitis, 2021).

However, the synonymous words, *acquire* and *obtain* are still unexplored. There are four reasons why it is worth studying and why they were purposely selected. First, they are both indexed in Coxhead's (2000) academic word list that was examined based on the range and frequency of the corpus of authentic 3.5 million running words, and Gardner & Davies' Academic Vocabulary List or AVL (2014) that was developed based on a compilation of a corpus of 120-million words, meaning that these two words play a vital role in an academic genre. Secondly, according to Cambridge Online Dictionary, *acquire* means 1) to get or obtain something, to buy a business or property, or buy shares in a business, 2) to gradually learn or gain something such as a language or skill, and 3) to develop a medical condition or catch a disease that you were not born while *obtain* means "to get something, especially by asking for it, buying it, working for it, or producing it from something else". They share same core meaning "to get something". Even though a dictionary provides some explanations of their meanings accompanied by examples, it lacks information on how to use these two words naturally. In other words, a limited number of noun collocates of both is given, resulting in a difficulty for learners to select an appropriate word. Thirdly, they were listed in the common 2,000 words in spoken English (Pearson Education, 2014), indicating that they are common words in which learners need to understand how to use them frequently. Fourthly, in terms of CEFR, they are both indexed at the B2 level, meaning their level of formality is not different based on the dictionary.

Research Objectives

1. To find out the distribution across genres in COCA of English verbs, *acquire* and *obtain* as well as their level of formality
2. To find out the similarities and differences between the synonymous verb, *acquire* and *obtain*, particularly, their collocation, semantic preference, and semantic prosody.

Research Methodology

COCA or the Corpus of Contemporary American English was purposely employed as a main source of data. Not only is COCA a large corpus composed of over 1 billion words, but it makes it possible to explore the distribution of words through eight different genres i.e., spoken English, fiction, magazines, newspapers, academic texts, TV and movies subtitles, blogs, and websites. COCA is regarded as one of the most frequently used corpora of the English language because of its big size.

To answer the first research question about the distribution, the target verb i.e., *acquire* and *obtain* were investigated through the occurrence across different genres in COCA. For the second research question, nouns frequently collocating with the target synonyms were searched and analyzed through COCA. To ensure the reliability of the association of the collocational

strength measurement, the MI or Mutual Information score is a statistical score used for identifying whether a word collocates by chance or not. Following Phoocharoensil (2020), the researchers considered the Mutual Information score which is a statistical score designed for the reliability of the collocational strength using a logarithmic scale to represent the ratio of the frequency of the collocation to the frequency of the two words occurring together whether they are collocated by chance or not (Kruawong & Phoocharoensil (2021). However, it has been critiqued that MI may be inadequate as Schmitt (2010) points out that collocations with very high MI scores could have relatively low overall corpus frequency. Note that in the current study, the frequency of noun collocates along with the MI value was employed to determine the strength of collocation to avoid including noun collocates that might be uncommon although they have a high MI score (Cheng, 2012; Gablasova, Brezina & McEnery, 2017). Therefore, we included only the top-40 frequency noun collocates of the two verbs with the frequency and MI score of ≥ 3 , which can indicate the significant value for the strength of collocational association. To report the top-40 frequency noun collocates, redundant words were removed in terms of their lemmatization. Since the automatic process may result in the redundancy of the collates e.g., *samples* and *sample*. Then, they were grouped into categories based on semantic preference. To analyze semantic prosody, the researchers manually consider top-40 frequency concordance lines of both verbs on COCA, which is categorized into three primary attitudinal tones, i.e., positive, negative, or neutral according to Louw (1993).

To ensure intercoder reliability of semantic preference and semantic prosody, the data was coded by one of the researchers holding a Ph.D. in Education who has been teaching English for over two decades until the researchers reach an agreement at 100%.

