Effectiveness of Corpus in Distinguishing Two Near-Synonymous Verbs: *Damage* and *Destroy*

Qiuyuan Song[1]

[1] Faculty of English and Education, Guangdong University of Foreign Studies, Guangzhou, China

Correspondence: Qiuyuan Song, Faculty of English and Education, Guangdong University of Foreign Studies, No. 178 Waihuan Road (East), Panyu District, Guangzhou City, China.

Abstract

This study aims to explore how corpus-based approaches can be used to address the distinctions of English near-synonyms effectively. Especially, it collected source data from the British National Corpus (BNC) and adopted Sketch Engine (SkE) as an analyzing tool to compare the near synonymous pair *damage* and *destroy* commonly misused by Chinese-speaking learners of English in terms of frequencies, genre distribution, colligation and collocation, differences in meanings and uses. It is found that *damage* and *destroy* are near-synonyms because they are relevant words and share most collocates but they are not fully intersubstitutable for certain contexts. Some words related to the human body or physical health are more collocated with *damage* and some such as military affairs and one's thought or belief more with *destroy*. In addition, the core meaning of *damage* gives more emphasis on something that can be recovered but does not work well as before, while *destroy* offers more senses for something that no longer exists. Furthermore, the British tend to collocate the two near-synonyms with the same word to create a build-up, because *destroy* is endowed with a stronger degree of destruction than *damage*. The study ends by suggesting corpus-based analysis should be promoted in language teaching and learning to improve the accurate use of English vocabulary by language learners.

Keywords: near-synonym, corpus, damage, destroy, BNC, Sketch Engine

1. Introduction

The English language has a large number of synonyms. This idea is echoed by Liu & Espino (2012) that rich synonyms enable English speakers "to convey meanings more precisely and effectively" (p. 198). But these kinds of variations in meanings and usage have greatly challenged English language teaching and learning (Ahmad et al., 2019). Traditionally, dictionaries are the main reference materials for language teachers and learners to discriminate synonyms. Although they could offer general and core meanings of the concepts of these synonyms, there is an absence of information on the nuances of near-synonyms or overlap of interpretations. For example, according to the Oxford Advanced Learner's Dictionary (OALD) (2009), *damage* denotes "to harm or spoil sth/sb" (p. 500), while *destroy* means "to damage sth so badly that it no longer exits, works, etc." (p. 543). In this descriptive comparison, *destroy* has shared some basic meanings with *damage*, since one is decoded by recourse to another. This semantic overlap between definitions may cause the potential for ambiguity. In addition, the definitions of *damage* and *destroy* do not sufficiently define contextual bounds. In light of this, synonym distinction and appropriate lexical choice really daunt language teachers and learners (Mackay, 1980).

The advent of corpus linguistics has made a great shift in vocabulary studies. Language educators and researchers can use a corpus, "a large collection of authentic texts that have been selected and organized following precise linguistic criteria" (Sinclair, 1991, 1996; Leech, 1991; Williams, 2003), to do linguistic analysis (e.g., lexis, multiword phrases). The corpus-based approach of language analysis is rationalized by many scholars and researchers (e.g., Shahzadi et al., 2019; Flowerdew, 2013; Albader, 2001; Richard & Tony, 2006). They believe that this approach is more reliable because authentic data, rather than intuition, can assist language teachers and researchers to find differences in the use of language. They also argue that people can get much larger amounts of text from an electronic corpus than the time when they do that manually. Moreover, it is an effective computational tool to reveal patterns that may not be obvious to the naked eye. Additionally, corpus analysis is suitable to be used to understand the similarities and differences among near-synonyms, and it helps

identify more specific criteria and suggestions for the usage of these apparently similar and interchangeable words. Therefore, this approach will be applied in the study to investigate two English near synonymous verbs *damage* and *destroy* commonly misused by Chinese-speaking learners of English in British National Corpus.

2. Literature Review

2.1 Synonyms

Synonymy is one of the relations that exist between different lexical items. Two types of synonyms, namely "perfect or absolute synonyms" and "near synonyms", are mentioned in previous studies. According to Lyons (1995), "prefer or absolute synonyms" refer to a pair of synonyms in which (a) "all meanings [being compared] are identical"; (b) two words are "synonymous in all contexts", and (c) they are "semantically equivalent on all dimensions of meaning, descriptive and non-descriptive" (p. 61). "Near synonyms" are defined by Cruse (1986) as "lexical items whose senses are identical in respect of 'central' semantic traits, but differ […] in 'minor' or 'peripheral traits'" (p. 237). Many researchers agree that most synonyms are likely to be near-synonyms. For example, Taylor (2003) argued that "it is commonly asserted that 'perfect', or 'full', synonyms do not exist, or if they do, they are exceedingly rare" (p. 265). Divjak & Gries (2006) also noted that "even if synonyms name one and the same thing, they name it in different ways; they present different perspectives on a situation" (p. 24). Thus "near synonyms" are more widely used by linguists than "synonyms". Since near-synonyms are "not fully intersubstitutable" (Inkpen & Hirts, 2006, p. 223), it is important to identify their variations in different contexts and perspectives through language analysis based on a corpus.

