A Corpus-Based Investigation of English Synonyms: *Disadvantage, Downside,* and *Drawback*

Piyaboot Sumonsriworakun

kunkusiaclub@hotmail.com, Chulalongkorn University, Thailand

Abstract

The study compares three synonymous nouns, *disadvantage*, *downside*, and *drawback,* in terms of their frequency, distribution patterns, and collocations, using data from the Corpus of Contemporary American English (COCA). The findings show that the frequency of *disadvantage* is the highest, followed by that of *downside* and *drawback*, respectively. Regarding their distribution across eight registers in COCA, *disadvantage* prevails in academic texts, whereas *downside* and *drawback* seem to be less formal as they are most often found in magazines. An analysis of semantic preferences of the verb collocates of the three synonyms revealed two common themes: CONSIDER and DEAL WITH. As for their adjective collocates, the three synonyms frequently co-occur with adjectives under the theme EXTENT. *Disadvantage* is more often preceded by adjectives subscribed to the theme ASPECT than *drawback* is, and while *downside* regularly combines with some adjectives representing counter-expectations, *drawback* tends to be accompanied by more adjective collocates organized into the theme PROMINENCE than the other two synonyms. It is advisable that English language teachers utilize these valuable insights to develop lessons and materials.

Introduction

Synonymy is especially important in languages since it enables language users to select a particular word over another with analogous meanings to convey their intentional meaning and add variety to their writing or speaking. A lexical item frequently has synonyms which are at odds with it in terms of certain nuances of meaning (Edmonds & Hirst, 2002). Since there are fine-grained distinctions between a word and its synonyms, it is often demanding for a native speaker of a language to distinguish the meaning of synonyms and put them into use with constant precision, and it is even more challenging for second language (L2) learners to select a synonym to fit in a particular context (Dewaele, 2008; Lee & Liu, 2009; Mackay, 1980). Put simply, the learners may find it difficult to convey their intended meaning through when faced with a choice of words very close in meaning. It was found that L2 learners, even those at the advanced language proficiency level, are muddled about the use of synonymous words, as they cannot recognize a slight difference between them (Lee & Liu, 2009; Martin, 1984). It is, therefore, important for L2 learners to analyze words with which synonyms typically co-occur in order that they can recognize similarities and differences in their meanings, thereby being able to express their views and emotions precisely and appropriately for successful communication (Edmonds & Hirst, 2002).

The synonyms selected for examination in the present study are the nouns *disadvantage, downside,* and *drawback*. It is difficult for English language learners and teachers alike to distinguish these three nouns as they carry almost identical meanings. Even worse, L2 learners of English who are insufficiently exposed to the English language may possibly consider the terms as having exactly the same meanings, with the result that their L2 use will deviate from common, yet natural, usage of the language (Szudarski, 2018).

The three nouns under investigation are defined by three online American English dictionaries, namely Oxford Advanced American Dictionary, Merriam-Webster Dictionary, and The American Heritage Dictionary of the English Language, as displayed in Table 1:

Table 1

Definitions and examples of disadvantage, downside, and drawback in three online American English dictionaries

	Oxford Advanced American Dictionary	Merriam-Webster Dictionary	The American Heritage Dictionary of the English Language
1. disadvantage	something that causes problems and tends to stop someone or something from succeeding or making progress *e.g., One major disadvantage of the area is the lack of public transportation.*	a quality or circumstance that makes achievement unusually difficult *e.g., His lack of formal schooling was a serious disadvantage.*	something that places one in an unfavorable condition or position *Synonym:* drawback *e.g., A disadvantage to living there is that you'd have no access to public transportation.*
2. downside	the disadvantages or less positive aspects of something	a negative aspect *e.g., The downside of living in the country is, of course, the long commute to work.*	a disadvantageous aspect *e.g., An option with a downside as well as benefits*
3. drawback	a disadvantage or problem that makes something a less attractive idea *Synonym:* disadvantage *e.g., The main drawback to it is the cost.*	an objectionable feature *Synonym:* disadvantage *e.g., The plan's only drawback is its cost.*	a disadvantage or inconvenience

According to the foregoing definitions, the three nouns are viewed as near-synonyms of one another, meaning that they are similar in meaning, yet not universally interchangeable. Some examples and collocational information concerning the nouns are also provided; however, they are not exhaustive. More specifically, one verb collocate

of *disadvantage*, i.e., *outweigh*, and some of its adjective collocates, namely *serious, severe, considerable, main,* and *major*, are available in the Oxford Advanced American Dictionary. Nevertheless, it should be noted that the two other dictionaries do not offer any collocational behavior of the three synonymous nouns since they occasionally treat a lexical item as a separate entity, irrespective of contexts, including words that combine with a particular lexical item.

The inadequacy of information offered by the dictionaries may cause L2 learners difficulty in differentiating English near-synonyms and using them to fit a particular environment (Lee & Liu, 2009). In light of this inadequacy, it is hoped that this corpus-based research will bridge the gap by offering insights into the usage of the three synonymous targets.

In the following, the concept of synonymy, collocations, and previous research on English synonyms will be discussed.

Literature Review

Synonymy

Synonyms typically refer to terms which are semantically the same, yet whose sound patterns are different (Jaszczolt, 2002). Generally, there are two major categories of synonyms: absolute synonyms and near-synonyms.

Absolute Synonyms

Absolute synonyms, also known as *total synonyms*, refer to those that can be interchangeably used in any context with identical meaning. However, some scholars, e.g., Quine (1951) and Goodman (1952), assert that true synonymy does not exist. Others have said that if absolute synonyms were to be found, they would be extremely rare (Edmonds & Hirst, 2002; Hornby, 2014). Even if absolute synonyms may include words in different varieties of English, for example, *tap* (British English) and *faucet* (American English), such lexical units would alter the style of discourse (Edmonds & Hirst, 2002).