Results

To answer research question No.1, Table No.1 reveals that not only is the total frequency of the verb *obtain* (21,173 tokens) greater than *acquire* (8,234 tokens) but it also occurs with more frequency in seven genres. The only genre that *obtain* occurs less frequently than *acquire* is fiction with a slight difference (555 and 490 tokens). Considering the degree of formality, it was found that both words occur the most in academic texts, meaning that they are used formally. To consider the top-5 genre where both words occur, they have the same orders ranked by the frequency namely, academic texts, websites, blogs, magazines, and newspapers, meaning not only do they seem to share the same level of the degree of formality, but they also distribute across different genres at the same order. The corpus-informed data, moreover, also indicates that both words seem to be less common in spoken genres, i.e., TV/Movie, spoken and fiction as these three genres are the least 3 genres that these two words occur. Our findings also are in line with Coxhead's (2000) academic word and Gardner & Davies' Academic Vocabulary List or AVL (2014) in which these two words are indexed in written academic genre due to their high degree of formality and they occur less in spoken discourse.

Table 1. The Distribution of *acquire* and *obtain* across eight genres on COCA

Genre	Acquire		Obtain	
	Frequency	Per million	Frequency	Per million
blog	1,549	12.04	2,909	22.62
web-genl	1,991	16.02	4,561	36.71
tv/movies	386	3.01	415	3.24
spoken	528	4.19	736	5.83
fiction	555	4.69	490	4.14
magazine	1,640	13.01	2,400	19.03
newspaper	1,585	13.02	2,164	17.78
academic	3,583	12.04	7,498	62.59
Total	8,234		21,173	

Table 2. Noun Collocates of *Acquire* and *Obtain*

No.	Acquire			Obtain		
	Noun collocates	Frequency	MI score	Noun collocates	Frequency	Mi score
1	skills	944	6.32	information	1801	4.66
2	knowledge	509	5.82	data	816	3.94
3	weapons	412	5.71	results	459	3.70
4	information	354	3.17	permission	452	6.45
5	land	277	4.12	consent	408	6.65
6	property	258	4.57	license	394	5.89
7	rights	214	3.53	degree	377	4.50
8	companies	202	3.32	copy	308	5.08
9	technology	180	3.39	warrant	298	6.62
10	trade	177	3.62	approval	291	5.54
11	language	153	3.34	records	290	4.19
12	status	151	3.81	sample	275	4.62
13	wealth	146	4.86	insurance	236	3.56
14	reputation	134	5.16	access	211	3.06

15	assets	131	5.03	documents	208	4.50
16	taste	113	4.23	credit	200	3.27
17	resources	112	3.07	search	192	3.24
18	understanding	94	3.02	employment	190	4.67
19	acres	94	5.04	scores	190	4.10
20	infection	91	4.87	funding	183	4.12
21	meaning	91	3.13	sources	176	3.59
22	equipment	89	3.59	weapons	173	3.52
23	stake	88	4.85	freedom	167	3.17
24	capability	85	5.66	participants	162	3.39
25	customers	82	3.23	coverage	160	3.87
26	citizenship	81	5.28	VISA	150	6.47
27	goods	79	3.82	benefits	148	3.04
28	expertise	75	4.74	abortion	126	3.86
29	materials	67	3.06	driver	121	3.39
30	significance	66	4.31	estimates	120	4.16
31	shares	64	3.90	loan	117	4.14
32	immunity	63	5.48	relief	110	3.64
33	habit	62	4.52	materials	109	3.19
34	talent	59	3.46	citizenship	101	5.07
35	capacity	58	3.20	certificate	99	5.49
36	ownership	57	4.31	assistance	98	3.54
37	license	56	3.58	financing	92	5.08
38	territory	52	3.70	chronicle	87	4.82
39	objects	51	3.14	interviews	77	3.50
40	vocabulary	50	4.64	authorization	77	6.20

Table 2 indicates the top-20 noun collocates extracted from COCA that frequently co-occur with the verb *acquire* and *occur* which were ranked by the frequency with a statical score of mutual information to ensure the collocational strength which is equal to or greater than 3 (Cheng, 2011). Drawing upon the top-40 noun collocates of the two verbs, it is obvious that there are five nouns that the words *acquire* and *obtain* share the same noun collocates, namely, *weapons*, *information*, *citizenship*, *materials*, and *license*. All five words seem to mean something that is of importance to *acquire* or *obtain* such as information or something that plays an essential role in which he/she needs such as citizenship. There are concrete nouns i.e., *materials*, *weapons*, *licenses*, and abstract nouns i.e., *information* and *citizenship*.

1. and 2 are extracted from COCA indicating how both words collocate with the noun, material.