2.2 Corpus-Based Approach for Discriminating Near Synonyms

Many scholars found that near-synonyms always differ from a semantic, syntactic, or pragmatic point of view (e.g., Cruse, 1986; Taylor, 2003; Divjak & Gries, 2006). In terms of these differences, corpus linguists make extensive use of computers to conduct a frequency or statistical approach on linguistic features of near-synonyms, such as their comparisons in collocation, colligation, semantic preference, and semantic prosody. The four parameters, which take different values and go from concrete to abstract, are assumed to be the internal structure of words by Sinclair (1996). His core notion is that lexical meaning is not purely ascribed at the level of words because a word, as the unit of meaning, is related with other words around it (Sinclair, 2004, p. 27).

The first parameter is collocation, which is defined as "the items in the environment set by the span" (McEnery & Hardie, 2012, p. 107). In corpus studies, collocation is regarded more in terms of probability, where "the strength of a particular collocation is assessed on the basis of how frequently it appears in a large representative sample of discourse" (Walker, 2011). Moreover, many corpus linguists assume that the term only refers to "significant collocations" which co-occur more frequently than "their respective frequencies and the length of the text in which they appear would predict" (Sinclair et al., 2004, p. 10). For a precise degree of significance to each co-occurrence, statistical measures, such as the MI (mutual information), z, t, log-likelihood, log-log, MI3 scores, are used to measure collocational strength (Richard & Tony, 2006). The second parameter is colligation. The concept refers to "the interrelation of grammatical categories in syntactical structure" (Firth, 1957, p. 12). The difference between colligation and collocation is that the former helps to study a word's grammatical functions while the latter emphasizes a word's lexical inter-relations. One example of collocation is that the word *powerful* is likely to collocate in a large general corpus with concrete nouns like *cars*, *computers*, *countries*, while *strong* is more closely associated with abstract nouns and concepts, such as *sense*, *feeling*, *belief* (Castello, 2014). Another example for colligation is that *consequence* has a very low likelihood of appearing as the object of a clause in contrast to *preference* and *use* (Hoey, 2005). The third parameter is semantic preference. It means "by a lexical set of frequently occurring collocates [sharing] some semantic feature" (Stubbs, 2002, p. 449). For instance, Partington (2004, p. 148) found that "absence/change of state" is a common feature of the collocates of maximizers such as *utterly*, *totally*, *completely*, and *entirely*. This finding unveils that semantic preference is beneficial in developing a profile of a word and understanding how certain collocates can be "bound together in extended units of meaning" (Sinclair, 1996). The fourth parameter is semantic prosody. Louw (1993), who popularized the term, defined it as a "consistent aura of meaning with which a form is imbued by its collocates" (p. 157). This means that semantic prosody is viewed as affective meanings of a given word with its typical collocates (Stubbs, 2001). It can be favorable, neutral, and unfavorable prosodies (Partington, 2004). *Cause*, for example, is an unfavorable semantic prosody because it co-occurs regularly with words like *accident*, *cancer*, *death*, etc. (Stubbs, 1996, pp. 173-174).

In addition to the parameters above, the non-linguistic features can also be found in the corpus, such as varieties defined by register and periods of time (Biber et al., 1998). For example, Cai (2012) concluded that *awesome*, *fabulous*, and *fantastic* have increasingly been used over time. Regarding the genre, *fabulous*, *fantastic*, *great*,

terrific, and *wonderful* were observed more in the spoken genre, whereas *awesome* and *excellent* highly occurred in magazines.

Overall, the relevant literature shows that different methods are available for researchers to study near-synonyms and they can choose the approach that best fits their goals.

2.3 Studies on English Near Synonymous Verbs

In the past decades, several corpus-based approaches on near synonymous verbs have been established. In earlier times, Church et al. (1994) carried out a corpus-based analysis comparing *ask for*, *request,* and *demand* in terms of substitutability. Biber et al. (1998) differentiated *begin* and *start* in their grammatical construction with regard to their different lexical associations across registers using the Longman-Lancaster Corpus.

Recently, more powerful tools are available for corpus-based language research. Lee & Liu (2009) adopted *VIEW* as a tool to focus on the syntactic pattern to compare and contrast *affect/influence* gathering data from BNC and COCA. Using *Sketch Engine*, Hu & Yang (2015) and Yang (2016) analysed the collocation, concordance, word sketches and sketch difference of synonyms *raise* and *increase*, *learn* and *acquire* in British National Corpus. In the same fashion, Shahzadi et al. (2019) used *Sketch Engine* to examine *arrive* and *reach*. Adopting different online tools such as *Sketch Engine*, *BNC Web*, and *Just the Word*, Gu (2017) examined *gain* and *obtain* in genre, collocation, colligation, and semantic prosody.

In different studies, the English causative verbs *get* and *have* were investigated by Gilquin (2003), and intra- and extralinguistic factors in the contexts of *hassle, brother*, and *annoy* were compared by Glynn (2007). In addition, selectional and collocational restrictions of the linguistic meanings between *create* and *produce* were inspected by Chung (2011) in the Brown Corpus and the Frown Corpus. Covering local speakers' corpus LOB and non-local speakers' corpus CLEC, Rui (2016) differentiated between two English action words *start* and *begin*. Furthermore, Lin & Chung (2021) attempted to explore the syntactic and semantic information of two synonymous verbs *propose* and *suggest* in a specific genre gathering from COCA.