Near-Synonyms

Near-synonyms, also referred to as *partial synonyms*, or *plesionym* used by Cruse (1986), are lexical items which are very similar in meaning, and yet they are not interchangeable (Liu, 2010), mainly because of their subtle nuances of meanings, collocations, emphasis, or registers (DiMarco et al., 1993; Partington, 2004; Xiao & McEnery, 2006). In contrast to absolute synonyms, near-synonyms abound in language. Take *beautiful*, *attractive*, *gorgeous*, and *charming* as examples; the words are near-synonyms which have similar conceptual or denotational meanings, but which are also likely dissimilar in collocational terms (Xiao & McEnery, 2006). Another case in point is *child* and *kid*. The former is prevalent apart from in informal contexts, where the latter is normally opted for (Hornby, 2014).

Corpus Linguistics and English Synonyms

Corpus linguistics is a methodology for investigating authentic language use (Lindquist, 2009). A corpus is characterized by a collection of naturally-occurring texts that allows quantitative and qualitative insights into authentic language use to be gained (Biber et al., 1998). With respect to quantitative information, corpus linguistics concentrates on the frequency of distribution patterns of words or phrases across various text types, e.g., academic, blogs, spoken, fiction, TV/movie subtitles. As far as qualitative data are concerned, two major kinds of semantic relationships, namely collocation and semantic preferences, can be investigated to discern the subtle differences between near-synonyms (Schmitt, 2010; Sinclair, 2004).

Collocation

The initial explanation of a collocation was that it comprised consecutive strings that needed to be learned as a single unit (Palmer, 1933). Later, the British linguist, Firth (1957), refined the concept, stating that a collocation refers to other words that often appear with one word and affect it. More specifically, a collocation is defined as "the more-frequent-than-average co-occurrence of two lexical items within five

words of text" (Sinclair et al. 2004, xiii). A collocation, then, typically refers to a relationship between a word and its collocates, where a collocate is a word that appears frequently with another particular word (Webb & Nation, 2017).

Semantic Preference

Semantic preference usually involves a connection between a term and a list of its collocates that share certain semantic properties (Stubbs, 2002). As an illustration, Stubbs (2001, p. 65) noted that *large* collocates with a list of semantically linked words representing sizes and quantities, such as *numbers, amounts, scale, quantities,* and *part*.

Semantic preference is closely interrelated to semantic prosody. The former deals with semantic properties of collocates, whereas the latter expresses speakers' or writers' views through the contexts in which a word is located. However, they are intricately intertwined in that semantic preference leads to the establishment of semantic prosody, which determines the surroundings that help limit the number of collocates that goes along with the "node item" (Partington, 2004, p. 151).

Previous Corpus-Based Studies on English Synonyms

Several studies have compared and contrasted synonyms. The current study reviews a few studies on synonyms. Stubbs (2001) explored the word *large* in COCA, and the findings showed that no less than one fourth of the 56,000 tokens of *large* found in COCA (the 200-million-word corpus at that time) collocated with items denoting amounts, numbers, and scale.

By looking into the distributional evidence of five near-synonymous adjectives in COCA, i.e., *chief, main, major, primary,* and *principal*, Liu (2010) identified a number of semantic and functional discrepancies. Firstly, the five adjectives were ranked in order of their overall frequencies: *major, main, chief, primary,* and *principal*. Moreover, it was found that, although all of the adjectives are generally used to describe abstract or dual nouns such as *concern* or *component*, it is virtually impossible for *principal* to collocate with ranked titles, e.g.,

executive. However, *principal* usually modifies unranked titles like *author*, suggesting the level of contributions made by the noun that *principal* modifies. On the other hand, *main* serves to describe concrete nouns such as *road*; in this case, *main* denotes importance, rather than size, as shown in many dictionaries. In terms of registers, the spectrum of word formality, listed from greatest to lowest, is *principal, chief, major, main,* and *primary.*

Jirananthiporn (2018) studied the frequencies of *problem* and *trouble*, their distribution patterns across genres in COCA, and their verb and adjective collocations. She found that *problem* seems to be used more pervasively than *trouble*. Furthermore, *problem* is discovered in academic texts, but *trouble* is largely seen in fiction and spoken discourse. The former is, therefore, more formal than the latter.

Jarunwaraphan and Mallikamas (2020) compared and contrasted the two near-synonyms and surrounding contexts in COCA: *chance* and *opportunity*. The results revealed that, among COCA's five text categories: academic, newspapers, magazines, fiction, and spoken, *opportunity* was found most commonly in academic genres and least encountered in fiction. On the other hand, *chance* was found most frequently in spoken language and appeared least in the academic genre. The registers of collocates with the two synonymous nouns were also obviously found to be the same as those of the nouns.

Phoocharoensil (2020) searched through COCA to look into patterns of registers and collocations in which three synonyms *consequence, result,* and *outcome* frequently appear. The findings showed that these synonymous nouns are frequently used in academic texts despite only the tiniest frequencies in informal text types in the corpus. It was also found that *consequence* often co-occurs with semantically negative verbs and adjectives, and that the collocates of *result* are semantically linked to research-related settings. Words accompanying *outcome* cover the most extensive range of semantic features, yet they tend not to subscribe to any particular circumstances.

Unlike previous research, this current study aims to unveil distribution patterns and collocational profile of the three synonymous nouns: *disadvantage*, *downside*, and *drawback*. This will enable English language learners to distinguish the subtle differences in meanings between the lexical items; hence, enabling those learners to select one

over the others in a given genre or collocational environment. Adopting the corpus-based approach, the present research seeks to address the following questions:

1. How are the three synonymous nouns distributed across various text types?
2. What are their frequent collocates?

Methodology

The data of the present study were collected from COCA, a one-billion-word American English corpus with balanced genres, developed between 1990 and 2019. This corpus embodies eight text types: blogs, web pages, TV and movie subtitles, spoken texts, fiction, popular magazines, newspapers, and academic journals. It is pervasively used by English language researchers.

