



Semantic Distribution and Collocation: Corpora Study of Synonyms *See* and *Watch*

¹Adibah Adibah, ²Beta Rida Pasaribu, ³Siti Sumarti

^{1,2,3} Program Studi S1 Pendidikan Bahasa Inggris, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Mataram, Indonesia

adibah@staff.unram.ac.id, betarida@staff.unram.ac.id, ssitis099@staff.unram.ac.id

INFO ARTIKEL

Riwayat Artikel:

Diterima: 30-06-2026

Disetujui: 02-07-2026

Kata Kunci:

distribusi semantik;

kolokasi;

korpus pembelajar;

sinonim

Keywords:

collocation;

learner corpus;

semantic distribution;

synonymy

ABSTRACT

Abstract Synonymy has long been a source of confusion for English language learners, particularly in distinguishing verbs that share similar meanings, such as *see* and *watch*. This study aims to use corpora to investigate the distinctions between these two synonymous verbs in terms of collocations and semantic distributions. Data were drawn from two written learner corpora: LOCNESS (The Louvain Corpus of Native English Essays), comprising 176 essays (149,574 words) written by American English learners, and GLC (Gachon Learner Corpus), comprising 212 essays (25,247 words) written by intermediate Korean English learners. Analysis was conducted using AntConc 3.5.9, applying loglikelihood statistics within a 4L-4R span, and cross-verified through the Collins Online Dictionary and Oxford Learner's Dictionary. The findings revealed that *see* was used considerably more often than *watch* in both corpora, consistent with its broader semantic range (14 senses in Collins and 18 in OLD, compared to 5 senses for *watch*). Regarding semantic distribution, Korean learners tended to favor literal, transparent senses of both verbs, while American learners employed a wider range of senses, including more opaque meanings of *see*. These findings carry important pedagogical implications, emphasizing the need to teach synonyms across their literal to the most opaque senses, and highlighting the value of corpus-based approaches in vocabulary instruction.

Abstrak Sinonim telah lama menjadi sumber kebingungan bagi pembelajar bahasa Inggris, terutama dalam membedakan verba yang memiliki makna serupa, seperti *see* dan *watch*. Penelitian ini bertujuan memanfaatkan korpus untuk menyelidiki perbedaan antara dua verba sinonim tersebut dari segi kolokasi, distribusi semantik, dan preferensi semantik (*semantic prosody*). Data dikumpulkan dari dua korpus tertulis, yaitu LOCNESS (*The Louvain Corpus of Native English Essays*) yang terdiri atas 176 esai (149.574 kata) yang ditulis oleh pembelajar bahasa Inggris Amerika, dan GLC (*Gachon Learner Corpus*) yang terdiri atas 212 esai (25.247 kata) yang ditulis oleh pembelajar bahasa Inggris Korea tingkat menengah. Analisis dilakukan menggunakan AntConc 3.5.9 dengan ukuran statistik loglikelihood pada rentang 4L-4R, serta diverifikasi melalui Collins Online Dictionary dan Oxford Learner's Dictionary. Hasil penelitian menunjukkan bahwa *see* digunakan jauh lebih sering dibandingkan *watch* pada kedua korpus, sejalan dengan cakupan makna *see* yang lebih luas (14 makna dalam Collins dan 18 makna dalam OLD, dibandingkan 5 makna untuk *watch*). Dari segi distribusi semantik, pembelajar Korea cenderung menggunakan makna literal dan transparan untuk kedua verba, sementara pembelajar Amerika menggunakan variasi makna yang lebih luas, termasuk makna yang lebih figuratif untuk *see*. Temuan ini memiliki implikasi pedagogis penting, yaitu perlunya pengajaran sinonim secara bertahap dari makna yang paling literal hingga makna yang lebih figuratif, serta pentingnya pendekatan berbasis korpus dalam pembelajaran kosakata bahasa Inggris.



A. INTRODUCTION

Synonymy has been a source of confusion for English language learners. Generally, it is challenging for them to differentiate between words that share similar senses, such as the most frequent verbs, *see* and *watch*. For instance, learners can write or say "I go seeing a movie" and "I go watching a movie" interchangeably, but they cannot write or say "I go watching a Broadway show." This difficulty is not unique to *see* and *watch*; near-synonymous verbs are widely recognized as a major source of errors for second-language learners, since the appropriate use of verb collocations constitutes one of the central challenges in L2 acquisition. Moreover, although dictionaries and thesauruses often present near-synonyms as though they were interchangeable, their semantic preferences and preferred syntactic structures are often underrepresented or only vaguely implied, which leaves learners with little explicit guidance on how to use such words naturally and appropriately. It is then essential to investigate the distinction between the two synonym verbs so that language learners can use them naturally and appropriately. Thus, this study aims to utilize corpora as tools to investigate the distinctions between two synonymous verbs, *see* and *watch*, in terms of their collocations, semantic distribution, and semantic preference or prosody.