However, more studies are needed on a various set of synonyms and near-synonyms (Cai, 2012; Uba, 2015). Accordingly, this research sheds insight and understanding on how two near synonymous verbs *damage* and *destroy* work in terms of frequencies, genre distribution, colligation and collocation, differences in meanings and uses.

3. Method

3.1 Data Collection

All data in this study were collected from British National Corpus (BNC). This is a monolingual, synchronic, general, and sample-based type of corpus containing 100 million words. Data in this corpus covers 90% written and 10% spoken texts from disciplines of a wide range from 1960 to 1990. The written genre includes, for instance, extracts from regional and national newspapers, specialist periodicals and journals for all ages and interests, popular fiction and academic books, published and unpublished letters and memoranda, school and university essays, etc. The spoken genre contains, for example, orthographic transcriptions of unscripted informal conversations as well as spoken language collected in different contexts, which range from formal business or government meetings to phone-ins and radio shows.

3.2 Corpus Tool

Sketch Engine (SkE) is a powerful tool for corpus-based language research (Kilgarriff, et. al., 2004). It was first used in lexicography and then applied to other different fields such as translation, discourse analysis, language teaching, terminology (Kilgarriff, et. al., 2014). SkE provides easy access to many ready-to-use corpora, for example, BNC is one of the sub-corpora. It can be used to perform different functions. In the present study, such functions are used: Thesaurus, Concordance, Collocation, word sketches, and Sketch Diff. Thesaurus automatically generates a list of synonyms or words belonging to the same category (semantic field). Concordance provides concordance lines showing keywords in context, which helps to define lexical and structural information about the keyword. Collocation provides the span, the minimum frequency of each collocate, and the strength of collocation. Word sketches offer a one-page summary of both word's grammatical and collocational behavior. Sketch Diff offers collocation differences in a straightforward setting.

3.3 Analysis Procedure

Identify that the two verbs *damage* and *destroy* are similar by the tool Thesaurus.

The frequencies for *damage* and *destroy* in BNC were gathered by using concordance. From the frequency, we can know how many times two words are used in communication.

The genre in which *damage* and *destroy* were used is easily retrieved from BNC by using TEXT TYPES, which allows a researcher to look for genres and sub-genres where a word appear.

The colligation of *damage* and *destroy* in BNC were based on Word sketches, which present the grammatical patterns of the two verbs.

For transitive verbs such as *damage* and *destroy*, the researcher focused on the noun collocates and adverb collocates based on syntactic patterns (v+n, n+v, adv+v, v+adv). The positional constraint adopted in this research is the left and the right horizon of the keyword within a span of five words. Only those collocations with a minimum frequency of 10 or above in the given range (-5, 5) were considered. When the top list of most frequent collocates is retrieved, the collocates are further graded by their logDice scores, which is a reasonable, stable, and reliable interpretation (Rychly, 2008).

In order to get a better understanding of the words in question, the use of *damage* and *destroy* for the same reference in a given context is compared and examined, given that the collocation of near-synonyms with the same word can best show their differences in nature (Taylor, 2003).

In addition, this study pays attention to subtle meaning differences across *damage* and *destroy* based on examination in context.

4. Results and Analysis

4.1 Thesaurus for Identifying Damage and Destroy

The thesaurus entry for the verb *damage* is shown in Table 1.

Table 1. Lemma of similar words of the verb *damage* in BNC

Rank	Lemma	Score	Freq
1	destroy	0.343	6040
2	injure	0.214	2690
3	affect	0.214	13095
4	hurt	0.206	4708
5	undermine	0.198	2080
6	ruin	0.197	1675

Table 1 shows that the top six similar words of the verb *damage* in BNC are *destroy*, *injure*, *affect*, *hurt*, *undermine*, and *ruin*. The score is a percentage of the shared collocates and used to sort the verbs by their similarity to *damage*. *Destroy* appears on the top of the thesaurus. This means that the two verbs *destroy* and *damage* are relevant words and share most collocates.

4.2 The Frequencies of Damage and Destroy

It is necessary to figure out the overall frequency of two near synonymous words in a corpus. These are shown in Table 2.

Table 2. The frequency of *damage* and *destroy* in BNC

	damage	destroy
Total	3,296	6,040
Per million	29.34	53.76

From Table 2, we can see that *destroy* is more commonly used in both spoken and written communication than *damage*.

4.3 The Genre Difference of Damage and Destroy

Table 3 and Table 4 demonstrate the genre comparison of *damage* and *destroy* in terms of raw frequency and relative text type frequency in BNC. A frequency limit of 5 is chosen. Rel (%) (the number of relative text type frequency) means the relative frequency of the query result divided by the relative size of the particular text type. Above 100% refers to typical of this text type; below 100% is the opposite.