COCA is employed in this study for several major reasons. First and foremost, it is a balanced-genre corpus. This essentially means that it contains an equal proportion of each of its eight text types, thereby possibly yielding non-skewed results of lexical distribution across different genres. A further significant reason is that the corpus is representative of contemporary American English language use. Gradually developed between 1990 and 2019, it comprises one billion words, which is considerably larger than any other currently existing American English corpus. Lastly, COCA is a monitor corpus, which differs from a static corpus (Davies, 2010). The former is an updated resource to which new texts have been added from year to year since it was first generated, while the latter is not updated after its creation.

COCA was searched in an effort to determine the frequency of use of the three target synonyms across the eight text genres in COCA. In order to identify subtle distinctions between the lexical items, COCA was then searched in order to find their verb collocates whose frequency is at least 2, and whose MI scores are a minimum of 3. The frequency count in conjunction with the MI value works well in identifying nouns that are typically modified by adjectives or preceded by verbs as it favors verbs or adjectives with a general high frequency in the corpus; nevertheless, it underrates those verbs or adjectives that have low overall frequency, yet

often accompany the target synonymous nouns (Szudarski, 2018). Consequently, the frequency count amasses verbs and adjectives that regularly emerge with the nouns, yet which do not strongly collocate with them. The MI value, acquired from calculating the dependence of the two words, can play a fundamental role in addressing this flaw, although it tends to promote words which have a low general frequency, but which strongly collocate with the synonyms under investigation (Church et al., 1994), serving as useful data for further investigation in terms of semantic preferences of collocations. Additionally, the MI score is selected for use over T-score in this study as the latter captures high frequency function-word collocations with moderate relations, thereby failing to identify certain low frequency strongly related word combinations, whereas the former stresses less common content-word collocations with very strong relations (Church et al., 1994; Gablasova et al, 2017; Liu, 2010). It is also preferred over log-likelihoood, which focuses mainly on genre-specific collocations rather than their general English counterparts (Pojanapunya & Todd, 2018).

In the present study, the MI scores are set at a minimum of 3 simply because the established value can be indicative of the habitual co-occurrence of a lexical pair (Cheng, 2012; Hunston, 2002; Liu, 2010; Phoocharoensil, 2021). Twenty of the verb collocates selected according to the criteria were then categorized according to their semantic preference and analyzed qualitatively.

Subsequently, adjective collocates of the three synonyms were collected from the corpus, employing the same criteria as those used in selecting the verb collocates. Also, the selection threshold was raised to 30 tokens as the number of the adjective collocates exceeded that of the verb counterparts. Then the adjective collocates were grouped into their semantic preferences and further compared.

Results and Discussion

To fulfil the objectives of the present study, firstly, the findings concerning the overall frequencies and distribution patterns of the three synonyms across eight different text types in COCA are summarized and discussed, followed by discussion of the findings regarding the verb and adjective collocates of the three synonyms.

Frequencies and Distribution Patterns

Both raw frequency and normalized frequency—frequency per million (PM) words—of the terms under investigation should be reported quantitatively (McEnery & Hardie, 2012). Hence, the frequencies of the three synonymous nouns across eight genres in COCA are summarized in Table 2. As can be seen, *disadvantage* ranks first, with 4,820 tokens (39.06 PM), followed closely by *downside*, which accounts for 4,443 tokens (35.52 PM). The word *drawback* is by far the least frequent, occurring at less than one-third of the frequency of the other two synonyms (1,690 instances, or 13.65 PM).

Table 2

Frequency and distribution of disadvantage, downside, and drawback across the text genres in COCA, from highest to lowest

Genre	disadvantage		Genre	downside		Genre	drawback	
	Frequency	Per million		Frequency	Per million		Frequency	Per million
Academic	1,431	11.95	Magazines	1,097	8.70	Magazines	541	4.29
Web pages	690	5.55	Blogs	840	6.53	Academic	316	2.64
Blogs	687	5.34	Newspapers	704	5.78	Newspapers	273	2.24
Magazines	620	4.92	Web pages	648	5.22	Web pages	191	1.54
Newspapers	598	4.91	Spoken	569	4.51	Blogs	185	1.44
Spoken	429	3.40	Academic	228	1.90	Fiction	87	0.74
Fiction	208	1.76	TV/movies subtitles	210	1.64	Spoken	53	0.42
TV/movies subtitles	157	1.23	Fiction	147	1.24	TV/movies subtitles	44	0.34
Total	4,820	39.06		4,443	35.52		1,690	13.65

With respect to the degree of formality that the three synonyms convey, Table 2 suggests that the word *disadvantage* appears most formal as it was found mostly in academic texts, comprising 1,431 tokens, or 11.95 PM. The word is also seen in less formal genres than academic journals, i.e., general web pages, blogs, magazines, and newspapers, with occurrences numbering 690 (5.55 PM), 687 (5.34 PM), 620 (4.92 PM), and 598 (4.91 PM), respectively. The three genres where the term occurs least frequently are all informal – spoken, fiction, and TV/Movie subtitles – which comprise 429 (3.40 PM), 208 (1.76 PM), and 157 (1.23 PM) instances, respectively.

In contrast to *disadvantage*, the word *downside* seems less formal as evidenced by its frequency of occurrence being highest in magazines (1,097 tokens, or 8.70 PM), followed by blogs, newspapers, web pages, and even spoken, with 840 (6.53 PM), 704 (5.78 PM), and 648 (5.22 PM) occurrences, respectively. This informality tends to be strongly reinforced by the word's considerably lower frequency in academic texts (228 tokens, or 1.90 PM), being less often seen in TV, and occurring the least in fiction, with 210 (1.64 PM) and 147 (1.24 PM) instances, respectively.

Similarly, the word *drawback* is often present in less formal registers, with its highest frequency in magazines (541 tokens, or 4.29 PM). The second and third largest number of tokens were found in academic texts and newspapers, making up 316 (2.64 PM) and 273 (2.24 PM) occurrences, respectively. The two least popular text types, also both extremely informal, are TV (44 instances, or 0.34 PM) and spoken (53 instances, or 0.42 PM).

The findings conform to a number of prior studies, such as Liu (2010), Phoocharoensil (2020), Jirananthiporn (2018), Jarunwaraphan and Mallikamas (2020), who argued that synonyms generally occur in varying degrees of formality.