A large number of studies have explored synonymous verbs in English by using a corpus-based approach (Phoocharoensil, 2021; Song, 2021; Yang, 2016). Song (2021) and Yang (2016) collected their data from the BNC (British National Corpus) and used SkE (Sketch Engine) to compare the synonyms of *damage* and *destroy* and *learn* and *acquire*, respectively. Yang (2016) found that various features of SkE can assist in distinguishing the two synonymous verbs, *learn* and *acquire*, by analyzing their concordance, collocation, word sketches, and sketch difference. One of the findings showed that although the two verbs share most collocates, *knowledge* and *skills* are nouns that most probably co-occur with *learn*, whereas *qualification* and *assets* collocate with *acquire*.

Similarly, Song (2021) examined the two near-synonymous words, *damage* and *destroy*, by analyzing their collocations, usage, senses, and genre distribution. It was found that *destroy* has a higher

degree of destruction than *damage* because the first verb refers to something that can no longer exist while the latter denotes something that can be fixed. Further, the results showed that *destroy* collocates with one's beliefs and thoughts as well as with military affairs while *damage* collocates with the human body and physical health.

Unlike the two previous studies, Phoocharoensil (2021) investigated collocations, genre distributions, and semantic preference or prosody of two near-synonym verbs, *persist* and *persevere*, in COCA (the Corpus of Contemporary American English). This study revealed that *persist* has collocations with negative senses of constantly uncomfortable circumstances, while *persevere* most probably co-occurs with words expressing a strong will to accomplish challenging tasks. This line of inquiry has continued to develop in more recent corpus-based research. Dushku and Paek (2021), studying 123 ESL learners ranging from intermediate to high proficiency, found that learners exhibited noticeable difficulty producing appropriate semantic prosody, even when they were able to recognize it, underscoring the pedagogical urgency of corpus-informed synonym instruction. In a related study, researchers examined the collocates, grammatical patterns, and semantic preferences of the near-synonymous verbs *affect* and *impact* using Sketch Engine to analyze lexical collocates, colligational profiles, and semantic prosody. Likewise, a corpus-based comparison of the near-synonymous verbs *type*, *kind*, and *sort* drew on online dictionary definitions alongside COCA-derived collocation and Mutual Information scores, finding that despite their semantic similarity, the three words could not be used interchangeably, a conclusion consistent with other related corpus studies on near-synonymous English vocabulary. Most recently, a corpus-based study extended this approach to the synonymous verbs *grasp*, *capture*, *seize*, *snatch*, and *take*, investigating their genre usage, collocational patterns, and semantic preferences, reaffirming that near-synonyms remain a persistent source of confusion for second language learners precisely because they are not interchangeable across most contexts, and substituting one for another risks misinterpretation or vagueness of meaning.

The studies above have examined English synonyms primarily in single corpora such as BNC or

COCA, and largely among general or native-speaker populations rather than specific groups of L1 English learners. This study aims to further the investigation by comparing collocation, semantic distribution, and semantic prosody of two synonymous verbs, *see* and *watch*, in two written corpora, LOCNESS (The Louvain Corpus of Native English Essays) and GLC (Gachon Learner Corpus).

B. METHOD

This study used two corpora, LOCNESS and GLC, to compare collocations, semantic distribution, and semantic preferences across two populations of L1 English learners. It collected 176 essays (149,574 words) written by American English learners in LOCNESS and 212 essays (25,247 words) written by intermediate Korean English learners in GLC. Purposive sampling was utilized to enable valid interpretation and generalizability of the study to the two specific populations.

After the data were collected from the two corpora, they were analyzed using AntConc 3.5.9 to determine the frequencies of *see* and *watch* and their collocations within the span of 4L-4R based on the loglikelihood statistical measure. Then, the concordance lines for these synonyms were entered into the Excel sheet and categorized into different senses by consulting two dictionaries, the Collins online dictionary and the Oxford Learner Dictionary (OLD). By doing so, the study examined the similarities or differences of the verbs used by these L1 English learners. Finally, to examine the semantic preference of synonyms, i.e., "a semantic environment in which vocabulary items commonly occur" (Phoocharoensil, 2021, p. 244), this study focused on nouns that co-occur on the right sides of the node (1R-4R) by comparing concordance lines in the two corpora.