Table 3. Comparison of frequency of *damage*/*destroy* in different text types of BNC

Text Type	Damage/Freq. (Rel (%))	Destroy/Freq. (Rel (%))
Written books and periodicals	2,714 (102.1)	5,273 (108.2)
Written miscellaneous	339 (136.1)	355 (77.7)
Written-to-be-spoken	111 (249.3)	184 (225.5)
Spoken context-governed	103 (50)	176 (46.6)
Spoken demographic	29 (21)	52 (20.6)

Table 3 highlights that the occurrence of both *damage* and *destroy* in written books and periodicals is significantly higher compared with in any other text types. The frequency of *destroy* is greater than *damage* in all text types. Moreover, both *damage* and *destroy* are used more frequently in Written-to-be-spoken text types (TV news scripts) than in the corpus.

The detailed comparison of frequencies of *damage* and *destroy* in different written texts is shown in table 4.

Table 4. Frequency of *damage* and *destroy* in different written text types of BNC

Text Type	Damage/Freq. (Rel (%))	Destroy/Freq. (Rel (%))
Informative: world affairs	913 (158.3)	1,811 (171.3)
Informative: leisure	586 (143.1)	808 (107.6)
Informative: social science	512 (108.4)	541 (62.5)
Informative: applied science	358 (149.3)	483 (109.9)
Informative: commerce & finance	237 (96.6)	257 (57.2)
Imaginative	216 (38.8)	904 (88.6)
Informative: natural & pure science	152 (119)	243 (103.9)
Informative: arts	143 (64.9)	486 (120.4)
Informative: belief & thought	47 (45.7)	279 (148.1)

Table 4 manifests that both *damage* and *destroy* are more frequently used in world affairs (e.g., business, politics, juridical matters) than in the whole corpus. *Damage* is 1.08 times as common in social science (e.g., Health, History, and Philosophy of Science) of written English than in the whole corpus, which is significantly more as compared to the frequency of *destroy* in these texts. But it is less frequently used in informative texts related to belief & thought and arts-related texts. *Destroy* is 3 times more than *damage* in belief and thought, and 1.8 times more in arts.

4.4 The Colligation Difference of Damage and Destroy

In terms of colligation, *damage* and *destroy* as verbs are summarized in the following patterns based on the Word Sketch of SkE (see Table 5).

Table 5. The colligation difference of *damage* and *destroy*

Colligation	Frequency of Damage	Frequency Ratio of Damage	Frequency of Destroy	Frequency Ratio of Destroy
object	1,951	47%	3,387	54%
subject	540	13%	1,137	18%
modifier	976	24%	972	15%
prep phrases	678	16%	789	13%
Total	4,145	100%	6,285	100%

Note. object (V + n), subject (n + V), modifier (adv+ V, V +adv), pp (V+ prep +obj)

From Table 5, it can be found that *damage* and *destroy* mainly are collocated with object nouns. These two words share a similar frequency ratio in the pattern of "n + V" and "V+ prep +obj". However, the difference in colligation lies in modifier. It shows that *damage* is more frequently used with adverbs.

4.5 The Collocation Difference of Damage and Destroy

4.5.1 Nouns Collocates

The first part will focus on the pattern of "V + n" to see the collocation with noun given the high frequency. The following table 6 illustrates the top 32 right side nouns collocates of *damage* and *destroy*.

Table 6. Comparison of object noun collocation of *damage* and *destroy*

Rank	Damage Collocate	Freq	logDice	Rank	Destroy Collocate	Freq	logDice
1	reputation	38	7.44359	1	enemy	43	7.06323
2	ligament	15	7.11776	2	ozone	27	6.91805
3	ozone	18	7.01132	3	forest	30	6.3251
4	health	73	6.81648	4	missile	19	6.30908
5	environment	36	6.32915	5	bomb	25	6.29402
6	morale	10	6.2714	6	crop	23	6.29204
7	credibility	10	6.26089	7	weapon	25	6.11907
8	prospect	19	6.16013	8	aircraft	21	5.85074
9	knee	14	5.91728	9	credibility	12	5.80832
10	property	35	5.90762	10	habitat	12	5.73947
11	confidence	15	5.5892	11	livelihood	10	5.6774
12	economy	20	5.44196	12	cell	29	5.64617
13	skin	13	5.32255	13	army	22	5.64045
14	brain	10	5.24044	14	Iraq	14	5.63602
15	image	16	5.23254	15	myth	12	5.62659
16	cell	18	5.18346	16	bacterium	11	5.59834
17	career	13	5.10274	17	confidence	19	5.58682
18	relation	19	4.89437	18	reputation	14	5.53458
19	baby	12	4.84434	19	bridge	16	5.53138
20	chance	15	4.69363	20	building	36	5.5173
21	car	29	4.69288	21	environment	23	5.43544
22	building	17	4.59736	22	job	48	5.39448
23	plant	14	4.51094	23	planet	10	5.31367
24	fish	10	4.51005	24	character	22	5.3009
25	business	26	4.43395	25	evidence	31	5.21215
26	window	12	4.41943	26	ship	16	5.21138
27	hair	11	4.39528	27	evil	10	5.20542
28	relationship	14	4.3928	28	life	69	5.08952
29	patient	17	4.28682	29	career	14	4.92118
30	site	11	4.23114	30	ability	15	4.90381
31	structure	11	4.03422	31	dream	13	4.88066
32	interest	20	3.99669	32	hope	31	4.86752