Verb Collocates and Distribution Patterns

This section reveals verb collocates that usually co-occur with the target synonyms *disadvantage*, *downside*, and *drawback* in COCA. The identified verb collocates have a minimum MI value of 3 (Cheng, 2012; Hunston, 2002; Liu, 2010; Phoocharoensil, 2021).

Table 3 shows that 8 verbs habitually collocate with *downside*, while only 5 verbs do so with *drawback* at the given frequency and MI value (at least 3); on the other hand, there are more than 20 verb collocates of *disadvantage*. Consequently, merely the top-20 verb collocates of *disadvantage* are listed in the table, along with all of the collocates for *downside* and *drawback*. Interestingly, the three synonyms share some verb collocates, namely *outweigh, offset,* and *minimize*. This evidence tends to reinforce their similarity of meaning.

Table 3

Verb collocates of disadvantage, downside, and drawback in COCA

	disadvantage			downside			drawback		
Rank	Verb collocates	Frequency	MI value	Verb collocates	Frequency	MI value	Verb collocates	Frequency	MI value
1	compare	97	3.73	limit	34	3.61	outweigh	36	7.95
2	face	89	3.21	outweigh	18	6.53	overcome	15	4.14
3	suffer	69	3.75	offset	6	4.23	offset	5	4.41
4	overcome	65	5.33	tilt	6	3.95	minimize	4	3.12
5	outweigh	57	7.69	minimize	6	3.28	mitigate	3	4.16
6	associate	42	3.37	underestimate	4	3.45			
7	compete	22	3.28	mitigate	3	3.73			
8	weigh	20	3.48	dwell	3	3.35			
9	impose	18	3.27						
10	compensate	17	4.76						
11	offset	16	5.14						
12	arise	14	3.01						
13	minimize	9	3.35						
14	labor	7	4.58						
15	confer	6	4.09						
16	compound	6	3.93						
17	exacerbate	5	3.93						
18	overturn	5	3.41						
19	remedy	4	4.31						
20	impede	4	4.05						

From Table 4, the nouns *disadvantage, downside, and drawback* present different distribution patterns of their verb collocates.

Table 4

Distributional patterns of verb collocates of disadvantage, downside, and drawback in COCA

disadvantage	verb + disadvantage	face, suffer, overcome, outweigh, weigh, impose, compensate, offset, minimize, labor, compound, confer, overturn, remedy, exacerbate
	disadvantage + verb	compare, associate, compete, arise, impede, exacerbate
downside	verb + downside	limit, outweigh, offset, tilt, minimize, underestimate, mitigate, dwell
drawback	verb + drawback	outweigh, overcome, offset, minimize, mitigate

Collocate Examples for Disadvantage

(1) When you have the White House, you sort of ***face a disadvantage*** in midterm elections, and so Republicans want to use this to energize their voters, to mobilize them, and to get them out to the polls. (SPOK)

(2) A considerable procedural ***disadvantage arises*** from this informality and from the absence of any form of discovery or pleadings when middle or senior management employees bring claims involving complex legal issues and potentially large sums of money. (ACAD)

(3) Ensure that any such legislation or measures do not criminalise the behaviour of, stigmatise, or in any other way, ***exacerbate the disadvantage*** of those vulnerable to such practices. (WEB)

(4) Our results suggest that social ***disadvantage exacerbates*** the deleterious health effects of lead. (ACAD)

As can be seen in Table 4 and the example sentences above, the verb collocates of the noun *disadvantage* occur in two environments. Firstly, *disadvantage* generally follows the verbs *face, suffer, overcome, outweigh, weigh, impose, compensate, offset, minimize, labor, compound, confer, overturn*, and *remedy,* as in (1). Secondly, the noun *disadvantage* heads the verbs *compare, associate, compete, arise,* and *impede,* as illustrated in (2). Nonetheless, it is worth noticing that both of the distribution patterns apply to the verb *exacerbate,* as seen in (3) and (4).

Collocate Example for Downside

Although *disadvantage* displays two differing distribution patterns, its synonymous noun *downside* has only one, which involves its placement after the verbs *limit, outweigh, offset, tilt, minimize, underestimate, mitigate,* and *dwell*, as exemplified in (5).

(5) We also have a big position in Microsoft (MSFT) – the release of Windows 8 looks underwhelming – but its valuation and monopoly-like positions **limit the downside**, in our opinion. (WEB)

Collocate Example for Drawback

As with the placement of *downside*, the noun *drawback* has a single distribution pattern throughout COCA, i.e., that following the verbs *outweigh, overcome, offset, minimize,* and *mitigate,* as in (6)*.*

(6) The process was designed to **overcome drawbacks** of previous processes such as slow cycle time, manual spray-up, difficult demolding, poor reinforcement placement consistency, waste, and energy consumption. (ACAD)

Semantic Preferences

A further stage of the present study is examining the semantic preferences of the three synonyms by categorizing their verb collocates according to their semantic similarities. Semantic preference typically

refs to the co-occurrence of a term and its collocates that can be subsumed under the same semantic category (Sinclair, 2004).

Table 5

Semantic preferences of verb collocates of disadvantage

Semantic preferences	Verb collocates of *disadvantage*
1. CONSIDER	compare, outweigh, weigh, offset, confer
2. DEAL WITH	face, overcome, compete, labor, overturn, impede, compensate, minimize, remedy
3. WORSEN	exacerbate, compound
4. HAVE	suffer, arise, impose
5. RELATE	associate

(7) Overall, the benefits of conducting research with student teachers **outweigh the disadvantages**. (ACAD)

(8) Race is not a proxy for disadvantage, because not all and not only blacks (or Mexican-Americans) have **suffered disadvantage**. (BLOG)

(9) I guess most of the **disadvantages that are associated** with a small company relate to budgets and resources. (WEB)

Through an in-depth exploration of the semantic preferences of *disadvantage,* five thematic classifications of the verb collocates of the noun under study emerged, as shown in Table 5. The first thematic classification CONSIDER contains the majority of verb collocates, denoting mental activities being performed in response to situations that cause problems or setbacks to somebody, namely *compare, outweigh, weigh, offset,* and *confer*, as exemplified in (7). The second theme DEAL WITH, which accounts for the largest number of the collocates, namely *face, overcome, compete, labor, overturn, impede, compensate, minimize,* and *remedy,* is concerned with ways in which an unfavorable circumstance is removed or reduced to a minimum. The third theme is

WORSEN, which comprises the verb lemmas which demonstrate an aggravating effect of an unpleasant condition, or an adverse situation which leads to a more severe effect, with *exacerbate* and *compound* subscribing to this thematic category. Fourthly, the theme HAVE, generally referring to experience, or an unfavorable circumstance or situation, is composed of the verbs *suffer*, *arise*, and *impose*, as exemplified in (8). The final thematic category is RELATE, which has only one member, *associate*, as illustrated in (9).