This study had several limitations. First, because it collected data from only two small corpora (a total of 174,821 words), the findings may not generalize to larger populations of American and Korean English learners. Further, the specific topics of essays in the two corpora, i.e., abstract and philosophical issues in LOCNESS and general daily topics in GLC, may also influence words elicited in their writings. Besides differences in topic, their significantly different levels of proficiency may also affect the results of

collocation, semantic distribution, and semantic preferences of the two synonyms.

C. ANALYSIS AND DISCUSSION

1. The Frequency of SEE and WATCH in GLC and LOCNESS Corpora

The frequency of *see* and *watch* is searched for using their lemmas in AntConc 3.5.9. The regular expressions *see/saw/seen/sees/seeing* and *watch/watched* capture all the lemmas. The results show that, out of 25,357 tokens in GLC, Korean intermediate English learners used 82 of SEE and 34 of WATCH in their essays. Further, out of 149,574 tokens, American learners wrote their essays with 246 of SEE and 46 of WATCH.

Table 1. Normalized Frequency of SEE and WATCH in GLC and LOCNESS corpora (per 10,000 tokens)

	SEE	WATCH
GLC	32	13
LOCNESS	16	3

In their normalized frequencies as shown in Table 1, SEE was used almost three times as often as WATCH in GLC and more than five times as often in LOCNESS. The higher frequency of the first synonym is strongly related to its nature of having more senses than the latter.

2. The Collocation of SEE and WATCH in GLC and LOCNESS Corpora

Besides frequencies of the synonyms, this study also searches their collocations by measuring the Loglikelihood statistics with the span of 4L-4R in the AntConc version 3.5.9. There are 653 collocate tokens with 276 collocate types of SEE, and 268 collocate tokens with 115 collocate types of WATCH in GLC. Further, LOCNESS has 1,968 collocate tokens with 684 collocate types of SEE, and 368 collocate tokens with 178 collocate types of WATCH.

Table 2. Normalized Collocations of SEE and WATCH in GLC and LOCNESS corpora (per 10,000 tokens)

	SEE	WATCH
GLC	258	106
LOCNESS	132	25

As shown in Table 2, the normalized total number of collocations indicates that the word *see* consistently has more collocations than its synonym in the two corpora. In other words, as *see* is mainly used in the two corpora, it can attract many co-

occurring words. The total collocations seem to align with the total frequencies.

Table 3. Top 10 Collocations of SEE in GLC and LOCNESS Corpora

GLC	Rank	Freq	Freq (L)	Freq (R)	Stat	Collocate
SEE	1	37	31	6	225.68	i
	2	29	7	22	182.06	the
	3	15	3	12	146.02	movie
	4	17	9	8	114.29	that
	5	13	11	2	114.23	children
	6	8	8	0	92.141	ever
	7	12	12	0	89.471	when
	8	12	7	5	78.684	they
	9	8	1	7	72.29	movies

LOC NES S	Rank	Freq	Freq (L)	Freq (R)	Stat	Collocate
SEE	1	112	39	73	365.42	the
	2	57	9	48	339.26	as
	3	54	11	43	202.19	a
	4	37	28	9	181.79	they
	5	30	27	3	181.5	can
	6	39	13	26	181.34	it
	7	30	25	5	170.22	i
	8	54	45	9	166.77	to
	9	25	15	10	155.2	what

As shown in Table 3, the verb *see* revealed different kinds of words in the two corpora. GLC showed many content words, such as *movie*, *children*, and *movies*, compared to LOCNESS. Interestingly, as shown in Table 4, in terms of *watch*, both corpora contained many content words, such as *movies*, *violent*, *children* in GLC, and *television*, *favorite*, *children* in LOCNESS. The word *movie(s)* appears in the top 10 words of *see* and *watch*, showing that this word is one of the most probable collocations in the GLC. This study will also examine the semantic distribution and preferences of the two synonyms in the following sections to confirm this finding.