From the table above, we can see that the dominant collocates of *damage* are object nouns which can be classified into six groups:

    a)   Human body or physical health: ligament, knee, skin, brain, hair, health;

    b)   Medical term: cell;

c) Natural environment: ozone, environment; fish, plant;

d) Commerce and Finance: property, economy, business;

e) Concrete concept: baby, patient, car, building, window, site;

f) Abstract concept: reputation, morale, credibility, prospect, confidence, image, career, relation, chance, relationship, structure, interest

It is also clear that the main categories of the right nouns collocates of *destroy* may be classified into six groups:

a) Military affairs: enemy, missile, bomb, weapon, army, Iraq;

b) Natural environment: ozone, forest, crop, habitat, environment;

c) Medical term: cell, bacterium;

d) Proper noun: planet;

e) Concrete concept: aircraft, bridge, building, ship;

f) Abstract concept: credibility, livelihood, myth, confidence, reputation, job, character, evidence, evil, life, career, ability, dream, hope

From the evidence of those object noun collocations above, it can be inferred that *damage* is more collocated with human body or physical health, while *destroy* is more contingent with military affairs. In addition, *destroy* is used more frequently with abstract nouns related with one's thought or belief, such as *myth*, *dream*, and *hope*. Skethc Diff echoes the findings. Table 7 shows the difference in terms of noun objects. It means that it is more usual to say "damage health" and "damage brain", while it is more fluent to say "destroy enemy" and "destroy army", or "destroy hope" and "destroy myth". It also explains in some degree the more frequent distribution of *damage* in the domain of social science (e.g., Health) as well as *destroy* in the domain of belief and thought as indicated in table 4.

Table 7. Comparison of object collocation

Collocation (Noun)	Frequency of Damage	Frequency of Destroy	logDice of Damage	logDice of Destroy
health	53	0	9.0	—
brain	45	0	8.9	—
ligament	15	0	7.9	—
skin	11	0	6.6	—
tissue	7	0	6.5	—
myth	0	10	—	6.4
dream	0	13	—	6.4
army	0	14	—	6.4
fleet	0	11	—	6.5
enemy	0	16	—	6.8
hope	0	23	—	7.0

The second part of difference lies in collocation with subjects. Sketch Diff shows the difference in following table (see Table 8). It can be observed that the subjects of *damage* can be natural disaster, such as *frost*, *smoke*, *heat*, *rain*, *water*, man-made disaster, such as *blast*, *vandal*, or abstract concepts, such as *episode*, *experience*, *statement*. The subjects of *destroy* are natural disaster, such as *earthquake*, *flood*, *disease*, man-made disaster, such as *mob*, or disaster of war, such as *explosion*, *bombing*, *aircraft*, *missile*, *force*, *irap*, *attack*, *army*. It once again demonstrates the relevance of *destroy* with the domain of military affairs (related with politics). It to some extent explains the high frequency of *destroy* used in world affairs as shown in table 4.

Table 8. Comparison of subject collocation

Collocation (Noun)	Frequency of Damage	Frequency of Destroy	logDice of Damage	logDice of Destroy
blast	5	0	7.8	—
frost	5	0	7.7	—
vandal	4	0	7.6	—
episode	4	0	7.1	—
smoke	4	0	6.8	—
heat	4	0	6.6	—
experience	7	0	6.0	—
statement	5	0	5.7	—
rain	4	0	5.5	—
water	4	0	4.6	—
army	0	5	—	5.5
attack	0	4	—	5.7
disease	0	5	—	5.7
iraq	0	4	—	6.0
force	0	12	—	6.1
missile	0	4	—	6.6
aircraft	0	10	—	6.9
mob	0	5	—	6.9
flood	0	7	—	7.3
bombing	0	9	—	7.8
explosion	0	12	—	8.0
earthquake	0	11	—	8.1

4.5.2 Comparison of Damage and Destroy for the Same Reference

Examples containing both *damage* and *destroy* for the same references were collected. The constraint was that one of the near synonyms appeared within five words on either side of another of the near synonyms in this group. The frequencies are shown in Table 9.

Table 9. Frequencies of collocations

Rank	Damage Collocate	Freq	logDice	Rank	Destroy Collocate	Freq	logDice
4	destroy	66	7.85581	10	damage	79	7.09121

Note. The "Rank" column shows the number which appears in the list of the top 20 collocates of the verb *damage* and *destroy*.

From this table, we can see that *damage* and *destroy* are used commonly to go with each other among native English speakers. This provides a good basis for comparison. The difference underpinned *damage* and *destroy* can be found out after exploring the KWIC in context (see examples).

Example 1

You will sometimes find that if your employer acts towards you in an unsatisfactory way, he is in breach of his implied duty of mutual trust and confidence. </s><s> Employers should not, without good cause, behave in a manner that is likely to **destroy** or seriously **damage** the employment relationship. </s><s> The law reports provide many illustrations of this kind of conduct.

Example 2

There were no bridges left standing between Verona and Rome and, according to Eric, who from now on was to travel this way once a week, every village south of Florence on the main road to Viterbo and Rome had been either **destroyed** or severely **damaged**. </s><s> The churches had suffered particularly badly. </s><s> The tally of religious paintings and statues lost was immense. </s><s> Everything was broken; it is difficult now to imagine the devastation.