Table 6

Semantic preferences of verb collocates of downside

Semantic preferences	Verb collocates of *downside*
1. CONSIDER	outweigh, offset, underestimate, dwell
2. DEAL WITH	limit, minimize, mitigate
3. WORSEN	tilt

(10) Their confidence in that wizardry and their own ideas may lead them to **underestimate the downsides** and even dangers of the work they are funding, say some science philosophers, historians and economists. (NEWS)

(11) Officials said that risks to its outlook for inflation have **tilted to the downside** since its previous meeting, which could lead policy makers to take a more cautious tack when it comes to monetary policy. (MAG)

Compared to the verb collocates of *disadvantage*, those of *downside* are smaller in number, which possibly results from the lower proportion of its total tokens than that of *disadvantage,* as can be seen in Table 6. A closer examination of the verb collocates of *downside* reveals three central themes based upon the term's semantic preferences. These themes are parallel to three counterparts emerging from the semantic preferences of *disadvantage*: CONSIDER, DEAL WITH, and WORSEN. Nevertheless, certain verb collocates of *downside* that adhere to these thematic classifications appear to differ from those of *disadvantage*. The

first and most popular theme emerging from the investigation of semantic preferences of *downside* is CONSIDER, which includes four different verb collocates, *outweigh, offset, underestimate,* and *dwell*, as exemplified in (10). A next theme DEAL WITH contains three verb collocates, i.e., *limit*, *minimize*, and *mitigate*, and the last theme is WORSEN, with *tilt* being its sole member, as seen in (11).

Table 7

Semantic preferences of verb collocates of drawback

Semantic preferences	Verb collocates of *drawback*
1. CONSIDER	outweigh, offset
2. DEAL WITH	overcome, minimize, mitigate

(12) Deciding whether the benefits of suing **outweigh the drawbacks** is a very individual, fact-specific decision. (ACAD)

(13) Researchers now report in ACS Applied Nano Materials a new, calcium-based conservation treatment inspired by nature that **overcomes many drawbacks** of currently used methods. (MAG)

As regards the semantic preferences of the verb collocates of *drawback*, as shown in Table 7, two themes are derived from the corpus-informed data, which is fewer than the number associated with either *disadvantage* or *downside*. This may be partly attributed to the smaller proportion of the instances of *drawback* than its two counterparts being analyzed. It is also of interest to note that the two themes and the verb collocates of *drawback* are included on the lists of themes and members pertaining to *disadvantage* and *downside*. This suggests synonymy applies to the three nouns under investigation. CONSIDER, which is the first theme of *drawback*, consists of *outweigh* and *offset*, as in (12), and the second theme DEAL WITH encompasses the verb collocates *overcome*, *minimize*, and *mitigate*, as in (13).

Adjective Collocates

After the common verb collocates of *disadvantage*, *downside*, and *drawback*, were studied, adjectives frequently co-occurring with the target nouns were, in turn, identified from the corpus data. The findings are summarized in Table 8.

Table 8

Adjective collocates of *disadvantage*, *downside*, and *drawback* in COCA

Rank	disadvantage			downside			drawback		
	Adjective collocates	Frequency	MI value	Adjective collocates	Frequency	MI value	Adjective collocates	Frequency	MI value
1	competitive	237	7.05	only	269	4.18	only	211	4.11
2	economic	166	4.02	potential	106	5.05	major	155	4.8
3	distinct	102	6.36	significant	49	3.25	main	89	4.75
4	significant	87	3.57	obvious	33	3.74	serious	80	4.09
5	huge	72	3.38	minor	11	3.22	potential	78	5.03
6	serious	69	3.11	minimal	8	3.8	significant	48	3.66
7	severe	64	4.98	considerable	7	3.05	obvious	41	4.38
8	potential	54	3.56	steep	5	3.24	minor	21	4.45
9	relative	53	5.33	definite	4	3.89	inherent	16	5.55
10	socioeconomic	43	6.5	unmentioned	4	3.86	chief	14	3.05
11	obvious	43	3.61	glaring	3	4.91	principal	11	4.32
12	inherent	27	5.35	unintended	3	4.24	slight	11	3.84
13	slight	23	4	unanticipated	2	5.45	practical	9	3.12
14	unfair	22	4.63	pernicious	2	5.11	apparent	8	3.25
15	perceived	20	4.81	pesky	2	4.75	definite	5	4.56
16	comparative	19	5.44	demonic	2	4.5	perceived	5	3.77
17	tremendous	19	3.83	scant	2	4.03	well-known	4	3.45
18	considerable	17	3.82	staggering	2	3.78	sole	4	3.15

19	concentrated	16	6.39	nonexistent	2	3.35	noticeable	3	4.4
20	enormous	16	3.05				nutritional	3	4.18
21	decided	15	8.76				immense	3	3.58
22	structural	15	9.35				mega	2	4.64
23	reciprocal	11	6.09				utmost	2	4.3
24	marked	10	4.76				daunting	2	3.66
25	systematic	9	3.76				daunting	2	3.66
26	cumulative	8	4.76						
27	definite	8	4.38						
28	tactical	8	4.28						
29	demographic	8	3.24						
30	profound	8	3.08						

The corpus-informed data displayed in Table 8 validate the synonymy of *disadvantage*, *downside*, and *drawback* as the three lexical items frequently co-occur with some adjectives, viz., *significant, obvious, definite,* and *potential*. Apart from this, however, *considerable* was found to be the only adjective collocate that *disadvantage* and *downside* have in common, although it usually does not precede *drawback*. Additionally, the nouns *disadvantage* and *drawback* share three adjectives – *perceived, serious,* and *inherent* – that are not regular collocates of *downside*. Lastly, the pre-modifier *only*, which is regularly used with both *downside* and *drawback*, does not appear to collocate with *disadvantage*.