(1) Kills are closely linked to the ability to acquire material, and hence also social, benefit.

(2) However, e-mail and the Internet should not be used to obtain material that is offensive.

The concordance lines (3) and (4) examples the collocations of *acquire* + weapon and *obtain* + weapon.

(3) Al Qaeda leaders have sought contacts in Iraq who could help them acquire weapon.

(4) Once in power, they would develop or obtain weapons of mass destruction

Here are two examples of pairs collocates of *acquire* + *license* and *obtain* + *license*.

(5) Texans won the right to obtain licenses to carry concealed weapons.

(6) More school psychologists will acquire licenses for private practice to expand their roles in schools.

Below are two extracted concordance lines of *acquire* information and *obtain* information

(7) If investors actively try to acquire information about the contents of forthcoming public reports.

(8) ...if newspapers or television receive or obtain information fairly from a reliable and responsible source...

The last noun collocate that both words share is citizenship.

(9) Apart from marriage to a Saudi national, there are no provisions for long-term foreign residents to acquire citizenship.

(10) However, they may obtain citizenship for their non-Jordanian husbands who may then confer citizenship on the children.

The next step is to categorize all the noun collocates based on semantic meaning. To scrutinize the data, it is suggested to categorize collocates into semantic preference sets based on the near meaning (Sinclair & Carter, 2004). Then the sets of semantic preferences of both words' noun collocates were categorized.

Table 3. Semantic Preference about Data

Word	Noun Collocates
acquire	information resources
obtain	information data sources chronicle results record score

Regarding Table 3, *acquire* and *obtain* share are similar as they share the noun collocate information. Moreover, while the word *acquire* can collocate with *information* and *resources*, *obtain* can be followed by seven nouns including *information*, *data*, *sources*, *chronicle*, *results*, *record*, and *score*. Judging by the semantic perspective, *acquire* seem to co-occur with the meaning of information or resource that can be used to supply something while the *obtain* is likely to collocate with the nouns that deals with a wider range of words as mentioned earlier. The concordance lines No. (11) and (12) are examples of them.

- (11) For instance, inspectors' powers to acquire information about accidents needed to be extended to include inquiries made to determine whether an investigation was appropriate, rather than being confined to those cases.
- (12) Another study of employers' use of temporary workers using US survey data (Magnum/Mayall/Nelson, 1985) obtained results very similar to our own with respect to fixed-term contract workers.

Table 4. Semantic Preference about Things or objects

Word	Noun Collocates
acquire	equipment goods materials objects weapon
obtain	copy documents materials sample weapon

Table No.4 illustrates the noun collocates of *acquire* and *obtain* in which the meaning is related to a thing or object. It is worth noticing that *materials* can be followed by both words. Particularly, *acquire* seem to collocate with materials relating to teaching, learning, or library-involved activity as in (13) and (14) while *obtain* + noun seems to occur variously as in (15) and (16).

(13) As the Germanic Librarian at Cal, I can add that our ability to acquire materials from the areas I represent has declined dramatically in the past five years.

(14) P.L.480 was a law enacted by Congress in 1958 that allowed LC to acquire materials on the behalf of other libraries and research centers in the United States.

(15) If the company you have hired cannot obtain materials from a builder's merchant without any advance payment there is something wrong.

(16) We can help you find information and obtain materials from elsewhere whether books, journal articles, or research reports.

The word *weapon* is another word that *acquire* and *obtain* share the same noun collocate, indicating the weapon of the government of army as in (3) and (4). In addition, *equipment*, *goods*, and *materials* are other noun meaning a thing or object in which people can make use of that can neighbor with *acquire* while *obtain* co-occurs with words relating to a piece of paper or record or sample such as *copy documents* and *sample*.

Table 5. Semantic Preference about Right or Allowance

Words	Noun Collocate
acquire	citizenship license ownership rights status
obtain	access approval authorization citizenship certificate consent credit degree license permission VISA

Concerning semantic preference of right or allowance as can be seen in Table 5, there are two nouns *acquire* and *obtain* share the same collocates including *citizenship* and *license* as in (9) and (10) and (5) and (6) respectively. Regarding *citizenship*, they seem to have no distinction about the use between two words as they are used in a context relating to the status to live in a foreign area. Similarly, for the word *license*, both synonymous verbs tend to occur in the context of status or permission to do something. To distinguish these two words, it can be noted that *acquire* collocates with the words allowing a person to do something such as *citizenship*, *license*, *ownership*, *rights*, and *status* whereas the word *obtain* seems to have a wider range of collocation indicating do something such as *approval*, *authorization*, *citizenship*, *credit*, *degree* *permission* as well as a document allowing him/her a permission to do something such as *certificate*, *consent*, and *VISA*.