In example 1 and 2, *destroy(ed)* and *seriously damage/ severely damaged* are used in a parallel structure. The two instances of *destroy* can be deemed as equivalents of *seriously/severely damage*. So, *destroy* is stronger than *damage* in degree of destruction.

Example 3

The more phosphate that is available, the more algae grow and multiply, sometimes to such an extent that they form a dense mat over the surface of the water. </s><s> The consequence of this is that aquatic life below is seriously **damaged** and even **destroyed**. </s><s> The algae on the surface prevent sunlight from reaching deeper plants so they don't grow well. </s><s> When they eventually start decaying, they use up oxygen dissolved in the water. </s><s> This, of course, makes life for aquatic animals extremely difficult.

Example 4

There are risks involved in credit trading. </s><s> Apart from the chance that the customer may default on his payments (perhaps even go bankrupt), there is the risk that he may also sell, **damage** or even **destroy** the goods. </s><s> In the typical transaction, it is the finance company which runs these risks. </s><s> The finance company can to some extent safeguard itself by including certain terms in the contract it makes with the customer.

In example 3 and 4, there is a building up in degree of "destruction" shown by "The consequence of this is that aquatic life below is seriously damaged and **even** destroyed." and "there is the risk that he may also sell, damage or **even** destroy the goods." This again proves that *destroy* is stronger in degree than *damage*.

Example 5

The whole Steam Tank is protected by a thick armoured skin, making it immune to fire from arrows and light missiles and impervious to blows from most warriors. </s><s> Machines which have been **destroyed** or **damaged** in battle have so far been recovered and rebuilt, although since Leonardo's disappearance many of the secrets of their construction have been lost and the engines become increasingly unreliable and inefficient.

Example 6

Rule (although weekly maintenance including 5-10% part water changes are also to be recommended) should keep your aquatic system and livestock in steady conditions and peak of health. </s><s> Never rinse reusable filter materials in tap water and never use detergents or you will **damage** or totally **destroy** the beneficial bacteria that have built up the efficiency of the filter unit. </s><s> Always rinse gently to remove the build-up of mulm and debris in a clean bucket of aquarium or pond water.

Example 7

They will also generate 50,000 new vehicle journeys a day, adding to pollution and local traffic congestion. </s><s> A total of 47 of the sites affected are ancient woodland, the RSNC claim. </s><s> Some of the areas which are not directly **destroyed** could be **damaged** to the point where they are no longer viable as areas of conservation importance, says the report. </s><s> The organization believes that the new plans could be in breach of European law requiring full environmental impact studies to be carried out.

In example 5, "Machines which have been destroyed or damaged in battle have so far been recovered and rebuilt, … and the engines become increasingly unreliable and inefficient.", two points are observed. First, *destroy* emphasizes something that no longer exists. This can be seen from its collocation with *rebuilt* in the sentence. What's more, in example 6, "… you will damage or totally destroy the beneficial bacteria…" the helper of the adverb "totally" provides information about the degree in which something disappears. Second, *damage* underscores something that can be recovered but does not work well as before. This can be proven by the cooccurrence of *recovered*, *unreliable* and *inefficient* in the sentence. In addition, in example 7, "Some of the areas which are not directly destroyed could be damaged to the point where they are no longer viable as areas of conservation importance", the clause following *damage* serves as a further explanation of the degree of destruction: "no longer viable", which means that something is not capable of doing what it is intended to do.

In a nutshell, from the comparison of *damage* and *destroy* in a given context, it can be identified that the use of *destroy* is to deepen the degree of destruction when native English speakers collocate the two near synonyms with the same word. What's more, the fact that *destroy* is stronger than *damage* in degree may explain why *damage* is more frequently used with adverbs as indicated in table 5. As Cai (2012) put it, "adverbs, functioning as intensifiers, are used to provide more information about degree or show emphasize, amplify, or down-tone." That is to say, certain words in English, such as *fabulous*, *perfect*, *destroy*, *catastrophe*, have extremely positive or negative meaning themselves even when unmodified. Thus, these words do not need much modification. Moreover, this fact may also explain why the subjects of *destroy* are natural disaster that will cause much greater destruction, such as *earthquake*, *flood,* while those of *damage* are *frost*, *rain*, *heat* and so on. And it is not surprising that we say "earthquake destroyed sth.", rather than "earthquake damaged sth.", or "frost damaged sth.", rather than "frost destroyed sth.". Another finding is that the meaning underpinned *destroy* is something that no longer exists, while *damage* implies something that can be recovered but does not work well as before.

4.5.3 Adverb Collocates

The near synonyms in question can be modified by adverbs. The collocation frequency in the "modifier" pattern in Sketch-Diff is shown in table 10.