The adjective collocates of *disadvantage* presented in Table 8 were subsequently categorized into a variety of themes according to their semantic preferences. The findings are displayed in Table 9 below.

Semantic Preferences of Adjective Collocates of Disadvantage

Table 9

Semantic preferences of adjective collocates of disadvantage

Semantic preferences	Adjective collocates of *disadvantage*
1. EXTENT	significant, huge, tremendous, considerable, concentrated, enormous, decided, cumulative, profound, relative, comparative, slight
2. ASPECT	competitive, economic, socioeconomic, structural, demographic, tactical, systematic
3. PROMINENCE	distinct, obvious, perceived, marked, definite
4. NEGATIVE SENSE	serious, severe, unfair
5. POSSIBILITY	potential
6. PART	inherent
7. RELATIONSHIP	reciprocal

(14) Most resorts have developed their own water sources, such as ponds and small lakes, which provide near limitless quantities of water to create snow. That puts resorts without water at a **huge disadvantage** early in the season. (NEWS)

(15) Mr. Obama also faces **demographic disadvantages**. While he draws young people and those with college degrees, Pennsylvania has one of the highest concentrations of people over 65 (15 percent, compared with a national average of 12 percent) and one of the lowest of people with college degrees (22 percent, compared with a national average of 24 percent). (NEWS)

Table 9 indicates seven themes of the adjective collocates of *disadvantage*: EXTENT, ASPECT, PROMINENCE, NEGATIVE SENSE, POSSIBILITY, PART, and RELATIONSHIP. Of 30 adjective collocates of *disadvantage*, twelve, consisting of *significant, huge, tremendous, considerable, concentrated, enormous, decided, cumulative, profound, relative, comparative,* and *slight,* are categorized into the theme EXTENT, which involves how great, important, or severe a disadvantage is, as in (14). It is worth mentioning that the adjectives subsumed under this theme can be grouped into two meanings, namely being large and small, with the former containing considerably more members than the latter. This theme corresponds to virtually all themes of verb collocates of *disadvantage*, i.e., CONSIDER, DEAL WITH, WORSEN, and HAVE, thereby

suggesting the degree of a disadvantage that may affect a situation or people involved. The second thematic category ASPECT encompasses adjectives which denote several different facets or subject areas, namely *competitive, economic, socioeconomic, structural, demographic, tactical,* and *systematic*, as in (15)*.*

(16) If some test-takers are enhancing their mental performance, arguably, artificially, through the use of psychoactive drugs, then it puts those competing with them but not using the drugs at a ***distinct disadvantage***. (ACAD)

(17) Mrs. Jones, who has a lawsuit pending against the school system, would also like to change the grading system that she believes places Clayton students at an ***unfair disadvantage*** with students from other systems. (NEWS)

(18) Finally, the temperature of eardrops is a ***potential disadvantage***. Cold solutions placed in the external auditory canal can be quite uncomfortable, especially in children. Children may have difficulty distinguishing between pain and the unpleasant sensation of cold medication. (ACAD)

(19) A small Third World nation cannot function in this age of interdependence without some trade interaction with the outside world. But it has ***inherent disadvantages*** in this interaction. To begin with, most of its export items are likely to be primary products which usually do not command high prices in the international market. Secondly, especially if it has a colonial background, a small nation will probably have a higher ratio of foreign trade to domestic production than the industrialized countries. (ACAD)

(20) Only the political leadership can effectively explain to the masses the common incentives of cooperation and the ***reciprocal disadvantages*** stemming from discord. (ACAD)

As far as the theme PROMINENCE is concerned, there are five adjectives which refer to the state of a disadvantage being easily seen or

well known: *distinct, obvious, perceived, marked,* and *definite*, as exemplified in (16). Moreover, the thematic classification NEGATIVE SENSE suggesting unfavorable and deplorable circumstances consists of three adjectives, namely *serious, severe,* and *unfair*. The word *unfair* was found to indicate a situation in which a disadvantage does not apply equally to everybody involved, as in (17). The three remaining themes encompassing POSSIBILITY, PART, and RELATIONSHIP each contain a single member. POSSIBILITY, which refers to the chance that a disadvantage might arise or occur to somebody, is composed of the adjective collocate *potential*, as in (18), while PART, denoting a disadvantage's natural occurrence or that it is impossible to avoid, and RELATIONSHIP, indicating the way in which two people or groups are affected by a disadvantage, consist of *inherent* and *reciprocal*, respectively, as in (19) and (20).

Semantic preferences of adjective collocates of downside

Under thorough scrutiny of the adjective collocates of *downside*, five themes emerged according to their semantic preference. The results are summarized in Table 10.

Table 10

Semantic preferences of adjective collocates of downside

Semantic preferences	Adjective collocates of *downside*
1. EXTENT	significant, steep, considerable, staggering, only, minor, minimal, scant
2. PROMINENCE	obvious, definite, glaring
3. NEGATIVE SENSE	pernicious, pesky, demonic
4. POSSIBILITY	potential, nonexistent
5. EXPECTATION	unintended, unanticipated, unmentioned

From Table 10, the adjectives co-existing with *downside* are arranged into five central themes, being slightly lower in number than themes associated with *disadvantage*. The most dominant theme EXTENT

constitutes the largest proportion of the adjective collocates, that is, eight out of 19, comprising *significant, steep, considerable, staggering, only, minor, minimal,* and *scant*. This corresponds to the results of verb collocates of *downside*, revolving predominantly around the degree of a *downside's* severity. Like the adjective collocates of *disadvantage*, their counterparts related to *downside* roughly fall into two meanings: large and small, albeit with the former containing a slightly higher number of adjectives than the latter. It is obvious that *downside* is described by more adjectives, signifying a smaller amount than *disadvantage*, yet it is not found to be modified by any adjectives under the ASPECT theme, as is *disadvantage*.