Discussion

Table 6. Other categorizations of semantic preference of *acquire*

Semantic Preference	Noun Collocates
knowledge	knowledge language understanding meaning vocabulary
land	land property assets acres territory
ability	skills status capability expertise habit talent capacity
business	companies trade stake customers shares
richness	wealth resource
negative meaning	infection
others	reputation taste technology significance immunity

According to Table No. 6, the word *acquire* is composed of 7 groups of semantic preference in which the word *obtain* is not. Considering the meaning of knowledge, it is found that *acquire* co-occurs with the words meaning to gradually learn or gain something especially knowledge or language including *knowledge*, *language*, *meaning*, *understanding*, and *vocabulary*. In addition, *acquire* collocates with the words relating to land or area such as *land*, *property*, *assets*, *acres*, and *territory*. Nouns relating to business such as companies, trade, stake, customers, and *shares* are collocates of *acquire*. Moreover, *acquire* co-occurs with a wide range of semantic preference sets such as richness and others including *reputation*, *taste*, *technology*, *significance*, and *immunity*. It can be noted that the noun collocates of *acquire* are likely to indicate, neutral or positive meaning resulting in neutral or positive semantic prosody. However, it is noticed that only *acquire* tends to collocate with nouns indicating negative meaning or negative semantic prosody, namely *infection* as can be seen in (17) and (18).

(17) For the same reasons that the female is more likely to acquire infection in the throat than her male sexual partner.

(18) Horses are thought to acquire infection mainly from pastures contaminated by donkeys during the summer months.

Table 7. Other categorizations of semantic preference of *obtain*

Semantic Preference	Noun Collocates
money	funding loan financing
to guarantee	warrant insurance
etc.	employment search estimates freedom coverage benefits abortion relief degree assistance participants interviews DNA

Table No.7 indicates that *obtain* tends to collocate with the words relating to money such as *funding, financing, and loan*, as well as the words meaning to guarantee such as *warrant and insurance* which contrast with those of *acquire*. In addition, there is a set of Etc. semantic preference that *obtain* can neighbor with including a wide range of words i.e., *employment search estimates freedom coverage benefits abortion relief degree assistance participants interview and DNA*.

Conclusion

This current study examined the synonymous verbs, *acquire*, and *obtain* regarding their degree of formality through their distribution across genres as well as their similarities and differences particularly, their collocation, semantic preference and semantic prosody, resulting in the different collocational behaviors. The results of this study indicate that not only do the verbs, *acquire* and *obtain* tend to not only occur heavily in the written academic genre, but also appear in other written genres such as blogs, websites, magazines, and newspapers. In addition, they both rarely appear in spoken genres, as a result, they seem to have a high degree of formality. For research question No.2, these two words share the same core meaning according to an online dictionary and COCA. Another interesting point is that corpus-informed evidence sheds some light on the similarities and differences. In general, *obtain* is likely to be more common than *acquire* based on the tokens found on COCA. In terms of semantic preference, they share some similar sets of words meaning data or information, things or objects, and right or allowance to do something. However, there are only five words that these words shared the same noun collocates i.e., *weapons, information, citizenship, materials, and license*. Thus, English language users especially should carefully consider the noun collocate to best suit the context. For example, *source* is a noun collocate of *obtain* but it seems not to collocate with *acquire*. Despite the shared sets of semantic preferences, it is obvious that these two words have their noun collocates which L2 should carefully consider when using in a context. Therefore, we can conclude that *acquire* and *obtain* are near or lose synonyms in which they cannot be interchangeably substituted in every context. Considering their semantic preference, another worth mentioning point is that *acquire* is likely to co-occur with something that a person gradually gets while the word *obtain* is found to be used with words relating to a paper or authorization to do something or allowance to get or do something. Moreover, while *acquire*

collocates with the business words relating to ownership, *obtain* are found with words relating to money. Apart from semantic preference, both words are likely to behave with either neutral or positive meanings. However, regarding their semantic prosody, only one word distinguishing between them is that *acquire* collocates with a noun, *infection* which tends to lead to negative cognitive meaning.