Table 4. Top 10 Collocations of WATCH in GLC and LOCNESS corpora

GLC	Rank	Freq	Freq (L)	Freq (R)	Stat	Collocate
WA TCH	1	20	0	20	264.56	movies
	2	19	0	19	238.64	violent
	3	14	10	4	154.34	children
	4	8	8	0	133.52	allowed
	5	8	8	0	71.537	not
	6	8	8	0	71.208	be
	7	10	10	0	64.916	to
	8	8	4	4	62.405	they
	9	10	4	6	62.083	the

LOC NES S	Rank	Freq	Freq (L)	Freq (R)	Stat	Collocate
WA TCH	1	23	20	3	92.32	to
	2	7	2	5	59.521	television
	3	20	7	13	51.838	the
	4	3	0	3	41.236	favorite
	5	9	5	4	39.712	for
	6	5	5	0	38.829	child
	7	5	5	0	33.245	children
	8	10	4	6	31.166	a
	9	3	0	3	30.729	v

3. Semantic Distribution of SEE and WATCH in GLC and LOCNESS Corpora

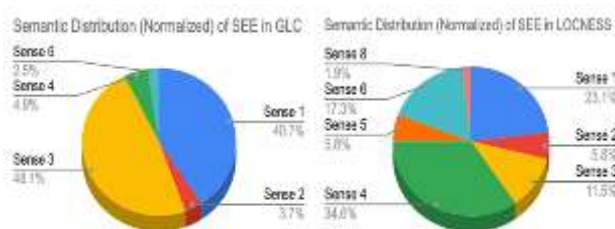
Before exploring the semantic distribution of the two synonyms, the definitions of *see* and *watch* are discussed. According to the Collins online dictionary, the verb *see* has 14 meanings, while *watch* has 5. The first three entries of Collins depict more transparent senses of *see* while the last eleven entries display more opaque ones, for example, "I haven't been to *see* a movie in 10 years" (entry 3) vs. "I really don't *see* any reason for changing it" (entry 5). Further, the first two entries of *watch* demonstrate literal senses. In contrast, the last three entries show more metaphorical ones, for instance, "I'd stayed up late to *watch* the movie" (entry 2) vs. "Human rights groups have been closely *watching* the case" (entry 3).

To enrich the results, the study also checks their senses in the Oxford Learner's Dictionary (OLD) and finds that the verb *see* yields 18 meanings while *watch* shows 5 meanings. The first seven entries of *see* in OLD depict more transparent senses of *see*, while the remaining eleven meanings show more opaque meanings, for instance, "In the evening we went to *see* a movie" (entry 3) vs. "He changed the way we *see* the world around us" (entry 9). Similarly, the first two entries of *watch* show literal senses while the third to fifth entries demonstrate more metaphorical ones, for example, "I was in the living room, *watching* TV" (entry 1) vs. "*Watch* yourself!" (entry 3). Overall, the results of the two dictionaries show that the

meanings of the words gradually shift from transparent to opaque and that the word *see* covers a broader range of senses than *watch*.

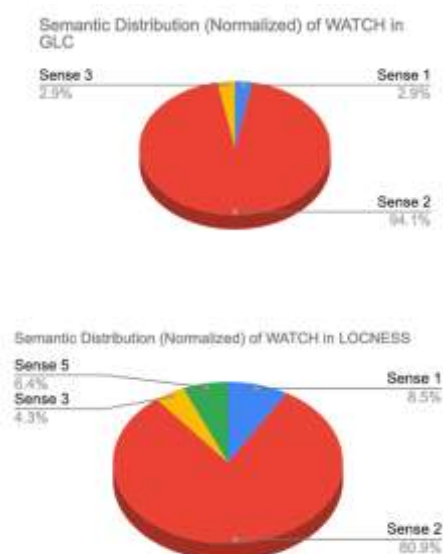
Further, based on the two references, the synonym of *see* and *watch* only overlaps in one specific context. First, Collins shows that the two verbs overlap in its third entry of *see*, "If you **see** an entertainment such as a play, movie, concert, or sports game, you **watch** it." and its second entry of *watch* "If you **watch** something on television or an event such as a sports contest, you spend time looking at it, especially when you **see** it from the beginning to the end." Further, OLD also confirms this and shows that senses of *see* and *watch* only overlap in one meaning, in its third entry of *see* "to **watch** a game, television program, performance, etc." and the first entry of *watch* "to look at somebody/something for a time, paying attention to what happens."

Chart 1. Semantic Distribution of *see* in GLC and LOCNESS (Normalized per 100 tokens)



As shown in Chart 1, Korean intermediate learners in GLC used 5 of the 14 meanings of *see* in Collins, whereas American learners used 7. Interestingly, while the central tendency among Korean learners was toward sense 3 (48.1%) and sense 1 (40.7%), which indicate transparent meanings, American learners tended to use more opaque definitions of *see*, like sense 4 (34.6%) and sense 6 (17.3%), while maintaining the moderate use of sense 1 (23.1%). The results show that American learners used a wider variety of senses in their writing than Korean learners did. Further, Korean learners tend to write in more literal terms, like perceiving with one's eyes (entry 1) and watching a movie (entry 3).