An additional theme of the adjectives co-occurring with *downside* is PROMINENCE, comprised of three lexical items, i.e., *obvious, definite,* and *glaring*. This might demonstrate that *downside* is often clearly seen. The third theme, NEGATIVE SENSE, embodies the three adjectives *pernicious, pesky,* and *demonic*. It is noteworthy that, according to Cambridge Dictionary accessed online, *pernicious* is regarded as a formal word, yet *pesky* is not, and *demonic* appears untagged. This is in line with the fact that the noun *downside* is oftentimes discovered in a mix of text genres, ranging from academic to magazines, and blogs to spoken. The fourth theme, POSSIBILITY, has two members: *potential* and *nonexistent*. Lastly and interestingly, the theme EXPECTATION, which is related to one's expectations about a downside, comprises three collocates, *unintended, unanticipated,* and *unmentioned*, as exemplified in (21).

(21) But there has been an **unintended downside**: That ease, combined with the huge pool of job seekers, now means that employers are overwhelmed with job applications. (WEB)

It is worth noticing that each of the collocates include the prefix -*un*, which would imply the way in which a downside runs contrary to one's expectations.

Semantic preferences of adjective collocates of drawback

Table 11

Semantic preferences of adjective collocates of drawback

Semantic preferences	Adjective collocates of *drawback*
1. EXTENT	major, main, significant, immense, mega, chief, principal, utmost, only, sole, slight, minor
2. PROMINENCE	obvious, apparent, definite, perceived, noticeable, well-known
3. ASPECT	nutritional, practical
4. NEGATIVE SENSE	serious, daunting
5. POSSIBILITY	potential
6. PART	inherent

As regards *drawback*, six principal themes emerged from the analysis of its 24 adjective collocates. The overarching theme EXTENT embraces ten adjectives: *major, main, significant, immense, mega, chief, principal, utmost, slight,* and *minor.* The second most preponderant theme PROMINENCE covers six adjectives, i.e., *obvious, apparent, definite, perceived, noticeable,* and *well-known*. It is worthy of note that the adjective collocates of *drawback* that are classified under the theme PROMINENCE exceed the number of their counterparts adhering to the same theme and related to *disadvantage* and *downside*, although the total of *drawback*'s adjective collocates is somewhat lower than that of its synonym, *disadvantage*. Next, two adjective collocates of *drawback* pertain to ASPECT, namely *nutritional* and *practical*. Nonetheless, it should be pointed out that the two adjective collocates of *drawback* connected with the theme ASPECT are substantially more restricted in number than the seven counterparts of *disadvantage* related to ASPECT. What is more, two lexical items usually preceding *drawback*, i.e., *serious* and *daunting*, are incorporated into NEGATIVE SENSE. Another theme, POSSIBILITY, includes one adjective collocate, *potential*, while the final theme, PART, contains a single lexical unit, *inherent*.

In sum, multiple adjective collocates of the three nouns *disadvantage, downside,* and *drawback* have four themes in common: EXTENT, PROMINENCE, NEGATIVE SENSE, and POSSIBILITY, which attests to the synonymous quality of the three lexical items. Notwithstanding the shared themes of the adjectives placed before the target nouns, there exist certain subtle divergences of semantic preference to which the

nouns are related. Adjectives collocating with the noun *disadvantage* often indicate aspects or subject matter. Meanwhile, those co-occurring with *downside* represent a downside that has arisen beyond expectation, or one which has been treated differently from how it should have been treated. Finally, the adjectives accompanying *drawback* are frequently associated with public attention that a drawback has attracted. This substantiates the notion put forward by previous studies such as those of Phoocharoensil (2020), Jirananthiporn (2018), Hornby (2014), that some collocates tend to come with a synonym, but do not do so with another synonym in the same set.

Conclusion

The present study examined similarities and differences between the three synonymous nouns *disadvantage*, *downside*, and *drawback*, paying attention to their frequency and occurrences across various text types in COCA, and their common verb and adjective collocates. It was found that *disadvantage* ranks first in overall frequency, followed by the frequency *downside* and *drawback*, respectively. More precisely, the total frequency of *disadvantage* is slightly more than that of *downside*, but is almost three times that of *drawback*. With respect to registers, *disadvantage* is predominantly found in formal or academic texts, while *downside* and *drawback* seem to be less formal since they prevail in magazines and other less formal sources.

In terms of verb collocates of the three target synonyms in COCA, *disadvantage* comprises more common verb collocates than *downside* and *drawback*. Given the semantic preferences of their verb collocates, the three synonyms share two themes, CONSIDER, suggesting that they involve careful consideration, and DEAL WITH, suggesting required action.

It should be noted, however, that although *disadvantage* has the largest number of adjective collocates, *downside* has fewer common adjective collocates than *drawback*. A rigorous investigation into the semantic preferences of adjective collocates discloses fine-grained distinctions among the three target synonyms. It was discovered that while *disadvantage*, *downside*, and *drawback* mainly co-occur with adjectives under the theme EXTENT, being enormous degrees in

particular, *downside* tends to co-occur with more adjectives representing a small degree than does *disadvantage*, reflecting increased flexibility when the concept of extent of a downside is expressed. Intriguingly, *downside* frequently combines with some adjectives representing the notion that the downside appears contrary to an individual's expectation about it. Additionally, *disadvantage* is more often preceded by adjectives connected to the theme ASPECT than *drawback* is. This is not the case for *downside*, i.e., it is not described by any adjectives denoting aspects or types. It is also worth mentioning that *drawback* is more likely to be accompanied by adjective collocates organized into the theme PROMINENCE than are the other two synonyms.