Table 10. Comparison of adverb collocation

Collocation (Adverb)	Frequency of Damage	Frequency of Destroy	logDice of Damage	logDice of Destroy
badly	197	0	10.2	—
severely	111	0	10.2	—
seriously	87	0	9.0	—
extensively	27	0	8.7	—
irreparably	13	0	8.7	—
permanently	16	0	8.0	—
humanely	0	5	—	7.3
utterly	0	11	—	7.3
ruthlessly	0	6	—	7.3
virtually	0	27	—	7.7
totally	0	50	—	7.9
completely	0	72	—	8.0

Table 10 reveals that the adverb to describe *damage* can be categorized into adverbs of degree, such as *badly*, *severely*, *seriously*, *extensively*, *irreparably*, *permanently*, which are used to intensify the degree of destruction for *damage*. The common adverbs to describe *destroy* are adverbs of positive denotation such as *humanely*, adverbs of negative denotation such as *ruthlessly*, and adverbs of degree, such as *completely*, *totally*, *virtually*, *utterly*, which are used to provide more information about degree for *destroy*. The findings correspond to results of the comparison of *damage* and *destroy* in a given context in the last section.

5. Conclusion and Implication

Based on the corpus analysis, the differences of two near synonymous verbs, *damage* and *destroy*, lie in frequency, genre, colligation, collocation, subtle meanings and uses. This study echoes previous claims that near synonyms are "not fully intersubstitutable". The overall findings can be summarized as follows:

First, although near synonym *destroy* has wider occurrence in BNC than *damage*, they both tend to occur more often in written books and periodicals, and TV news scripts.

Second, there is a significant difference in collocation of *damage* and *destroy* with different nouns. *Damage* is more collocated with human body or physical health, which partly explains its more frequent usage in social science (e.g., Health). *Destroy* is more widely used with military affairs and one's thought or belief. It is often used in the text types of world affairs (e.g., politics) and belief and thought.

Third, the British tend to collocate the two near-synonyms with the same word to create a build-up, because *destroy* is endowed with a stronger degree of destruction than *damage*. This may be the potential reason why *damage* is more frequently used with intensifying adverbs, and why *destroy* is more generally used with natural

disaster, such as *earthquake*, *flood*, while *damage* is used more with *frost* or *rain*, and so on. What is also noticeable is the fact that the use of *destroy* tends to refer to something that no longer exists and the use of *damage* for something that can be recovered but does not work well as before.

The findings of this study implicate that the use of corpora is a good supplementary tool for not only beginners but also advanced learners when they feel confused between and among lexical choices, especially near-synonyms, of their second language. A corpus study is of great value because traditional dictionaries and language teaching will not suffice for vocabulary learning. First, the traditional dictionaries and language teachers failed to add genre distribution. Genre is essential because it helps language learners to recognize a particular setting and communicative function for words (e.g., *damage* and *destroy* are often used for news report). Second, the dictionaries and teachers may not provide the frequently used collocation, such as "destroy myth". As a consequence, language learners are likely to have vocabulary issues, for example, they misuse synonyms, in speaking or writing. Third, the dictionaries and teachers have not always noted important context information related to words, especially authentic usage of near-synonyms, though they have made an attempt to give illustrative sentences. For native speakers, they are able to acquire vocabulary naturally in first-language contexts; however, for second language learners, especially those where there is little or no exposure to first-language contexts, it is necessary to provide the authentic context to a greater extent. Therefore, corpus-based in language teaching and learning should be promoted and generalized so that it could be beneficial for more people.

Acknowledgments

References

Ahmad, S. S., Mahmood, M. A., & Ahmad, M. (2019). A corpus based analysis of conflicting synonyms in technical writings. *International Journal of Educational sciences, 25*(1-3). https://doi.org/10.31901/24566322.2019/25.1-3.1078

Albader, Y. B. (2001). *Comparing the usage and distribution of two near-synonymous words: Error vs. mistake.* (MA), University of Manchester.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511804489

Cai, J. (2012). *Is it "Great" Enough? A corpus-based Study of "Great" and its Near Synonyms.* (Unpublished MA thesis), Ball State University, Muncie: Indiana.

Castello, D. (2014). *A Corpus Study of Strong and Powerful.* (Unpublished MA), University of Birmingham Edgbaston.

Chung, S-F. (2011). A corpus-based analysis of "create" and "produce". *Chang Gung Journal of Humanities and Social Sciences, 4*(2), 399-425. Retrieved from https://www.researchgate.net/publication/265812963

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

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

Devitt, M. (2006). Linguistic Intuitions Revisited. *The British Journal for the Philosophy of Science, 61*(4), 833-865. https://doi.org/10.1093/bjps/axq018

Divjak, D., & Gries, S. Th. (2006). Ways of trying in Russian: Clustering behavioral profiles. *Journal of Corpus Linguistics and Linguistic Theory, 2*(1), 23-60. https://doi.org/10.1515/CLLT.2006.002

Firth, J. (1957). *Papers in linguistics*. Oxford: Oxford University Press.

Flowerdew, J. (2013). *Discourse in English Language Education*. London, England: Routledge. https://doi.org/10.4324/9780203080870

Gilquin, G. (2003). Causative 'get' and 'have': So close, so different. *Journal of English Linguistics, 31*(2), 125-148. https://doi.org/10.1177/0075424203031002002

Glynn, D. (2007). *Mapping meaning. Towards a usage-based methodology in Cognitive Semantics.* (Unpublished doctoral dissertation). University of Leuven, Leuven, Belgium.