To sum up, this current study sheds some light on the degree of formality as well as the distinction between the academic synonymous verbs, *acquire* and *obtain*. To master the use of common and natural collocation, L2 learners cannot only make use of either their intuition or a dictionary but the corpus-based evidence. When it comes to writing, L2 writers can make use of COCA or another corpus along with consulting a dictionary to select the appropriate word to suit each context. For the further research, researchers may consider studying another word or set of synonyms of these words such as *get*, *earn*, and *receive* to yield some distinction, especially between spoken and written genres.

It should be noted that this current study is not without limitations. Since the main source of data is corpus-based on COCA, which is an American corpus, the data would reflect only the use of American speakers of English. Further researchers may focus on corpora of other native speakers of English such as British, Australian, New Zealand or ESL or EFL context corpus (Phoocharoensil & Kanokpermpoon, 2021) or compare between or among two or more corpora. In addition, there would be some noun collocates that do not appear on the results reported which may limit the results of this study since this study employed MI score. Further researchers may consider another statistical criterion such as t-score, z-score, or log-likelihood rather than MI score (Phoocharoensil & Kanokpermpoon, 2021).

References

- Aroonmanakun, V. (2015). Quick or fast: A corpus based study of English synonyms. *LEARN Journal: Language Education and Acquisition Research Network*, 8 (1), 53-62.
- Boonraksa, T., & Naisena, S. (2022). A Study on English Collocation Errors of Thai EFL Students. *English language teaching*, 15 (1), 164-177.
- Carter, R. (2012). *Vocabulary: Applied linguistic perspectives*. Routledge.
- Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. Routledge
- Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 34 (2), 213-238.
- Edmonds, P., & Hirst, G. (2002). Near-synonymy and lexical choice. *Computational linguistics*, 28 (2), 105-144.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67 (1), 155-179. <https://dx.doi.org/10.1111/lang.12225>
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied linguistics*, 35 (3), 305-327.
- Kruawong, T., & Phoocharoensil, S. (2022). A genre and collocational analysis of the near-synonyms teach, educate and instruct: A corpus-based approach. *TEFLIN Journal*, 33 (1), 75-97.
- Liu, D. (2010). Is it a chief, main, major, primary, or principal concern?: A corpus-based behavioral profile study of the near-synonyms. *International Journal of Corpus Linguistics*, 15 (1), 56-87.

- Pearson Education. (2014). *Longman dictionary of contemporary English* (6th ed.). Harlow, UK: Pearson Education Limited.
- Phoocharoensil, S. (2010). A corpus-based study of English synonyms. *International Journal of Arts and Sciences*, 3 (10), 227-245.
- Phoocharoensil, S. (2011). Collocational errors in EFL learners' interlanguage. *Journal of education and practice*, 2 (3), 103-120.
- Phoocharoensil, S. (2021). Semantic prosody and collocation: A corpus study of the near-synonyms persist and persevere. *Eurasian Journal of Applied Linguistics*, 7 (1), 240–258.
- Phoocharoensil, S. (2021). Multiword Units and Synonymy: Interface between Collocations, Colligations, and Semantic Prosody. *GEMA Online Journal of Language Studies*, 21 (2).
- Phoocharoensil, S., & Kanokpermpoon, M. (2021). Distinguishing the near-synonyms 'increase' and 'rise': Genre and collocation investigation. *Kasetsart Journal of Social Sciences*, 42 (4), 968-975.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Springer.
- Selmistraitis, L. (2020). Semantic preference, prosody and distribution of synonymous adjectives in COCA. *Journal of Language Studies*, 20 (3), 1-18.
- Sinclair, J., & Carter, R. (2004). *Trust the text: Language, corpus and discourse*. Routledge.
- Thornbury, S. (2002). *How to teach vocabulary*. Harlow.
- Yeh, Y., Liou, H. C., & Li, Y. H. (2007). Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning*, 20 (2), 131-152.