The present study, nevertheless, has some limitations. Firstly, in analyzing semantic preferences of collocates, it is sometimes difficult to classify a few of them into one or another thematic category. This challenge can be addressed by examining the contexts in which collocates appear, coupled with reading the definitions of collocates provided in English dictionaries. Additionally, it concentrates solely on American English; therefore, its findings might not be able to be generalized to other English varieties. Finally, the present research does not consider associations between the synonymous items and grammatical words that co-occur with them such as English definite and indefinite articles. Future research should consider grammatical aspects of the three synonyms; alternatively, it may select other synonyms, either in this set of meaning – such as *snag* or *pitfall* – to provide a more comprehensive picture of the synonym usage, or in a different set in order to guide language learners toward more accurate and natural use of synonyms. Moreover, further research can be undertaken across English varieties other than American English.

It is expected that the findings of the study will be beneficial to English language teachers who wish to develop lessons and teaching materials for synonym usage. The results will also help explain the distribution patterns of the synonymous items across numerous genres and their collocation profile, intriguing individuals who intend to delve into similarities and differences in word meanings, using corpora as a source of information.

Acknowledgments

About the Author

Piyaboot Sumonsriworakun: An English language lecturer at Chulalongkorn University Language Institute, Thailand. His research interests include corpus linguistics, second language acquisition, and English language teaching.

References

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Cambridge University Press. (n.d.). In *Cambridge.org dictionary*. Retrieved March 10, 2022, from https://dictionary.cambridge.org/dictionary/english/

Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. Routledge.

Church, K.W., Gale, W., Hanks, P., Hindle, R., & Moon, R. (1994). Lexical substitutability. In B. T. S. Atkins & A. Zampolli (Eds.), *Computational approaches to the lexicon* (pp. 153–177). Oxford University Press.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing: Journal of the Association for Literary and Linguistic Computing, 25*(4), 447–464. https://doi.org/10.1093/llc/fqq018

Dewaele, J. (2008). Appropriateness in foreign language acquisition and use: Some theoretical, methodological and ethical considerations. *International Review of Applied Linguistics in Language Teaching, 46*(3), 245–265.

DiMarco, C., Hirst, G., & Stede, M. (1993). The semantic and stylistic differentiation of synonyms and near-synonyms. *AAAL Spring Symposium on Building Lexicons for Machine Translation*, *1*, 114–121.

Edmonds, P., & Hirst, G. (2002). Near-Synonymy and lexical choice. *Computational Linguistics, 28*(2), 105–144. https://doi.org/10.1162/089120102760173625

Firth, J. R. (1957), A synopsis of linguistic theory. In *Studies in Linguistic Analysis* (pp. 1930–1955). Blackwell.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning, 67*(S1), 155-179.

Goodman, N. (1952). On likeness of meaning. In L. Linsky (Ed.), *Semantics and the philosophy of language* (pp. 67–74). University of Illinois Press.

HarperCollins Publishers. (n.d.). In Ahdictionary.com. Retrieved March 10, 2022, from https://ahdictionary.com

Hornby, D. (2014). *Linguistics: A complete introduction*. Hachette.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Jarunwaraphan, B., & Mallikamas, P. (2020). A corpus-based study of English synonyms: 'Chance' and 'opportunity'. *REFLections, 27*(2), 218–245.

Jaszczolt, K. (2002). *Semantics and pragmatics: Meaning in language and discourse.* Longman.

Jirananthiporn, S. (2018). Is this problem giving you trouble? A corpus-based examination of the differences between the nouns 'problem' and 'trouble'. *Thoughts, 2*, 1-25.

Lee, C. & Liu, J. (2009). Effects of collocation information on learning lexical semantics for near synonym distinction. *Computational Linguistics and Chinese Language Processing, 14*(2), 205-220.

Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh University Press.

Liu, D. (2010). Is It a 'chief', 'main', 'major', 'primary', or 'principal' concern? A corpus-based behavioral profile study of near-synonyms. *International Journal of Corpus Linguistics, 15*(1), 56–87. https://doi.org/10.1075/ijcl.15.1.03liu

Mackay, S. (1980). Teaching the syntactic, semantic, and pragmatic dimensions of verbs. *TESOL Quarterly, 14*, 17-26.

Martin, M. (1984). Advanced vocabulary teaching: The problem of synonyms. *The Modern Language Journal, 68* (2), 130-137.

McEnery, T. & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Merriam-Webster. (n.d.). In Merriam-Webster.com dictionary. Retrieved March 10, 2022, from https://www.merriam-webster.com

Oxford University Press (n.d.). In Oxfordlearners.com dictionaries. Retrieved March 10, 2022, from https://www.oxfordlearnersdictionaries.com/definition/american_english

Palmer, H. E. (1933). *Second interim report on English collocations*. Kaitakusha.

Partington, A. (2004). Utterly content in each other's company: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics, 9*(1), 131-156.

Phoocharoensil, S. (2020). A genre and collocational analysis of consequence, result, and outcome. *3L: The Southeast Asian Journal of English Language Studies, 26*(3), 1–16. https://dx.doi.org/10.17576/3L-2020-2603-01

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. https://doi.org/10.32601/ejal.911269

Pojanapunya, P., & Todd, R. W. (2018). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory, 14*(1), 133-167. https://doi.org/10.1515/cllt-2015-0030

Quine, W. V. O. 1951. Two dogmas of empiricism. *Philosophical Review*, 60, 20–43.

Schmitt, N. (2010). *Research vocabulary: A Vocabulary research manual.* Palgrave Macmillan.

Sinclair, J. (2004). *Trust the text*. Routledge.

Sinclair, J., Susan J., & Robert D. (2004) *English collocation studies: The OSTI report* (Krishnamurthy, R. (Ed.). Continuum.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical*

semantics. Blackwell.

Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics, 7* (2), 215–244.

Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research*. Routledge.

Webb, S. & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Xiao, R. Z. & McEnery, A. M. (2006). Collocation, semantic prosody and near synonymy: A crosslinguistic perspective. *Applied Linguistics, 27*(1), 103-129. https://doi.org/10.1093/applin/ami045