GU, B. (2017). Corpus-based study of two synonyms-obtain and gain. *Sino-US English Teaching, 14*(8), 511-522. https://doi.org/10.17265/1539-8072/2017.08.006

Hoey, M. (2005). *Lexical Priming: a new theory of words and language*. Routledge. https://doi.org/10.4324/9780203327630

Hu, C., & Yang, B. (2015). Using Sketch Engine to investigate synonymous verbs. *International Journal of English Linguistics, 5*(4), 29-41. https://doi.org/10.5539/ijel.v5n4p29

Inkpen, D., & Hirst, G. (2006). Building and using a lexical knowledge base of near synonym differences. *Computational Linguistics, 5*(3), 223-262. https://doi.org/10.1162/coli.2006.32.2.223

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography, 1*(1), 7-36. https://doi.org/10.1007/s40607-014-0009-9

Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). *The Sketch Engine*. Proc Euralex, Lorient, France. Retrieved from https://www.researchgate.net/publication/267223981

Lee, H-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. Retrieved from https://www.aclweb.org/anthology/O09-4004

Leech, G. (1991). The state of the art in corpus linguistics. In Aijmer, K. & Altenberg, B. (Eds.), *English Corpus Linguistics*. London: Longman.

Lin, Y. Y., & Chung, S. F. (2021). A Corpus-Based Study on Two Near-Synonymous Verbs in Academic Journals: PROPOSE and SUGGEST. *English Teaching & Learning*. https://doi.org/10.1007/s42321-020-00072-0

Liu, D., & Espino, M. (2012). Actually, Genuinely, Really, and Truly: A corpus-based Behavioral Profile study of near-synonymous adverbs. *International Journal of Corpus Linguistics, 17*(2), 198-228. https://doi.org/10.1075/ijcl.17.2.03liu

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In Baker, M., Francis, G. & Tognini-Bonelli, E. (Eds.), *Text and Technology* (pp. 157-176). Amsterdam: John Benjamins. https://doi.org/10.1075/z.64.11lou

Lyons, J. (1995). *Linguistic Semantics: An Introduction*. Cambridge University Press. https://doi.org/10.1017/CBO9780511810213

Mackay, S. (1980). Teaching the syntactic, semantic, and pragmatic dimensions of Verbs. *TESOL Quarterly, 14*, 17-26. https://doi.org/10.2307/3586805

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511981395

Oxford Advanced Learner's Dictionary. (2009). Oxford: Oxford University Press.

Partington, A. (2004). Utterly content in each other's company: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics, 9*(1), 131-156. https://doi.org/10.1075/ijcl.9.1.07par

Richard, X., & Tony, M. E. (2006). Collocation, semantic prosody, and near synonymy: a cross-linguistic perspective. *Applied Lingus, 27*(1), 103-129. https://doi.org/10.1093/applin/ami045

Rui, W. (2016). *A Corpus-based Comparison Study of Synonym by Chinese EFL Learners: A case study of begin and start*. Retrieved from https://en.cnki.com.cn/Article_en/CJFDTOTAL-JZGC201619069.htm

Rychly, P. (2008). A lexicographer-friendly association score. In Sojka, P. & Horak, A. (Eds.), *Proceeding of Recent Advances in Slavonic Natural Language Processing, RASLAN 2008*. [Online]. Brno: Masaryk University, pp. 6-9. Retrieved from https:// www.fi.muni.cz/sojka/download/raslan2008/13.pdf

Shahzadi, A., Asghar, A., & Javed, S. (2019). Effectiveness of Croups in Teaching English Synonyms. *Journal of Corpus Linguistics-CJCL, 2*(1). Retrieved from https://www.researchgate.net/publication/340333700

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (1996). The Search for Units of Meaning. *TEXTUS, 9*(1), 75-106.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge. https://doi.org/10.4324/9780203594070

Sinclair, J., Jones, S., Daley, R., & Krishnamurthy, R. (2004). *English Collocational Studies: The OSTI Report*. London: Continuum.

Stubbs, M. (1996). *Text and Corpus Analysis*. Oxford: Blackwell.

Stubbs, M. (2001). *Words and Phrases*. Oxford: Blackwell.

Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics, 7*(2), 215-244. https://doi.org/10.1075/ijcl.7.2.04stu

Taylor, J. R. (2003). Near synonyms as co-extensive categories: 'high' and 'tall' revisited. *Language Sciences, 25*(3), 263-284. https://doi.org/10.1016/S0388-0001(02)00018-9

Uba, S. Y. (2015). A corpus-based behavioural profile study of near-synonyms: important, essential, vital, necessary and crucial. *International Journal of English Language and linguistics Research, 3*(5), 9-17.

Walker, C. (2011). How a corpus-based study of the factors which influence collocation can help in the teaching of business English. *English for Specific Purposes, 30*(2), 101-112. https://doi.org/10.1016/j.esp.2010.12.003

Williams, G. (2003). From meaning to words and back: Corpus linguistics and specialised lexicography. *ASp,* 39-40. https://doi.org/10.4000/asp.1320

Yang, B. (2016). A corpus-based comparative study of learn and acquire. *English Language Teaching, 9*(1). https://doi.org/10.5539/elt.v9n1p209