Chart 2. Semantic Distribution of *watch* in GLC and LOCNESS (Normalized per 100 tokens)



Interestingly, as shown in Chart 2, both Korean and American learners in GLC and LOCNESS dominantly used sense 2 in the Collins dictionary, which refers to "If you **watch** something on television or an event such as a sports contest, you spend time looking at it, especially when you **see** it from the beginning to the end." While the results show similar distributions, American learners used 4 of 5 meanings, whereas Koreans used only 3 out of 5 in their writings. Similar to the findings of *see*, American learners used a slightly greater variety of senses of *watch* when writing their essays.

D. CONCLUSION

This study investigates the synonymous verbs, *see* and *watch*, in two learner corpora, GLC and LOCNESS. Notably, it examines the synonym's frequency, collocations, and semantic distribution in the two corpora. Hence, this study has three research questions. First, in terms of frequency, both corpora show the high use of *see* compared to its synonym, *watch*. This finding is reasonable because *see* has a broader range of definitions (14 in Collins and 18 in OLD) than *watch* (5 in Collins and OLD). Next, regarding its collocations, while *watch* co-occurs with many content words in both corpora, *see* has different results. In GLC, *see* collocates with many content words, whereas in LOCNESS it collocates with plenty of function words.

In terms of semantic distribution, Korean intermediate learners tend to use more transparent meanings of *see* and *watch* in their essays. Interestingly, while American learners tend to utilize more literal meanings of *watch*, they use more metaphorical meanings of *see* in their writings. Furthermore, since the two dictionaries could not distinguish between the two synonyms, findings on semantic preference would assist English learners with this distinction (Alanazi, 2022; Phoocharoensil, 2021).

As mentioned in the limitation, the topics discussed in the two corpora may restrict the results of this study, i.e., collocations and semantic distribution. The different topic selections and genres may also limit the usage of *see* and *watch* for these learners. Further, their different proficiency level, intermediate for Koreans and native for Americans, may also limit their knowledge and familiarity with a wide range of senses of these two synonyms. Most importantly, because the data from the two corpora are relatively small, the results cannot be generalized to a wider population of Korean intermediate or American learners. In sum, the findings of this study need to be interpreted with caution.

The pedagogical implication is that it is important for English educators to gradually teach a variety of senses from the most to the least frequent meanings. Intermediate English learners, such as Korean learners in GLC, generally understand and use the literal meanings well in their spoken and written production. Thus, it is a great starting point for introducing more opaque word senses to help them distinguish the synonyms. It is also beneficial to incorporate a corpus-learning approach into English instruction to enable English learners to use corpus features, such as concordance lines, to become automatic learners.

REFERENCES

- [1] Alanazi, Z. (2022). Corpus-based analysis of near-synonymous verbs. *Asian-Pacific Journal of Second and Foreign Language Education*, 7(1), Article 15. <https://doi.org/10.1186/s40862-022-00138-5>
- [2] Carlstrom, B., & Price, N. (2012–2014). The Gachon learner corpus. Retrieved from <http://koreanlearnercorpusblog.blogspot.kr/p/corpus.html>
- [3] Collins dictionary. (n.d.). In Collins online dictionary. Retrieved June 30, 2026, from <https://www.collinsdictionary.com>
- [4] Dushku, S., & Paek, Y. (2021). Investigating ESL learners' awareness of semantic prosody across proficiency levels. *Language Awareness*, 30(3), 234–256. <https://doi.org/10.1080/09658416.2020.1871360>
- [5] Granger, S., & Wynne, M. (1998). Louvain corpus of native English essays (LOCNESS) [Corpus]. Centre for English Corpus Linguistics, Université catholique de Louvain. <https://www.uclouvain.be/en/research-institutes/ilc/cecl/locness>
- [6] Oxford Learner's Dictionary. (n.d.). In Oxford learner's dictionaries. Retrieved June 30, 2026, from <https://www.oxfordlearnersdictionaries.com/>
- [7] 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>
- [8] Song, Q. (2021). Effectiveness of corpus in distinguishing two near-synonymous verbs: "Damage" and "Destroy." *English Language Teaching*, 14(7), 8–20. <https://doi.org/10.5539/elt.v14n7p8>
- [9] Yang, B. (2016). A corpus-based comparative study of "Learn" and "Acquire." *English Language Teaching*, 9(1), 209–220. <https://doi.org/10.5539/elt.v9n1p209>